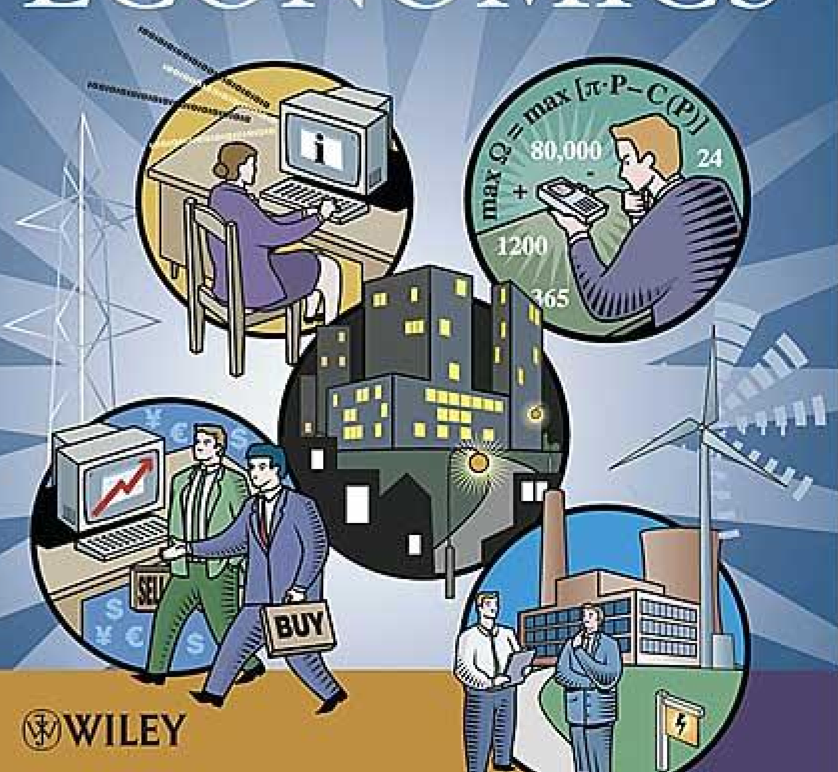
FUNDAMENTALS OF

POWER SYSTEM ECONOMICS



WILEY

DANIEL S. KIRSCHEN | GORAN STRBAC

Fundamentals of Power System Economics

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

Fundamentals of Power System Economics

Daniel Kirschen
Goran Strbac
University of Manchester Institute of Science & Technology (UMIST), UK

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-84572-4

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

For Penny and Philippe
For Dragana, Jelena and Anna

Contents

Preface

For about a hundred years, the electricity supply industry was in the hands of vertically integrated monopoly utilities. During that time, engineers treated the management of this industry as a set of challenging optimization problems. Over the years, these optimization problems grew in size, complexity and scope. New algorithms were developed, and ever more powerful computers were deployed to refine the planning and the operation of the power systems. With the introduction of competition in the electricity supply industry, a single organization is no longer in charge. Multiple actors with divergent or competing interests must interact to deliver electrical energy and keep the lights on. Conventional optimization problems are often no longer relevant. Instead, dozens of new questions are being asked about a physical system that has not changed. To deliver the promised benefits of competition, old issues must be addressed in radically new ways. To stay in business, new companies must maximize the value of the service they provide. Understanding the physics of the system is no longer enough. We must understand how the economics affect the physics and how the physics constrain the economics.

An environment with many independent participants evolves very rapidly. Over the last two decades, hundreds of technical papers, thousands of reports and a few books have been written to discuss these new issues and to propose solutions. The objective of this book is not to summarize or repeat what is in these documents. Instead, we have chosen to concentrate on delivering a clear and in-depth explanation of the fundamental issues. Our aim is to give the readers a solid understanding of the basics and help them develop innovative solutions to problems that vary in subtle ways from country to country, from market to market and from company to company. Therefore, we do not discuss the organization of specific markets. Neither do we attempt to describe all the solution techniques that have been proposed.

The plan of this book is simple. After introducing the participants in a restructured electricity supply industry, we discuss the concepts from microeconomics that are essential for the understanding of electricity markets. We then move on to the analysis of the operation of power systems in a competitive environment. To keep matters simple, we begin by ignoring the transmission network and we consider the operation of pure energy markets. We then discuss power system security and the effects that networks have on electricity prices. Finally, in the last two chapters, we address the issue of investments in power generation and transmission equipment in a competitive environment.

The typical reader we had in mind while writing this book was a first-year graduate student or a final-year undergraduate student specializing in power engineering. We have assumed that these students know the physical structure of power systems, understand the purpose and principles of a power flow calculation and are familiar with basic optimization theory. We believe that this book will also be valuable to engineers who are working on deregulation or competition issues and who want to acquire a broader perspective on these questions. Finally, this book might also be useful to economists and other professionals who want to understand the engineering perspective on these multidisciplinary issues.

Except when a specific source is cited, we have made no attempt to use or produce realistic numbers in the problems and examples. We have used $ as a unit for money because it is probably the best-known symbol for a currency. We could have used €, £ or ¥ instead without any change in meaning.

Some of our examples refer to the fictitious countries of Syldavia and Borduria, which are the product of the fertile imagination of the Belgian cartoonist Hergé, creator of the character Tintin.

This book stems from our research and teaching activities in power system economics at UMIST. We are grateful to our colleagues Ron Allan and Nick Jenkins for fostering an environment in which this work was able to flourish. We also thank Fiona Woolf for fascinating interdisciplinary discussions on transmission expansion.

A few of our students spent considerable time proofreading drafts of this book and checking answers to the problems. In particular, we thank Tan Yun Tiam, Miguel Ortega Vazquez, Su Chua Liang, Mmeli Fipaza, Irene Charalambous, Li Zhang, Jaime Maldonado Moniet, Danny Pudjianto and Joseph Mutale. Any remaining errors are our sole responsibility.

1 Introduction

1.1 Why Competition?

For most of the twentieth century, when consumers wanted to buy electrical energy, they had no choice. They had to buy it from the utility that held the monopoly for the supply of electricity in the area where these consumers were located. Some of these utilities were vertically integrated, which means that they generated the electrical energy, transmitted it from the power plants to the load centers and distributed it to individual consumers. In other cases, the utility from which consumers purchased electricity was responsible only for its sale and distribution in a local area. This distribution utility in turn had to purchase electrical energy from a generation and transmission utility that had a monopoly over a wider geographical area. In some parts of the world, these utilities were regulated private companies, while in others they were public companies or government agencies. Irrespective of ownership and the level of vertical integration, geographical monopolies were the norm.

Electric utilities operating under this model made truly remarkable contributions to economic activity and quality of life. Most people living in the industrialized world have access to an electricity distribution network. For several decades, the amount of energy delivered by these networks doubled about every eight years. At the same time, advances in engineering improved the reliability of the electricity supply to the point that in many parts of the world the average consumer is deprived of electricity for less than two minutes per year. These achievements were made possible by ceaseless technological advances. Among these, let us mention only the development and erection of transmission lines operating at over 1 000 000 V and spanning thousands of kilometers, the construction of power plants capable of generating more than 1000 MW and the online control of the networks connecting these plants to the consumers through these lines. Some readers will undoubtedly feel that on the basis of this record, it may have been premature to write the first paragraph of this book in the past tense.

In the 1980s, some economists started arguing that this model had run its course. They said that the monopoly status of the electric utilities removed the incentive to operate efficiently and encouraged unnecessary investments. They also argued that the cost of the mistakes that private utilities made should not be passed on to the consumers. Public utilities, on the other hand, were often too closely linked to the government. Politics could then interfere with good economics. For example, some

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

public utilities were treated as cash cows, and others were prevented from setting rates at a level that reflected costs or were deprived of the capital that they needed for essential investments.

These economists suggested that prices would be lower and that the economy as a whole would benefit if the supply of electricity became the object of market discipline rather than monopoly regulation or government policy. This proposal was made in the context of a general deregulation of western economies that had started in the late seventies. Before attention turned toward electricity, this movement had already affected airlines, transportation and the supply of gas. In all these sectors, regulated market or monopolies had been deemed the most efficient mean of delivering the "products" to the consumers. It was felt that their special characteristics made them unsuitable for trading on free markets. Advocates of deregulation argued that the special characteristics of these products were not insurmountable obstacles and that they could and should be treated like all other commodities. If companies were allowed to compete freely for the provision of electricity, the efficiency gains arising from this competition would ultimately benefit the consumers. In addition, competing companies would probably choose different technologies. It was therefore less likely that the consumers would be saddled with the consequences of unwise investments.

If electricity truly were a simple commodity[1], kilowatt-hours could be stacked on a shelf – like kilograms of flour or television sets – ready to be used as soon as the consumer turns on the light or starts the industrial process. Despite recent technological advances in electricity storage and microgeneration, this concept is not yet technically or commercially feasible. The reliable and continuous delivery of significant amounts of electrical energy still requires large generating plants connected to the consumer through transmission and distribution networks.

In this book, we will explore how the production and trading of electrical energy can be separated conceptually from the operation of this power system. The kilowatt-hours can then be treated as a commodity and traded on a deregulated market.

1.2 Dramatis Personae

Before we delve into the analysis of electricity markets, it is useful to introduce the types of companies and organizations that play a role in these markets. In the following chapters, we will, of course, discuss in much more detail the function and motivation of each of these participants. Since markets have evolved at different rates and in somewhat different directions in each country or region, not all these entities will be found in each market. In some cases, one company or organization may perform more than one of the functions described below.

Vertically integrated utilities own generating plants as well as a transmission and distribution network. In a traditional regulated environment, such a company has a monopoly for the supply of electricity over a given geographical area. Following the liberalization of the electricity market, its generation and network activities are likely to be separated.

[1] This book would also not have been written.

Generating companies (gencos) produce and sell electrical energy. They may also sell services such as regulation, voltage control and reserve that the system operator needs to maintain the quality and security of the electricity supply. A generating company can own a single plant or a portfolio of plants of different technologies. Generating companies that coexist with vertically integrated utilities are sometimes called *independent power producers* (IPP).

Distribution companies (discos) own and operate distribution networks. In a traditional environment, they have a monopoly for the sale of electrical energy to all consumers connected to their network. In a fully deregulated environment, the sale of energy to consumers is decoupled from the operation, maintenance and development of the distribution network. Retailers then compete to perform this energy sale activity. One of these retailers may be a subsidiary of the local distribution company.

Retailers buy electrical energy on the wholesale market and resell it to consumers who do not wish, or are not allowed, to participate in this wholesale market. Retailers do not have to own any power generation, transmission or distribution assets. Some retailers are subsidiaries of generation or distribution companies. All the customers of a retailer do not have to be connected to the network of the same distribution company.

A market operator (MO) typically runs a computer system that matches the bids and offers that buyers and sellers of electrical energy have submitted. It also takes care of the settlement of the accepted bids and offers. This means that it forwards payments from buyers to sellers following delivery of the energy. The independent system operator (ISO) is usually responsible for running the market of last resort, that is, the market in which load and generation are balanced in real time. Markets that close some time ahead of real time are typically run by independent for-profit market operators.

The independent system operator (ISO) has the primary responsibility of maintaining the security of the power system. It is called independent because in a competitive environment, the system must be operated in a manner that does not favor or penalize one market participant over another. An ISO would normally own only the computing and communications assets required to monitor and control the power system. An ISO usually combines its system operation responsibility with the role of the operator of the market of last resort.

Transmission companies (transco) own transmission assets such as lines, cables, transformers and reactive compensation devices. They operate this equipment according to the instructions of the independent system operator. Transmission companies are sometimes subsidiaries of companies that also own generating plants. An *independent transmission company* (ITC) is a transmission company that does not own generating plants and also acts as an independent system operator.

The regulator is the governmental body responsible for ensuring the fair and efficient operation of the electricity sector. It determines or approves the rules of the electricity market and investigates suspected cases of abuse of market power. The regulator also sets the prices for the products and services that are provided by monopolies.

Small consumers buy electrical energy from a retailer and lease a connection to the power system from their local distribution company. Their participation in the electricity market usually amounts to no more than choosing one retailer among others when they have this option.

Large consumers, on the other hand, will often take an active role in electricity markets by buying their electrical energy directly through the market. Some of them may offer their ability to control their load as a resource that the ISO can use to control the system. The largest consumers are sometimes connected directly to the transmission system.

1.3 Models of Competition

Hunt and Shuttleworth (1996) propose four models to chart the evolution of the electricity supply industry from a regulated monopoly to full competition.

1.3.1 Model 1: Monopoly

The first model, which is shown in Figure 1.1, corresponds to the traditional monopoly utility. Submodel (a) corresponds to the case where the utility integrates the generation, transmission and distribution of electricity. In submodel (b), generation and transmission are handled by one utility, which sells the energy to local monopoly distribution companies. This model does not preclude bilateral energy trades between utilities operating in different geographical areas. As illustrated in Figure 1.1, these trades take place at the wholesale level.

1.3.2 Model 2: Purchasing agency

Figure 1.2(a) shows a possible first step toward the introduction of competition in the electricity supply industry. The integrated utility no longer owns all the generation

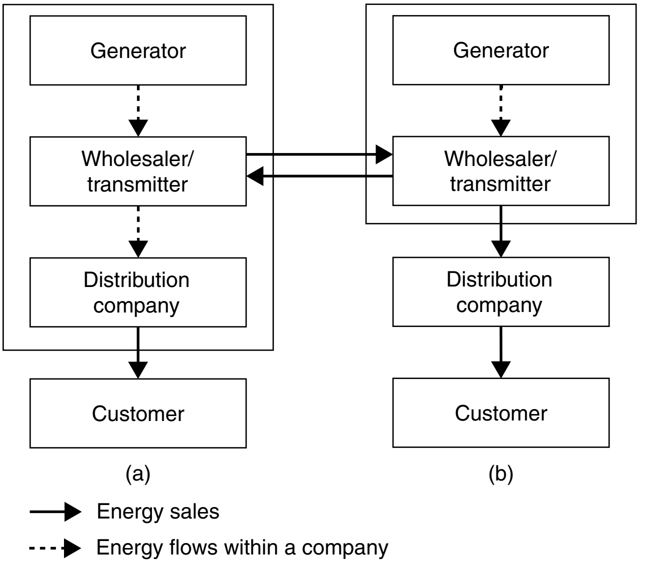


Figure 1.1 Monopoly model of electricity market based on (Hunt and Shuttleworth, 1996). In submodel (a), the utility is completely vertically integrated, while in submodel (b), the distribution is handled by one or more separate companies

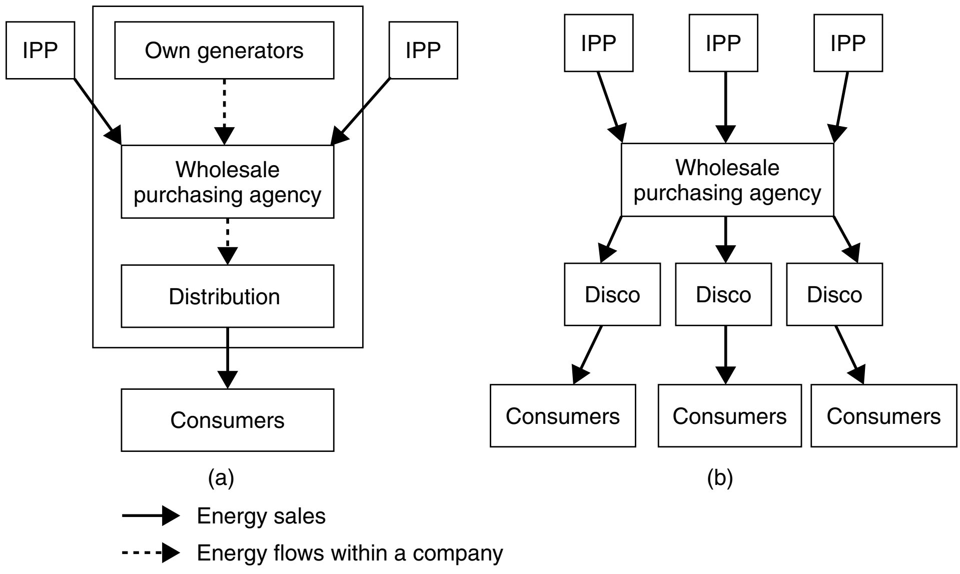


Figure 1.2 Purchasing agency model of electricity market based on (Hunt and Shuttleworth, 1996). (a) integrated version; (b) disaggregated version

capacity. Independent power producers (IPP) are connected to the network and sell their output to the utility that acts as a purchasing agent. Figure 1.2(b) shows a further evolution of this model where the utility no longer owns any generation capacity and purchases all its energy from the IPPs. The distribution and retail activities are also disaggregated. Discos then purchase the energy consumed by their customers from the wholesale purchasing agency. The rates set by the purchasing agency must be regulated because it has monopoly power over the discos and monopsony power toward the IPPs. This model therefore does not discover a cost-reflective price in the same way that a free market does (see Chapter 2). However, it has the advantage of introducing some competition between generators without the expense of setting up a competitive market as in the more complex models that we describe next.

1.3.3 Model 3: Wholesale competition

In this model, which is shown in Figure 1.3, no central organization is responsible for the provision of electrical energy. Instead, discos purchase the electrical energy consumed by their customers directly from generating companies. These transactions take place in a wholesale electricity market. The largest consumers are often allowed to purchase electrical energy directly on the wholesale market. As we will see in Chapter 3, this wholesale market can take the form of a pool or of bilateral transactions. At the wholesale level, the only functions that remain centralized are the operation of the spot market, and the operation of the transmission network. At the retail level, the system remains centralized because each disco not only operates the distribution network in its area but also purchases electrical energy on behalf of the consumers located in its service territory.

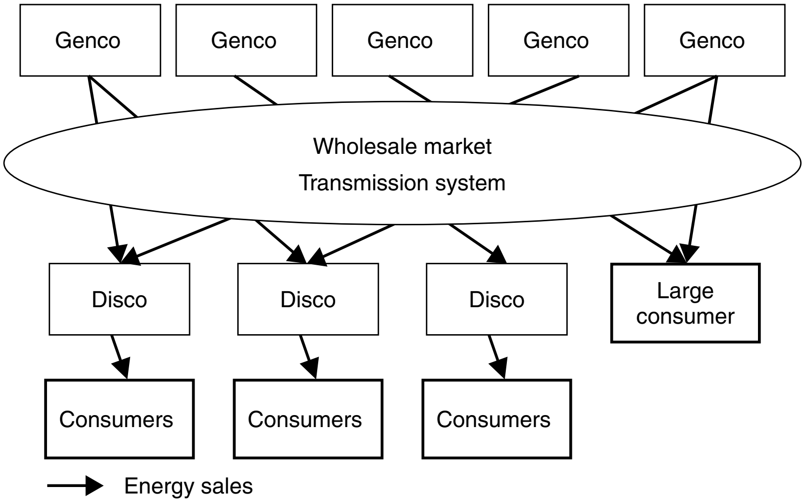


Figure 1.3 Wholesale competition model of electricity market based on (Hunt and Shuttleworth, 1996)

This model creates considerably more competition for the generating companies because the wholesale price is determined by the interplay of supply and demand. On the other hand, the retail price of electrical energy must remain regulated because small consumers cannot choose a competing supplier if they feel that the price is too high. This leaves the distribution companies exposed to sudden large increases in the wholesale price of energy.

1.3.4 Model 4: Retail competition

Figure 1.4 illustrates the ultimate form of competitive electricity market in which all consumers can choose their supplier. Because of the transaction costs, only the largest consumers choose to purchase energy directly on the wholesale market. Most small and medium consumers purchase it from retailers, who in turn buy it in the wholesale market. In this model, the "wires" activities of the distribution companies are normally separated from their retail activities because they no longer have a local monopoly for the supply of electrical energy in the area covered by their network. In this model, The only remaining monopoly functions are thus the provision and operation of the transmission and distribution networks.

Once sufficiently competitive markets have been established, the retail price no longer has to be regulated because small consumers can change retailer when they are offered a better price. As we will see in Chapter 2, from an economics perspective this model is the most satisfactory because energy prices are set through market interactions. Implementing this model, however, requires considerable amounts of metering, communication and data processing.

The cost of the transmission and distribution networks is still charged to all their users. This is done on a regulated basis because these networks remain monopolies.

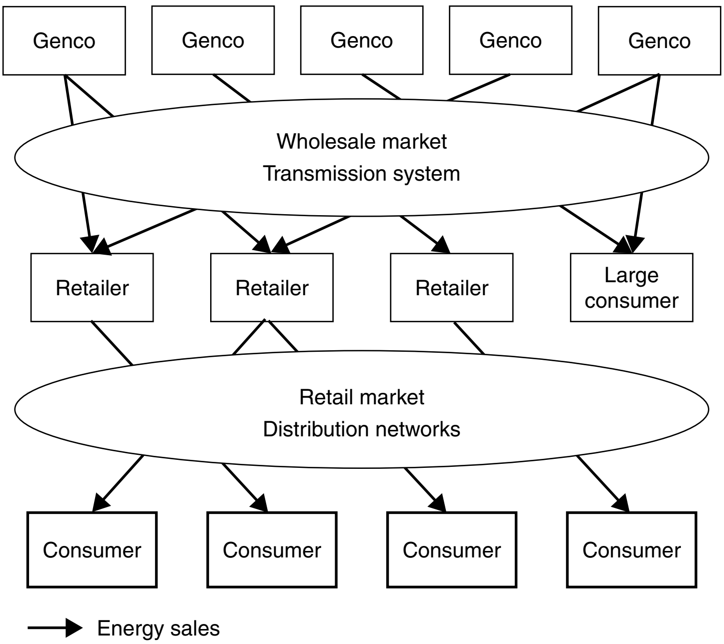


Figure 1.4 Retail competition model of electricity market based on (Hunt and Shuttleworth, 1996)

1.3.5 Competition and privatization

In many countries, the introduction of competition in the supply of electricity has been accompanied by the privatization of some or all components of the industry. Privatization is the process by which publicly owned utilities are sold by the government to private investors. These utilities then become private, for-profit companies. Privatization is not, however, a prerequisite for the introduction of competition. None of the four models of competition described above implies a certain form of ownership. Public utilities can, and in many instances do, compete with private companies.

1.4 Open Questions

In the monopoly utility model, all technical decisions regarding the operation and the development of the power system are taken within a single organization. In the short term, this means that, at least in theory, the operation of all the components of the system can be coordinated to achieve least cost operation. For example, the maintenance of the transmission system can be scheduled jointly with the maintenance of the generation units to minimize the effects of congestion. Similarly, the long-term development of the system can be planned to ensure that the transmission capacity and topology match the generation capacity and location.

Introducing competition implies renouncing centralized control and coordinated planning. A single integrated utility is replaced by a constellation of independent companies. Each of these decides independently what it will do to maximize its private objectives. When the idea of competitive electricity markets was first mooted, it was rejected by many on the grounds that such a disaggregated system could not keep the lights on. There is now ample evidence to demonstrate that separating the operation of generation from that of the transmission system does not necessarily reduce the reliability of the overall system.

What is considerably more difficult to prove is that a disaggregated, competitive system operates more efficiently than a centralized one. While it is clear that the profit motive encourages generating companies to take better care of their plants, it remains to be proven that this improvement in availability (and possibly efficiency) is sufficient to compensate for the loss of coordination between the plants.

In terms of long-term development, the argument in favor of competition is that central planners always get their forecast wrong. In particular, monopoly utilities have a tendency to overestimate the amount of generation capacity that will be needed. Their captive consumers are then obliged to pay for unnecessary investments. With the introduction of competition, it is hoped that the sum of the independent investment decisions of several profit-seeking companies will match the actual evolution of the demand more closely than the recommendations of a single planning department. In addition, underutilized investments by a company operating in a free market represent a risk for its owners and not its customers. Experience from around the world suggests that investors are willing to accept this risk. However, it remains to be seen if the growth in generation capacity smoothly matches the increase in demand or goes through "boom-and-bust" cycles.

Vertically integrated utilities can plan the development of their transmission network to suit the construction of new generating plants. In a competitive environment, the transmission company does not know years in advance where and when generating companies will build new plants. This uncertainty makes the transmission planning process much more difficult. Conversely, generating companies are not guaranteed that transmission capacity will remain available for the output of their plants. Other companies may indeed build new plants in the vicinity and compete for the available transmission capacity.

The transmission and distribution networks have so far been treated as natural monopolies. Having two separate and competing sets of transmission lines or distribution feeders clearly does not make sense. From both the economic and the reliability points of view, all lines, feeders and other components should be connected to the same system. On the other hand, some economists and some entrepreneurs have begun to argue that not all these components must be owned by the same company. They believe that new investments could be driven by investors who expand a network to satisfy specific needs for power transmission or distribution that they have identified. Taken individually, such opportunities could be lucrative for the investors. However, they must take place within a framework that maximizes the overall benefits derived by all users of the network. Such a framework remains to be developed.

1.5 Further Reading

Hunt S, Shuttleworth G, *Competition and Choice in Electricity*, Wiley, Chichester, 1996.

1.6 Problems

1.1 Using the classification proposed by Hunt and Shuttleworth, determine the level of competition that exists in your region or country or in another area for which you have access to sufficient information. Discuss any difference that you observe between the basic model and the electricity market implementation in this area.

1.2 Identify the companies that participate in the electricity market in the area that you chose for Problem 1.1. Map the basic functions defined in this chapter with these companies and discuss any difference that you observe. Identify clearly the companies that enjoy a monopoly status in some or all their activities.

1.3 Identify the regulatory agencies that oversee the electricity supply industry in the area that you chose for Problem 1.1.

1.4 Identify the organizations that fulfill the functions of market operator and system operator in the area that you chose for Problem 1.1.

1.5 The reasons invoked for implementing a competitive electricity market or a certain model of electricity market depend on local circumstances. Identify and discuss the reasons that were invoked in the region that you chose for Problem 1.1.

2
Basic Concepts from Economics

2.1 Introduction

In this chapter, we introduce the concepts from the theory of microeconomics that are needed to understand electricity markets. We also take this opportunity to explain some of the economics terminology that is becoming more common in the engineering literature. This chapter has a limited and utilitarian scope and does not pretend to provide a complete or rigorous course in microeconomics. The reader who feels the need or the inclination to study this subject in more depth is encouraged to consult one of the specialized textbooks listed at the end of this chapter.

As we will see in the following chapters, electricity is not a simple commodity and electricity markets are more complex than markets for other products. To avoid unnecessary complications, we will introduce the basic concepts of microeconomics using examples that have nothing to do with electricity.

2.2 Fundamentals of Markets

Markets are a very old invention that can be found in most civilizations. Over the years, they have evolved from being simply a location where a few people would occasionally gather to barter goods to virtual environments where information circulates electronically and deals are made with a click of a mouse. Despite these technological changes, the fundamental principle has not changed: a market is a place where buyers and sellers meet to see if deals can be made.

To explain how markets function, we will first develop a model that describes the behavior of the consumers. Then, we will develop a model explaining the activities of the producers. By combining these two models, we will be able to show under what conditions deals can be struck.

2.2.1 Modeling the consumers

2.2.1.1 Individual demand

Let us begin with a simple example: suppose that you work close enough to a farmer's market to be able to walk there during your midmorning break. While the farmers

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

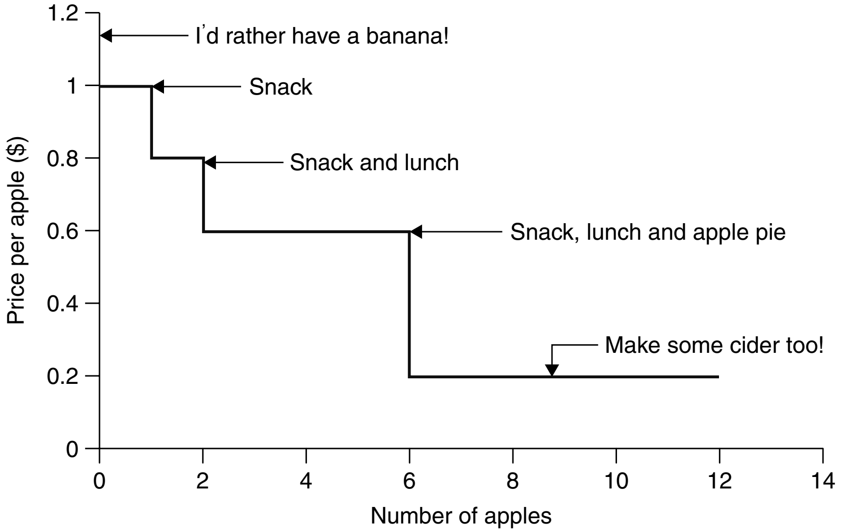


Figure 2.1 Typical relation between the price of apples and the demand of a particular customer

sell different types of fruit and vegetables on this market, today you are looking at the apples. The number of apples you purchase depends on their current price. There is certainly a price above which you will decide to forego your daily snack or buy another type of fruit instead. If the price is below that threshold but still quite high, you will probably buy only one apple to eat on your way back to work. If the price is lower still, you may buy one for now and another for lunch. At even lower prices, you may decide to purchase apples to make a pie for dinner. Finally, if the price is lower than you have ever seen it before, this may be the opportunity to experiment with the cider-making kit that your brother-in-law gave you for your last birthday. Figure 2.1 summarizes how your demand for apples varies with price. Traditionally (and, at first, counterintuitively), the price is plotted on the vertical axis in such graphs. This curve shows what the price should be for a consumer to purchase a certain amount. It is drawn assuming that the consumer's income and the price of other commodities remain constant.

You might argue that your decision to buy apples would also be influenced by the quality of those that are for sale. This is an important point and we will take care of it by assuming that all the non-price characteristics of the commodity considered (type, size and quality) are precisely defined.

2.2.1.2 Surplus

Let us suppose that when you get to the market, the price is \$0.40 per apple. At that price, as shown by Figure 2.2, you decide to buy six apples. We can calculate the *gross consumer's surplus* that you, as a consumer, achieve by buying these apples. This represents the total value that you attach to these apples. The calculation goes as follows:

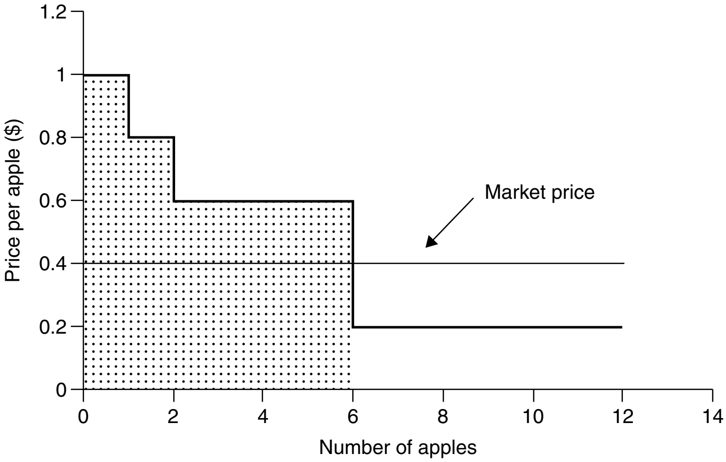


Figure 2.2 Gross surplus of purchasing apples

Value of the first apple:	$1 \times \$1.00 =$	\$1.00
Value of the second apple:	$1 \times \$0.80 =$	\$0.80
Value of the next four apples:	$4 \times \$0.60 =$	\$2.40
Gross surplus:		\$4.20

As Figure 2.2 shows, the gross consumer's surplus is equal to the area under the curve shown in Figure 2.1. However, you have had to pay $6 \times \$0.40 = \$2.40$ to purchase these apples and this represents money that you no longer have for other purposes. We define the *net consumer's surplus* (or simply *consumer's surplus)* as the difference between the gross consumer's surplus and the expense of purchasing the goods. Graphically, as illustrated in Figure 2.3, the net consumer's surplus is equal to the area between the curve and the horizontal line at the market price. The net consumer's surplus represents the "extra value" that you get from being able to buy all the apples at the same market price, even though the value you attach to them (except the last one) is higher than the market price.

2.2.1.3 Demand and inverse demand functions

It is very unlikely that all the consumers going to the market have exactly the same appetite for apples as you do. Some of them would pay much more for the same number of apples, while others buy apples only when they are cheap. If we aggregate the demand characteristics of a sufficiently large number of consumers, the discontinuities introduced by the individual decisions are smoothed away, leading to a curve like the one shown in Figure 2.4. This curve represents the *inverse demand function* of the

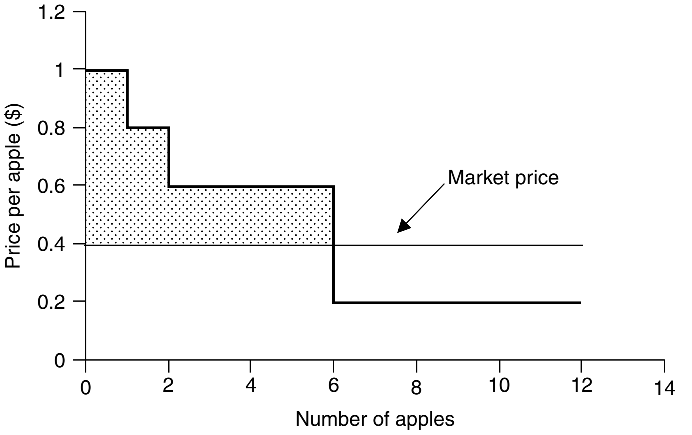


Figure 2.3 Net consumer's surplus resulting from the purchase of apples

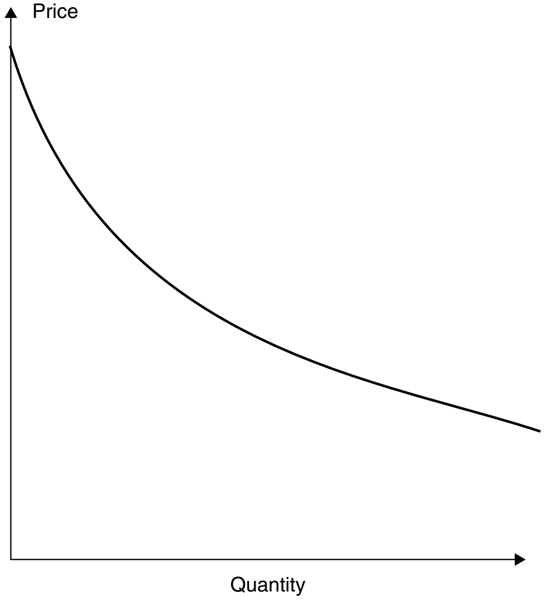


Figure 2.4 Typical relation between the price of a commodity and the demand for this commodity by a group of consumers. This curve is called the inverse demand function or the demand function depending on the perspective adopted

customers taken as a whole. If q represents the quantity consumed and π the price of the commodity, we can write

$$\pi = D^{-1}(q) \tag{2.1}$$

If we look at the same curve from the other direction, we have the *demand function* for this commodity:

$$q = D(\pi) \tag{2.2}$$

For most, if not all, practical commodities, the demand function is downward sloping, that is, the amount consumed decreases as the price increases. The inverse demand function has an important economic interpretation. For a given consumption level, it measures how much money the consumers would be willing to pay to have a small additional amount of the good considered. Turning this around, it also tells how much money these same consumers would want to receive in compensation for a reduced consumption. Not spending this amount of money on this commodity would allow them to purchase more of another commodity or save it for purchasing something at a later date. In other words, the demand curve gives the *marginal value* that consumers attach to the commodity. The typical downward-sloping shape of the curve indicates that consumers are usually willing to pay more for additional quantities of a commodity when they have only a small amount of this commodity. Their marginal willingness to pay for this commodity decreases as their consumption increases.

The concepts of gross and net consumer's surplus that we defined above for a single consumer can be extended to the gross and net surpluses of a group of consumers. As Figure 2.5 illustrates, the gross surplus is represented graphically by the area below the inverse demand function up to the quantity that the consumers purchase at the current market price. The net surplus corresponds to the area between the inverse demand function and the horizontal line at the market price.

The concept of net surplus is much more important than the calculation of an absolute value for this quantity. Calculating the absolute value of the net surplus is quite difficult because the inverse demand function is not known accurately.

Examining how this net surplus varies with the market price is much more interesting. Figure 2.6 illustrates the change in net surplus when the market price increases. If the market price is π_1, the consumers purchase a quantity q_1 and, the net surplus is equal to the shaded area. If the price increases to π_2, the consumption level decreases to q_2, and the consumers' net surplus is reduced to the roughly triangular area labeled A. Two effects contribute to this reduction in net surplus. First, because the price is higher, consumption decreases from q_1 to q_2. This loss of net surplus is equal to the area labeled C. Second, because consumers have to pay a higher price for the

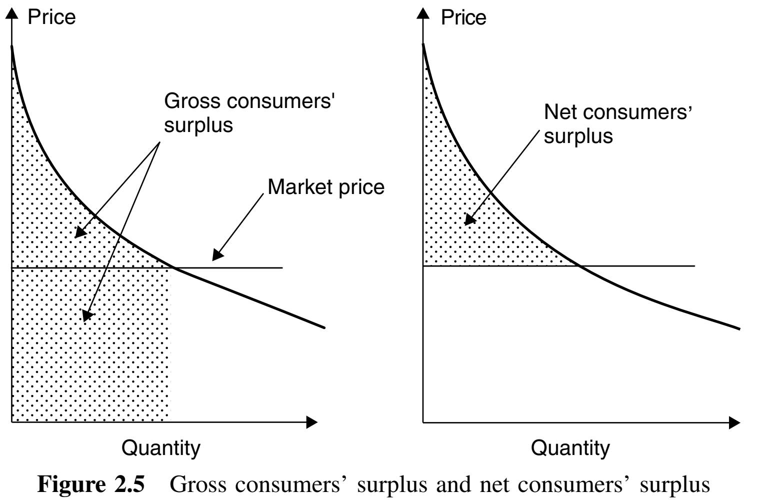


Figure 2.5 Gross consumers' surplus and net consumers' surplus

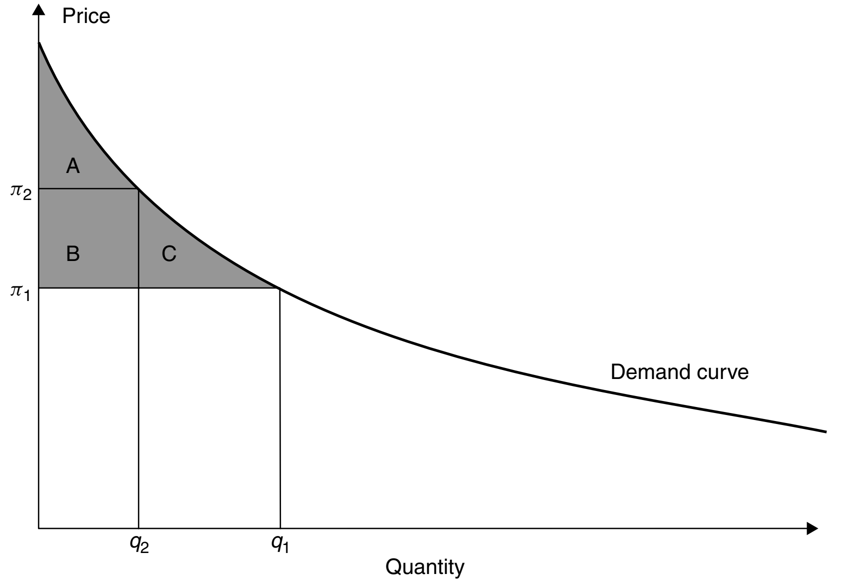


Figure 2.6 Change in the net consumers' surplus resulting from an increase in the market price

quantity q_2 that they still purchase, they loose an additional amount of surplus represented by the area labeled B.

2.2.1.4 Elasticity of demand

Increasing the price of a commodity even by a small amount will clearly decrease demand. But by how much? To answer this question, we could use the derivative $\frac{dq}{d\pi}$ of the demand curve. Using this slope directly presents the problem that the numerical value depends on the units that we use to measure the quantity and the price. Comparing the demand's response to price changes for various commodities would be impossible. To get around this difficulty, we define the *price elasticity of demand* as the ratio of the relative change in demand to the relative change in price:

$$\varepsilon = \frac{\frac{dq}{q}}{\frac{d\pi}{\pi}} = \frac{\pi}{q}\frac{dq}{d\pi} \tag{2.3}$$

The demand for a commodity is said to be *elastic* if a given percentage change in price produces a larger percentage change in demand. On the other hand, if the relative change in demand is smaller than the relative change in price, the demand is said to be *inelastic*. Finally, if the elasticity is equal to -1, the demand is *unit elastic*.

The elasticity of the demand for a commodity depends in large part on the availability of substitutes. For example, the elasticity of the demand for coffee would be much smaller if consumers did not have the option to drink tea. When discussing elasticities and substitutes, one has to be clear about the timescale for substitutions. Suppose that electric heating is widespread in a region. In the short run, the price elasticity of the demand for electricity is very low because consumers do not have a choice if they want to stay warm. In the long run, however, they can install gas-fired heating and the price elasticity of the demand for electricity will be much higher.

The concept of substitute products can be quantified by defining the *cross-elasticity* between the demand for commodity i and the price of commodity j:

$$\varepsilon_{ij} = \frac{\dfrac{\mathrm{d}q_i}{q_i}}{\dfrac{\mathrm{d}\pi_j}{\pi_j}} = \frac{\pi_j}{q_i}\frac{\mathrm{d}q_i}{\mathrm{d}\pi_j} \tag{2.4}$$

While the elasticity of a commodity to its own price (its *self-elasticity*) is always negative, cross-elasticities between substitute products are positive because an increase in the price of one will spur the demand for the other. If two commodities are *complements*, a change in the demand for one will be accompanied by a similar change in the demand for the other. Electricity and electric heaters are clearly complements. The cross-elasticities of complementary commodities are negative.

2.2.2 Modeling the producers

2.2.2.1 Opportunity cost

Our model of the consumers' behavior is based on the assumption that these consumers can choose how much of a commodity they purchase. We also argued that the consumption level is such that the marginal benefit that consumers get from this commodity is equal to the price that they have to pay to obtain it. A similar argument can be used to develop our model of the producers.

Let us consider one of the apple growers who brings her products to the market that we visited earlier. There is a price below which she will decide that selling apples is not worthwhile. There are several reasons why she could conclude that this revenue is insufficient. First, it might be less than the cost of producing the apples. Second, it might be less than the revenue she could get by using these apples for some other purposes, such as selling them to a cider-making factory. Finally, she could decide that she would rather devote the resources needed to produce apples (money, land, machinery and her own time) into some other activity, such as growing pears or opening a bed-and-breakfast. One can summarize these possibilities by saying that the revenue from the sale of apples is less than the *opportunity cost* associated with the production of these apples.

2.2.2.2 Supply and inverse supply functions

On the other hand, if the market price for apples is higher, our producer may decide that it is worthwhile to increase the amount of apples that she brings to the market.

Other producers have different opportunity costs and will therefore decide to adjust the amount they supply at different price thresholds. If we aggregate the amounts supplied by a sufficiently large number of producers, we get a smooth, upward-sloping curve such as the one shown in Figure 2.7. This curve represents the *inverse supply function* for this commodity:

$$\pi = S^{-1}(q) \tag{2.5}$$

This function indicates the value that the market price should take to make it worthwhile for the aggregated producers to supply a certain quantity of the commodity to the market. We can, of course, look at the same curve from the other direction and define the *supply function*, which gives us the quantity supplied as a function of the market price:

$$q = S(\pi) \tag{2.6}$$

As depicted in Figure 2.8, goods produced by different producers (or by the same producer but using different means of production) are located on different parts of the supply curve. The marginal producer is the producer whose opportunity cost is equal to the market price. If this market price decreases even by a small amount, this producer would decide that it is not worthwhile to continue production. Extra-marginal production refers to production that could become worthwhile if the market price were to increase. On the other hand, the opportunity cost of the infra-marginal producers is below the market price. These producers are thus able to sell at a price that is higher than the lowest price at which they would find it worthwhile to produce.

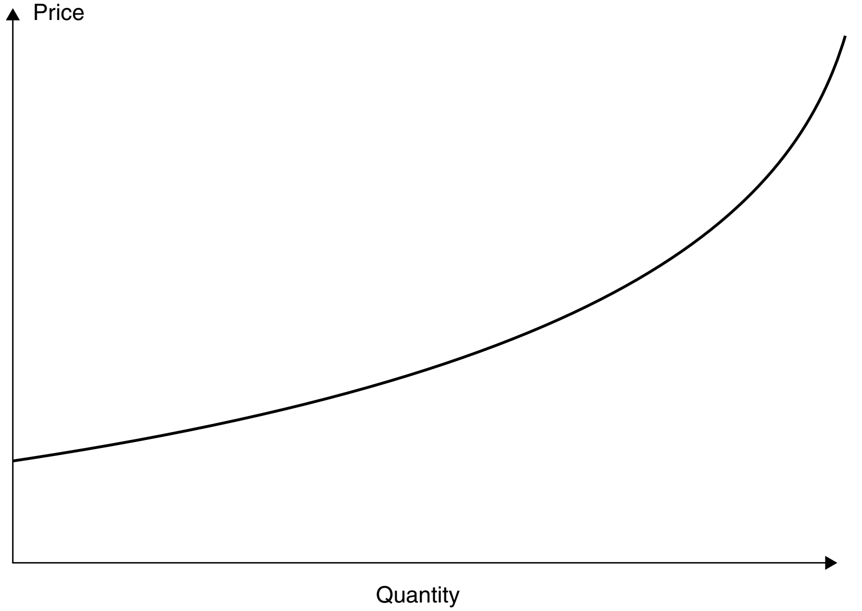


Figure 2.7 Typical supply curve

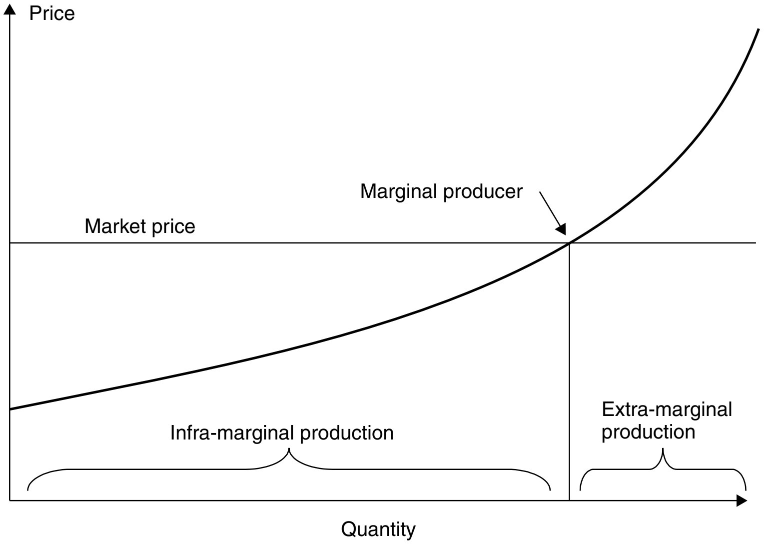


Figure 2.8 Marginal production is such that its opportunity cost is equal to the market price

2.2.2.3 Producers' revenue

Since the entire supply of the commodity is traded at the same price, the *producers' revenue* is equal to the product of the traded quantity q_1 and the market price π_1. This quantity is thus equal to the shaded area in Figure 2.9. The *producers' net surplus* or *producers' profit* arises from the fact that all the goods (except for the marginal production) are traded at a price that is higher than their opportunity cost. As Figure 2.10 shows, this net surplus or profit is equal to the area between the supply curve and the horizontal line at the market price. Producers with a low opportunity cost capture a proportionately larger share of the profit than those who have a higher opportunity cost. The marginal producer does not reap any profit.

Figure 2.11 shows that an increase in the market price from π_1 to π_2 affects the net producers' surplus in two ways. It increases the quantity that they supply to the market from q_1 to q_2 (area labeled C) and increases the revenue on all the quantity that they supplied to the market at the original price (area labeled B).

2.2.2.4 Elasticity of supply

An increase in the price of a commodity encourages suppliers to make available larger quantities of this commodity. The *price elasticity of supply* quantifies this relation. Its definition is similar to the definition of the price elasticity of the demand, but it involves the derivative of the supply curve rather than the derivative of the demand curve:

$$\varepsilon = \frac{\frac{dq}{q}}{\frac{d\pi}{\pi}} = \frac{\pi}{q}\frac{dq}{d\pi} \tag{2.7}$$

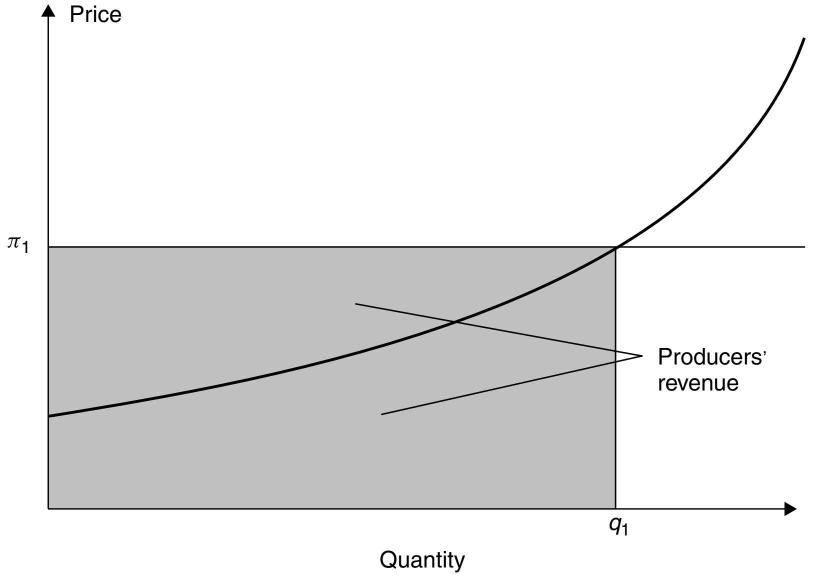


Figure 2.9 The producers' revenue is equal to the product of the market price π_1 and the traded quantity q_1

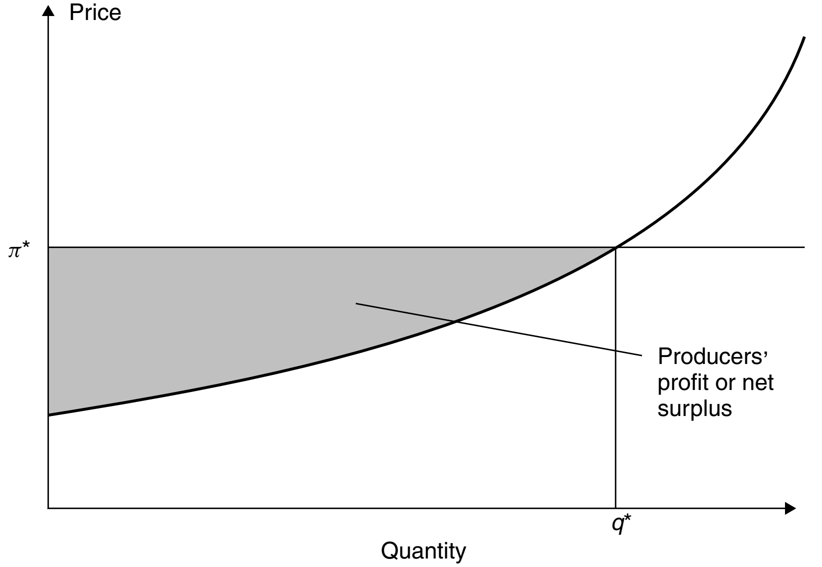


Figure 2.10 The producers' profit or net surplus is due to the ability of the producers to sell their commodity at a price higher than their opportunity cost

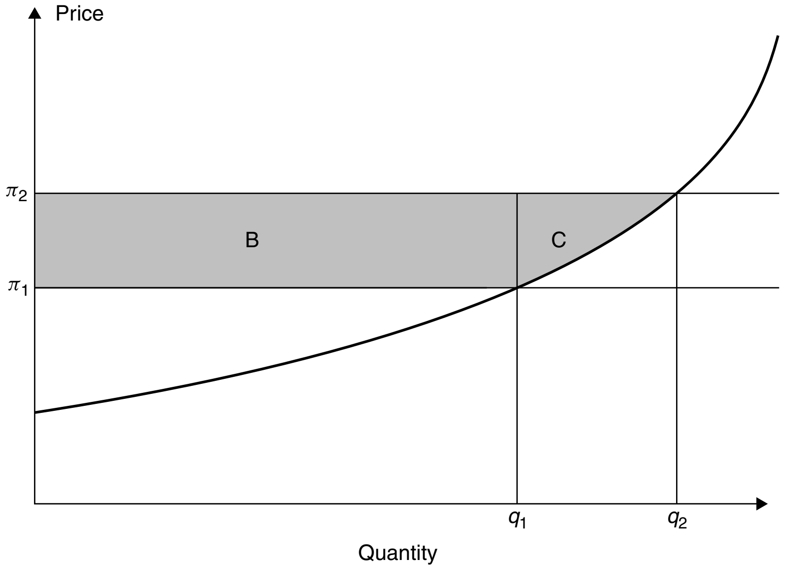


Figure 2.11 Change in the producers' profit or net surplus when the market price changes

The elasticity of supply is always positive. It will usually be higher in the long run than in the short run because suppliers have the opportunity to increase the means of production.

2.2.3 Market equilibrium

So far, we have considered producers and consumers separately. It is time to see how they interact in a market. In this section, we make the assumption that each supplier or consumer cannot affect the price by its actions. In other words, all market participants take the price as given. If this assumption is true, the market is said to be a *perfectly competitive market*. This assumption is usually not true for electricity markets. We will thus discuss in a later section how markets operate when some participants can influence the price through their actions.

In a competitive market, it is the combined action of all the consumers on one side and of all the suppliers on the other side that determines the price. The *equilibrium price* or *market clearing price* π^* is such that the quantity that the suppliers are willing to provide is equal to the quantity that the consumers wish to obtain. It is thus the solution of the following equation:

$$D(\pi^*) = S(\pi^*) \tag{2.8}$$

This equilibrium can also be defined in terms of the inverse demand function and the inverse supply function. The equilibrium quantity q^* is such that the price that the consumers are willing to pay for that quantity is equal to the price that producers must receive to supply that quantity:

$$D^{-1}(q^*) = S^{-1}(q^*) \tag{2.9}$$

Figure 2.12 illustrates these concepts. So far, we have shown that at the market equilibrium, the behaviors of the consumers and the suppliers are consistent. We have not yet shown, however, that this point represents a stable equilibrium. To demonstrate this, let us show that the market will inevitably settle at that point. Suppose, as shown in Figure 2.13, that the market price is $\pi_1 < \pi^*$, where the demand is greater than the supply. Some suppliers will inevitably realize that there are some unsatisfied customers to whom they could sell their goods at more than the going price. The traded quantity will increase and so will the price until the equilibrium conditions are reached. Similarly, if the market price is $\pi_2 > \pi^*$, the supply exceeds the demand and some suppliers are left with goods for which they cannot find buyers. To avoid being caught in this situation, they will reduce their production until the amount that producers are willing to sell is equal to the amount that consumers are willing to buy.

2.2.4 Pareto efficiency

When a system is completely under the control of a single organization, this organization normally attempts to optimize some measure of the benefit it derives from the system. On the other hand, when a system depends on the interactions of various organizations with diverging interests, conventional optimization is not applicable and must be replaced by *Pareto efficiency*. An economic situation is Pareto efficient if the benefit derived by any of the parties can be increased only by reducing the benefit enjoyed by one of the other parties.

The equilibrium situation in a competitive market is Pareto efficient in terms of both the quantity of goods exchanged and the allocation of these goods. Let us first consider the quantity exchanged with the help of Figure 2.14. Suppose that the quantity exchanged is q, which is less than the equilibrium quantity q^*. At that quantity, there

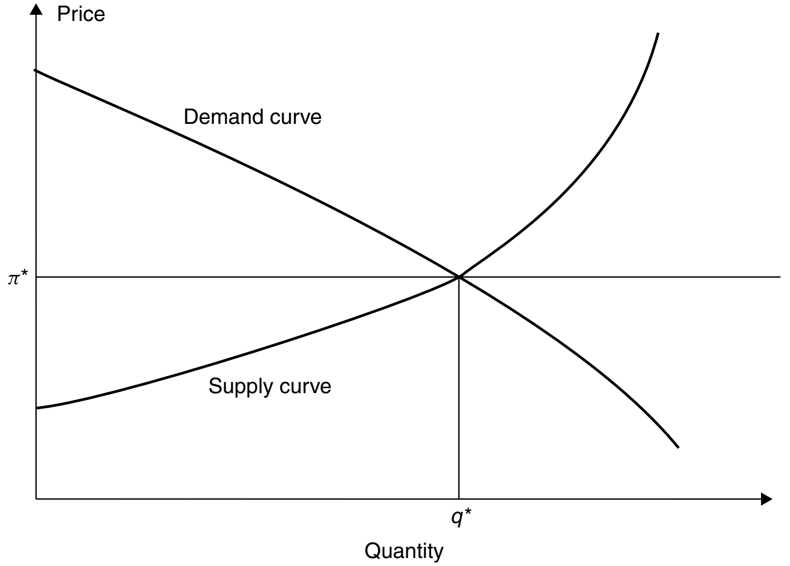


Figure 2.12 Market equilibrium

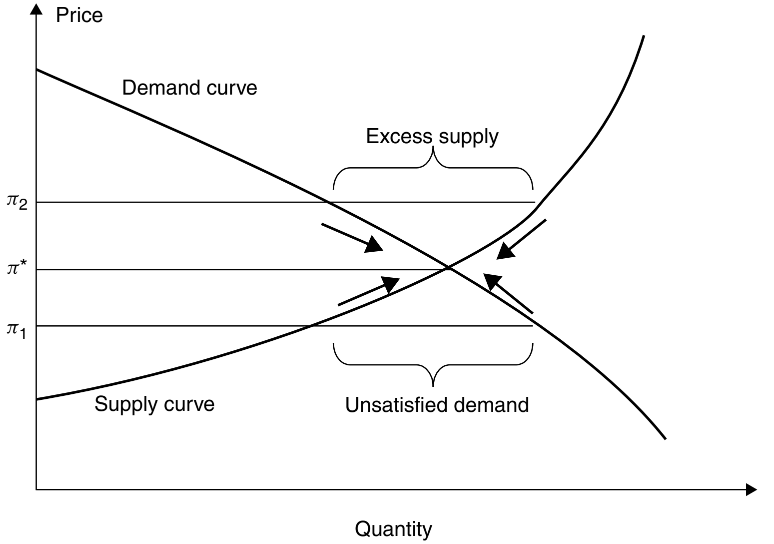


Figure 2.13 Stability of the market equilibrium

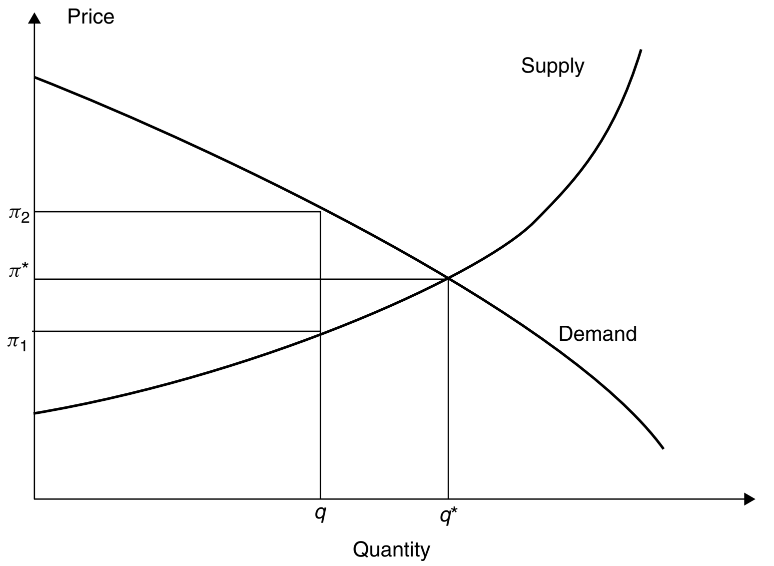


Figure 2.14 Pareto efficiency of the market equilibrium

is someone willing to sell extra units of the good considered at a price π_1, which is less than the price π_2 that someone else is willing to pay for that extra unit. If a trade can be arranged between these two parties at any price between π_1 and π_2, both parties will be better off. Thus, if the total amount traded is less than the equilibrium q^*, the situation is not Pareto efficient. Similarly, any amount in excess of the equilibrium value is not Pareto efficient because the price that someone would be willing to pay for an extra unit is lower than the price that it would take to get it supplied.

Let us now consider the efficiency of the allocation of the goods. In a competitive market, all units of a given commodity are traded at the same price and this price represents the marginal rate of substitution between this good and all other goods. Consumer A's willingness to pay this price means that he values the last unit he purchased of this good more than other goods. On the other hand, consumer B may decide that at that price, she would rather buy other goods. Suppose now that the goods are not allocated on the basis of the willingness to pay the market price but on some other basis. Consumer A may find himself in a situation in which he may be willing to pay $10 to buy an extra unit of the good in addition to those that he has been allocated. On the other hand, consumer B may have received an allocation such that she only values the last unit that she has at $8. Since these two consumers place different values on one unit of the same good, they would both be better off if they could trade this unit at any price between $8 and $10. It is thus only when goods are allocated on the basis of a single marginal rate of substitution, as happens in a competitive market, that Pareto efficiency is achieved.

2.2.5 Global welfare and deadweight loss

The sum of the net consumers' surplus and of the producers' profit is called the *global welfare*. It quantifies the overall benefit that arises from trading. We will now show that the global welfare is maximum when a competitive market is allowed to operate freely and the price settles at the intersection of the supply and demand curves. Under these conditions, Figure 2.15 shows that the consumers' surplus is equal to the sum of the areas labeled A, B and E and the producers' profit to the sum of the areas labeled C, D and F.

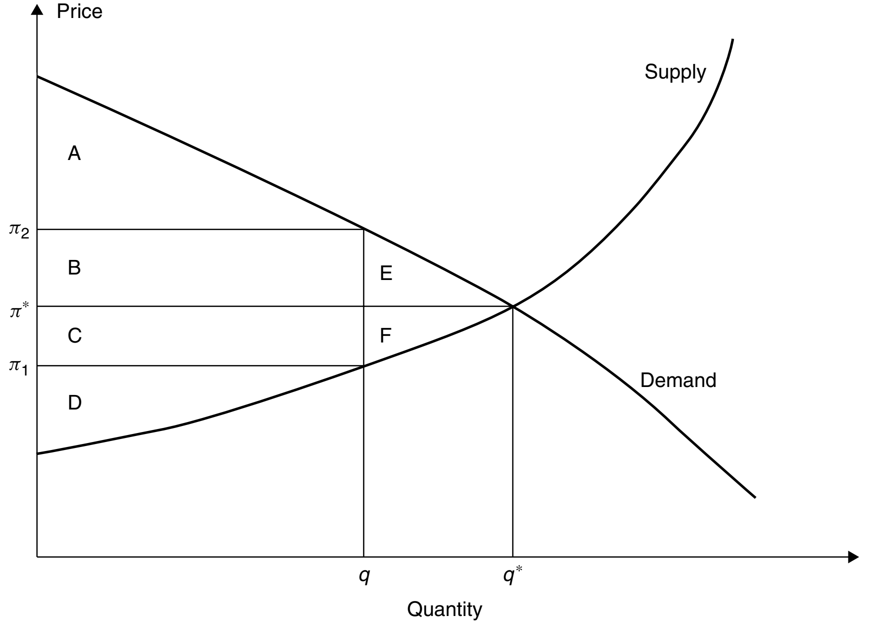


Figure 2.15 Global welfare and deadweight loss

External intervention sometimes prevents the price of a good from settling at the equilibrium value that would result from a free and competitive market. First, in an effort to help producers, the government could set a minimum price for a commodity. If this price is set at a value π_2 that is higher than the competitive market-clearing price π^*, this minimum price becomes the market price and consumers reduce their consumption from q^* to q. Under these conditions, the consumers' surplus shrinks to area A, while the producers' surplus is represented by the sum of the areas B, C and D.

Similarly, the government could enforce a maximum price for a good. If this price is set at a value π_1 that is lower than the competitive market clearing price π^*, producers will cut their output to q. In this case, the consumers enjoy a net surplus equal to the sum of areas A, B and C while the producers' surplus is only area D.

Finally, the government could decide to tax this commodity. If we assume that the tax is passed on to the consumers in its entirety, it creates a difference between the price paid by the consumers (say π_2) and the price received by the producers (say π_1). The government collects the difference $\pi_2 - \pi_1$ for each unit traded. Under these conditions, the demand again drops from q^* to q, the consumers' surplus contracts to area A and the producers' surplus to area D. The total amount collected by the government in taxes is equal to the sum of areas B and C.

External intervention redistributes the global welfare in favor of the producers, the consumers or the government, respectively. Unfortunately, all these interventions have the undesirable side effect of reducing the global welfare by an amount equal to the sum of the areas labeled E and F. This drop in global welfare is called the *deadweight loss* and is the result of the reduction in the amount traded caused by the price distortion. Note that for simplicity we have assumed the same drop in demand for all three forms of external intervention. Obviously, this does not have to be the case.

We will see in later chapters that, in some markets, the price of electrical energy is set through a centralized calculation and not through the direct interaction of producers and consumers. To maximize the benefits of trading, this centralized calculation should simulate the operation of a free market by maximizing the global welfare.

2.3 Concepts from the Theory of the Firm

Let us now take a more detailed look at the behavior of the firms that produce the goods that are traded on the market.

2.3.1 Inputs and outputs

For the sake of simplicity, we will consider a firm that produces a quantity y of a single good. In order to produce this output our firm needs some inputs, which are called *factors of production.* Factors of productions vary widely depending on the type of output produced by the firm. They can be classified into broad categories such as raw materials, labor, land, buildings or machines. We will assume that our firm needs only two factors of production. Its output is related to the inputs by a *production function* that reflects the technology used by the firm in the production of the good:

$$y = f(x_1, x_2) \tag{2.10}$$

For example, y might represent the amount of wheat produced by a farmer, with x_1 being the amount of fertilizer and x_2 the surface of land that this farmer uses to raise this wheat.

To get some insight into the shape of the production function, let us keep the second factor of production constant and progressively increase the first. At the beginning, the output y increases with x_1. However, for almost all goods and technologies, the rate of increase of y decreases as x_1 gets larger. This phenomenon is called the *law of diminishing marginal product.*

In our example, the yield of a fixed amount of land cultivated by our farmer will go up as he increases the amount of fertilizer. It is clear, however, that above a certain density, the effectiveness of fertilizer declines. Similarly, cultivating more land will increase the total amount of wheat produced. However, as the amount of land goes up, the rate of increase in output will inevitably decrease, as the fixed amount of fertilizer must be spread over a larger area.

2.3.2 Long run and short run

Some factors of production can be adjusted faster than others. For example, a horticulturist can increase her production of apples by increasing the amount of organic fertilizer she uses or by hiring more labor to harvest the fruit. The effect of these adjustments will be felt at the next harvest. She could also increase production by planting more trees. In this case the results will not materialize until these new trees have had time to mature, a process that obviously takes a few years.

There is, however, no specific deadline separating the short and long runs. Economists define the long run as being a period of time sufficiently long as to allow all factors of productions to be adjusted. On the other hand, in the short run, some of the factors of production are fixed. For example, if we assume that the second factor of production has a fixed value $\overline{x_2}$, the production function becomes a function of a single variable:

$$y = f(x_1, \overline{x_2}) \tag{2.11}$$

Figure 2.16 shows the shape of a typical production function.

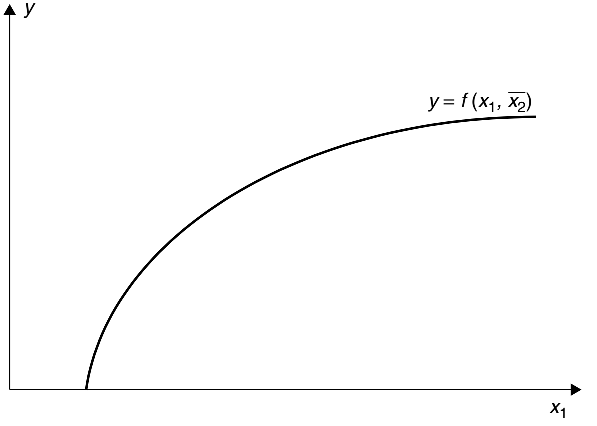


Figure 2.16 Typical short-run production function

In the short run, the output often depends on a single production factor. It is then convenient to define the input–output function, which is the inverse of the production function:

$$x_1 = g(y) \text{ for } x_2 = \overline{x_2} \tag{2.12}$$

The input–output function indicates how much of the variable production factor is required to produce a specified amount of goods. For example, the input–output curve of a thermal power plant shows how much fuel is required every hour to produce a given amount of power using this plant. We can then define the short-run cost function:

$$c_{SR}(y) = w_1 \cdot x_1 + w_2 \cdot \overline{x_2} = w_1 \cdot g(y) + w_2 \cdot \overline{x_2} \tag{2.13}$$

where w_1 and w_2 are the unit costs of the factors of production x_1 and x_2. Figure 2.17(a) illustrates a typical short-run cost function. The convexity of this function is due to the law of diminishing marginal products. Because of this convexity, the derivative of the cost function, which is called the *marginal cost function*, is a monotonically increasing

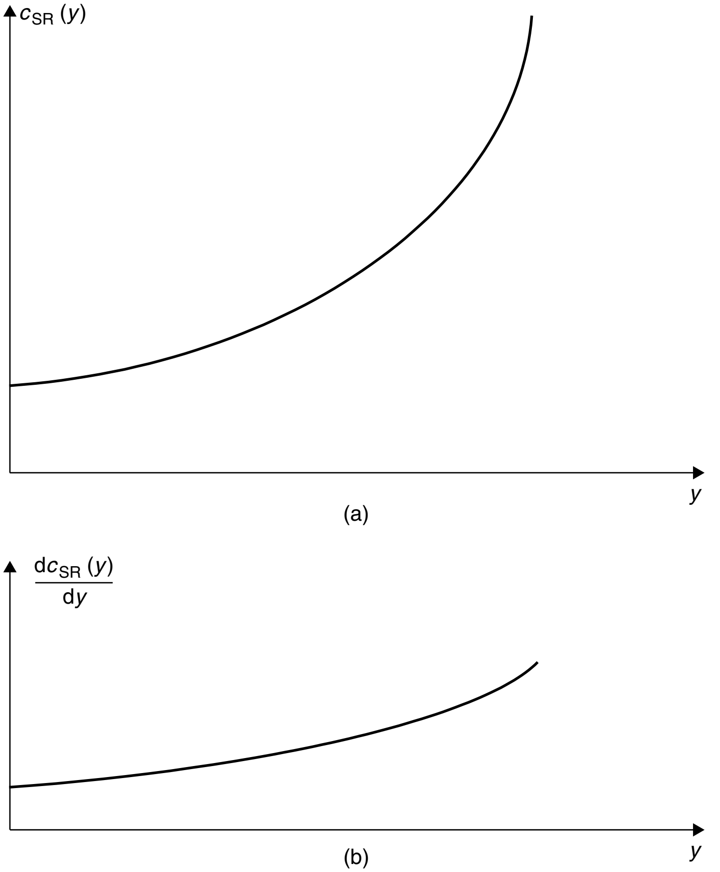


Figure 2.17 (a) Typical short-run cost function and (b) corresponding short-run marginal cost function

function of the quantity produced. The marginal cost function corresponding to the cost function of Figure 2.17(a) is shown in Figure 2.17(b). Note that if the cost of production is expressed in dollars, the marginal cost is expressed in dollars per unit produced. At a given level of production, the numerical value of the marginal cost function is equal to the cost of producing one more unit of the goods.

Using these functions, we can determine the short-run behavior of a firm in a perfectly competitive market. In such a market, no firm can influence the market price. Therefore, the only action that the firms can take to maximize their profits is to adjust their output. Since profit is defined as the difference between the firm's revenues and costs, the optimal level of production is given by

$$\max_{y}\{\pi \cdot y - c_{\mathrm{SR}}(y)\} \tag{2.14}$$

At the optimum, we must have

$$\frac{\mathrm{d}\{\pi \cdot y - c_{\mathrm{SR}}(y)\}}{\mathrm{d}y} = 0$$

or

$$\pi = \frac{\mathrm{d}c_{\mathrm{SR}}(y)}{\mathrm{d}y} \tag{2.15}$$

The firm will thus increase its production up to the point at which the marginal cost is equal to the market price. Gaining an intuitive understanding of this relation is useful. If the firm were operating at a point at which its marginal cost of production is less than the current market price, it could increase its profits by producing another unit and selling it on the market. Similarly, if the firm's marginal cost of production were higher than the market price, it would save money by not producing the last unit it sold.

Defining a long-run cost function is more complicated because, in the long run, the firm has more flexibility in deciding how it will produce. For example, a firm could decide to buy more expensive machines and reduce its labor costs or vice versa. The production function therefore cannot be treated as a function of a single variable. We will, however, assume that the firm behaves in an optimal manner. By this we mean that in the long run it chooses the combination of factor of production that yields any quantity of goods at minimum cost. The long-run cost function is thus the solution of an optimization problem and can be expressed as follows:

$$c_{\mathrm{LR}}(y) = \min_{x_1, x_2}(w_1 \cdot x_1 + w_2 \cdot x_2) \text{ such that } f(x_1, x_2) = y \tag{2.16}$$

where, for the sake of simplicity, we have considered only two factors of production.

In the first part of this book, we will use short-run cost functions because we will be concerned with the operation of an existing power system. Long-run cost functions will be used in the last two chapters of this book when we consider the expansion of the power system.

2.3.3 Costs

In this section, we define the various components of the production cost and introduce various curves that are used to characterize these costs.

In the short run, some factors of production are fixed. The cost associated with these factors does not depend on the amount produced and is thus a *fixed cost*. For example, if a generating company has bought land and built a power plant on this land, the costs of the land and the plant do not depend on the amount of energy that this plant produces. On the other hand, the quantity of fuel consumed by this plant and, to a certain extent, the manpower required to operate it depend on the amount of energy it produces. Fuel and manpower costs are thus examples of *variable cost*s. There is also a third class of costs called *quasi-fixed costs*. These are costs that the firm incurs if the plant produces any amount of output but does not incur if the plant produces nothing. For example, in the case of a generating plant, the cost of the fuel required to start up the plant is fixed in the sense that it does not depend on the amount of energy that the plant goes on producing. However, this start-up cost does not need to be paid if the plant stays idle.

In the long run, there are no fixed costs because the firm can decide on the amount of money it spends on all production factors. At the limit, the firm's long-run costs can be zero if it decides to produce nothing and goes out of business. A *sunk cost* is the difference between the amount of money a firm pays for a production factor and the amount of money it would get back if it sold this asset. For example, in the case of a power plant, the cost of the land on which the plant is built is not a sunk cost because land can always be resold. It is thus a *recoverable cost*. On the other hand, if production with this plant is no longer profitable, the difference between the cost of building the plant and its scrap metal value is a sunk cost.

2.3.3.1 Short-run costs

If we assume that the costs of the production factors are constant, the cost functions defined in the previous section can be expressed as a function of the level of output y:

$$c(y) = c_v(y) + c_f \tag{2.17}$$

where $c_v(y)$ represents the variable costs and c_f represents the fixed costs. The *average cost function* measures the cost per unit of output. It is equal to the sum of the *average variable cost* and the *average fixed cost:*

$$AC(y) = \frac{c(y)}{y} = \frac{c_v(y)}{y} + \frac{c_f}{y} = AVC(y) + AFC(y) \tag{2.18}$$

Let us sketch what these average cost curves might look like. Since the fixed costs do not depend on the production, the average fixed cost is infinite for a zero output. As production increases, these fixed costs are spread over an increasing output. The average fixed-cost curve is thus a monotonically decreasing function, as shown in Figure 2.18(a). For modest production levels, variable costs typically increase linearly with the output. The average variable cost is therefore constant. If production can be

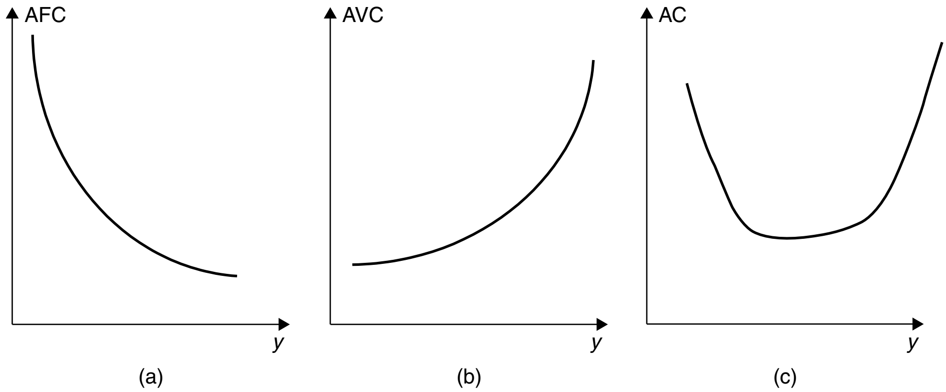


Figure 2.18 Typical shapes of the average cost functions

made more efficient as the output increases, the average variable cost might actually decrease somewhat as production increases. However, as the production increases, as Figure 2.18(b) shows, the average variable cost eventually and inevitably rises. This increase in the average variable cost arises because the fixed factors start constraining the production of goods. For example, the output of a manufacturing plant can often be increased beyond the capacity for which the plant was designed. However, this might require paying the workers overtime pay, maintaining the machines more frequently and generally adopting less-efficient procedures. Similarly, in the case of generating plants, the maximum efficiency is often achieved for an output that is somewhat below the maximum capacity of the plant. The average cost curve combines these two effects and has the typical U-shape shown in Figure 2.18(c).

It is essential to understand the difference between the average and the marginal costs. Both quantities are expressed in dollars per unit produced, but the marginal cost reflects only the cost of the last unit produced. On the other hand, the average cost factors in the cost of all the units already produced. Since the fixed costs are constant, they do not contribute to the marginal cost. Figure 2.19 illustrates the relation between the marginal cost curve and the average cost. For low production levels, the marginal cost is smaller than the average cost because of the influence of the fixed costs. On the other hand, for high production levels, the marginal cost is higher than the average cost. The marginal cost curve intersects the average cost curve at its minimum.

2.3.3.2 Long-run costs

We have argued above that, in the long run, there are no fixed costs because all the factors of production can be changed and the firm has the option to produce nothing and get out of business. However, the technology may be such that some costs are incurred independently of the level of production. There may therefore be some quasi-fixed costs in the long run. The long-run average cost curve therefore tends to have a U-shape, as shown in Figure 2.20. What can we say about the relation between the short-run cost and the long-run cost? In the long run, we can minimize the production cost for any level of output because we can adjust all the factors of production. On the other hand, in the short run, some of the production factors are fixed. The short-run

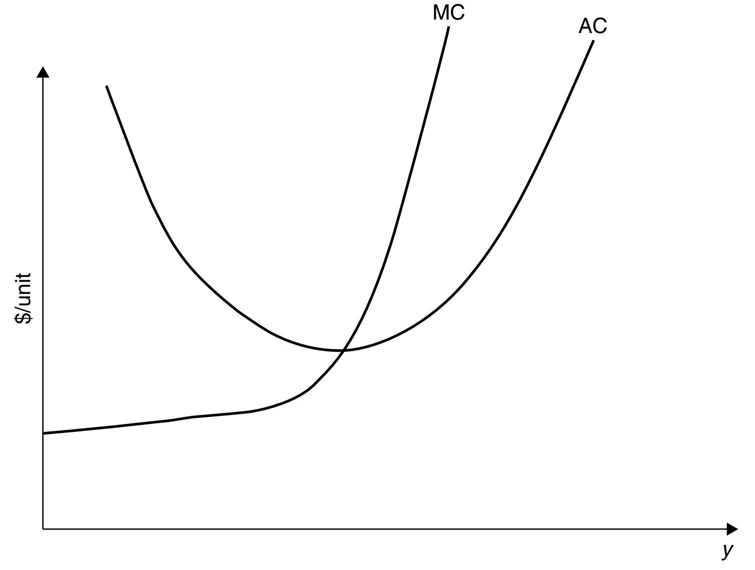


Figure 2.19 Typical relation between the average cost curve (AC) and the marginal cost curve (MC)

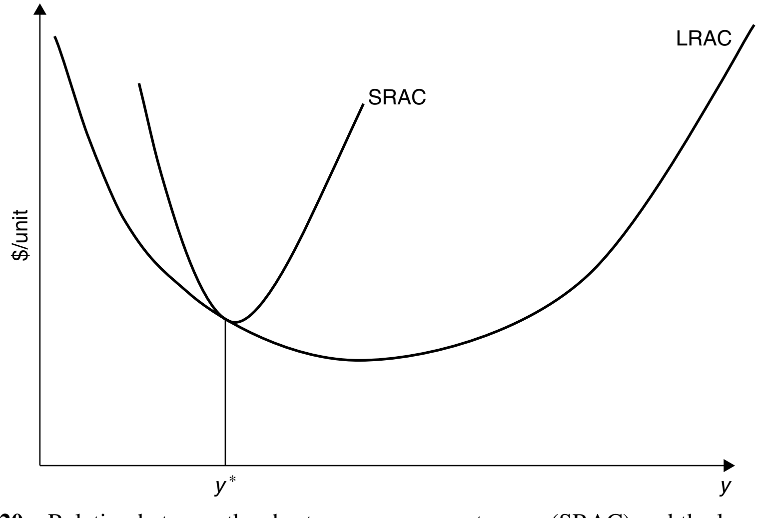


Figure 2.20 Relation between the short-run average cost curve (SRAC) and the long-run average cost curve (LRAC) if the fixed production factors are chosen to minimize the production cost for a value y^* of the output

production cost is therefore equal to the long-run production cost only for the value of output y^* for which the fixed production factors were optimized. For other levels of output, the short-run cost is higher than the long-run cost. The short-run average cost curve is therefore above the long-run average cost curve, except for the output for which the fixed production factors have been optimized. At that point, the two curves are tangent, as shown in Figure 2.20. We could, of course, select other sets of fixed production factors that would minimize the production cost for other values of the output $y_1, y_2, \ldots y_n$. In other words, we could build plants with other capacities. For each plant size, the short-run average cost would be equal to the long-run average cost only for the designed plant capacity. As Figure 2.21 shows, the long-run average cost curve is therefore the lower envelope of the short-run average cost curves.

When all factors of production can be adjusted, the cost of a unit increase in production is given by the *long-run marginal cost curve*. Figure 2.22 illustrates two observations about this long-run marginal cost curve. First, the long-run and short-run marginal costs are equal only for the production level y^* for which the fixed production

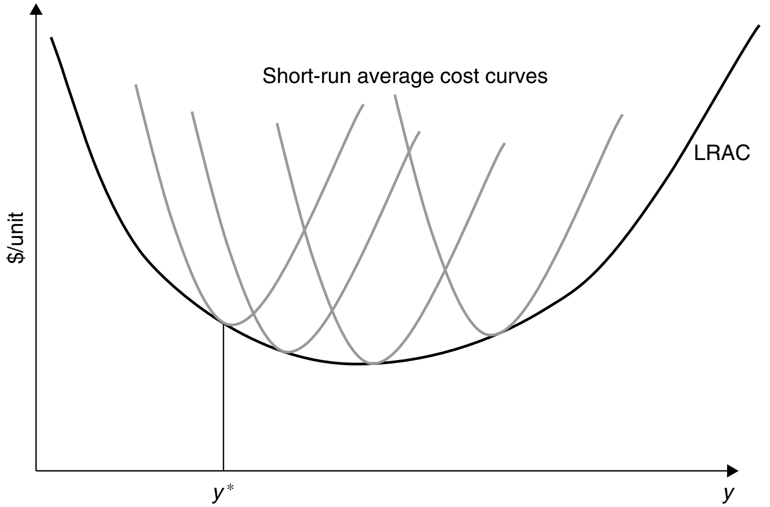


Figure 2.21 The long-run average cost curve (LRAC) is the lower envelope of the short-run average cost curves

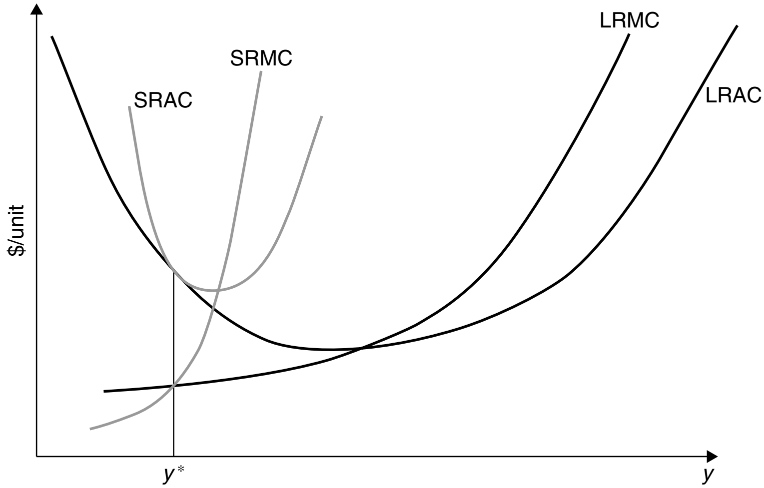


Figure 2.22 Relation between the short-run average cost (SRAC), the short-run marginal cost (SRMC), the long-run average cost curve (LRAC) and the long-run marginal cost (LRMC)

factors have been optimized. Second, the long-run marginal cost is equal to the long-run average cost for the production level that results in the minimum long-run average cost. As long as the long-run marginal cost is smaller than the long-run average cost, this long-run average cost decreases. As long as the average cost decreases, the production is said to exhibit *economies of scale*.

2.4 Types of Markets

So far, we have treated markets only as a mechanism for matching the supply and the demand for a commodity through the discovery of an equilibrium price. We are now going to discuss how a market might operate and how different types of markets serve different purposes.

Besides the obvious need to agree on the quality, quantity and price of the goods, three other important matters must be decided when a buyer and a seller arrange a trade:

- The date of delivery of the goods
- The mode of settlement
- Any conditions that might be attached to this transaction.

How buyers and sellers settle these matters defines the type of contract that they conclude and hence the type of market in which they participate.

2.4.1 Spot market

In a spot market, the seller delivers the goods immediately and the buyer pays for them "on the spot". No conditions are attached to the delivery. This means that neither party can back out of the deal. A fruit and vegetable market is a good example of a spot market: you inspect the quality of the produce and tell the vendor how many cucumbers you want, she hands them to you, you pay the price indicated and the transaction is complete. If later on you decide that you would rather eat lettuce, you probably would not even think of trying to return the cucumbers and getting your money back. On the surface, the rules of such markets may appear very informal. In fact, they have behind them the weight of centuries of tradition. Modern spot markets for commodities such as oil, coffee or barley are superficially more sophisticated because the quantities traded are much larger and because traders communicate electronically. However, the principles are exactly the same.

A spot market has the advantage of immediacy. As a producer, I can sell exactly the amount that I have available. As a consumer, I can purchase exactly the amount I need. Unfortunately, prices in a spot market tend to change quickly. A sudden increase in demand (or a drop in production) sends the price soaring because the stock of goods available for immediate delivery may be limited. Similarly, a glut in production or a dip in demand depresses the price. Spot markets also react to news about the future availability of a commodity. For example, a forecast about a bumper harvest of an agricultural commodity could send its price on the spot market (the spot price) plunging if enough consumers have the ability to wait until this harvest comes to market. Changes in the spot price are essentially unpredictable because if they were predictable, the market participants would anticipate them.

Large and unpredictable variations in the price of a commodity make life harder for both suppliers and consumers of this commodity. Both are running businesses and are thus facing a variety of risks. Bad weather or a pest can ruin a harvest. The breakdown of a machine can stop production. A strike can stop the shipment of finished goods. While being in business means taking some risks, an excessive amount of risk endangers the survival of a business. Most businesses will therefore try to reduce their exposure to price risks. For example, the producer of a commodity will try to avoid being forced to sell its output at a very low price. Similarly, a consumer does not want to be obliged to buy an essential commodity at a very high price. This desire to avoid being exposed to the wild price fluctuations that are common in spot markets has led to the introduction of other types of transactions and markets. These markets are described in the following sections.

2.4.2 Forward contracts and forward markets

Imagine that J. McDonald is a farmer who raises wheat. Even though it is early summer, he is very confident that he will be able to deliver 100 tons at harvest time. On the other hand, he is very concerned about price fluctuations. He would very much like to "lock in" an acceptable price now and stop worrying about having to sell at a low price when the wheat is ripe. Will he be able to find someone ready to agree to such a deal? Just like farmers are concerned about having to sell at a low price, the Pretty Good Breakfast food-processing company does not want to have to pay a high price for the wheat it uses to make its well-known pancake mix. If an acceptable price can be agreed, it is ready to sign a contract with farmer McDonald now for the delivery of his wheat harvest in a few months time. This *forward contract* specifies the following:

- The quantity and quality of the wheat to be delivered
- The date of delivery
- The date of payment following delivery
- The penalties if either party fails to honor its commitment
- The price to be paid.

On what basis can the farmer and the food-processing company agree on a price for a delivery of a commodity in a few months time when even the spot price is volatile? Both parties start by calculating their best estimate of what the spot price might be at the time of delivery. This estimate takes into account historical data about the spot price and any other information that the farmer and the food-processing company might have about harvest yields, long-term weather forecasts and demand forecasts. Since a lot of that information is publicly available, the estimates of both parties at any given time are unlikely to be very different. However, the price agreed for the contract may differ from the best estimates because of differences in bargaining positions. If farmer

McDonald is concerned about the possibility of a very low price on the spot market, he may agree to a price below his expected value of this spot market price. The difference between his expectation of the spot market price and the price agreed in the forward contract represents a *premium* that he is willing to pay to reduce his exposure to a downward price risk. On the other hand, if the food-processing company is vulnerable to an upward price risk, farmer McDonald might be able to extract a price that reflects a premium above his expectations of the spot market price.

If the spot price at the time of delivery is higher than the agreed price, the forward contract represents a loss for the seller and a profit for the buyer. On the other hand, if the spot price is lower than the agreed price, the forward contract represents a loss for the buyer and a profit for the seller. These profits and losses are "paper profits" and "paper losses" because they only reflect the fact that a party could have done better and the other worse by trading on the spot market. Nevertheless, a paper loss makes a company less competitive because it means that it has bought or sold a commodity at a worse price than some of its competitors did.

Forward contracts make it possible for parties to trade at a price acceptable to both sides and hence provides a way to share the price risk. Over the years, these two parties could enter into similar forward contracts with a premium over or below the expected spot price. If their estimates of future spot prices are unbiased, in the long run the difference between the average spot price and the average forward price should be equal to the average premium. The party that gets the premium is therefore being remunerated for accepting the price risk.

Going back to our agricultural example, suppose that the Pretty Good Breakfast Company signs every year a forward contract with farmer McDonald at a price that is below the expected spot price for wheat at the time of delivery. In the long run, the company should profit from accepting to shoulder this risk. In the short run, however, it may have to endure a string of large losses if the spot price moves in the wrong direction. To ride through such losses, it must have large financial reserves or demand a substantial premium. If the premium it demands is too large, farmer McDonald may decide that signing a forward contract with the Pretty Good Breakfast Company is not worthwhile. Could other food-processing companies offer him a better deal? Similarly, the Pretty Good Breakfast Company will look for other farmers who might agree to sign forward contracts. If enough farmers and food-processing companies are interested in trading wheat in advance of delivery, a *forward market* for wheat will develop. The establishment of such a market gives all parties access to a large number of possible trading partners and helps them determine whether the price they are being offered is reasonable.

In some cases, two parties may want to negotiate all the details of a forward contract. This approach is justified if the contract is designed to cover the delivery of a large quantity of a commodity over a long period of time or if special terms need to be discussed. Since such negotiations are expensive, many forward contracts use standardized terms and conditions. This standardization makes possible the resale of forward contracts. For example, let us suppose that the sale of a new Belgian waffle mix manufactured by the Pretty Good Breakfast Company does not meet expectations. Over the summer, the company realizes that it will not need all the wheat for which it has signed forward contracts. Rather than wait until the contracted date of delivery to sell the excess wheat on the spot market, it can resell the forward contracts it holds to other food-processing companies. Other producers will have signed contracts during the spring. As the summer goes by, some of them may realize that they have overestimated the quantity that they will be able to produce. If they cannot deliver the quantities specified in the contracts, they will have to cover the deficit by buying wheat on the spot market. Rather than hope that the spot price will be favorable on the date of delivery, these producers could buy the forward contract from the Pretty Good Breakfast Company to offset their anticipated deficit. The price at which forward contracts are traded will be the current market price for forward contracts with the same delivery date. Depending on the market's view of the evolution of the spot price, this resale price may be higher or lower than the price agreed by the originators of the contract.

2.4.3 Future contracts and futures markets

The existence of a *secondary market* where producers and consumers of the commodity can buy and sell standardized forward contracts helps these parties manage their exposure to fluctuations in the spot price. Participation in this market does not have to be limited to firms that produce or consume the commodity. Parties that cannot take physical delivery of the commodity may also want to take part in such a market. These parties are speculators who want to buy a contract for delivery at a future date, in the hope of being able to sell it later at a higher price. Similarly a speculator can sell a contract first, hoping to buy another one later at a lower price. Since these contracts are not backed by physical delivery, they are called *future contracts* rather than forwards. As the date of delivery approaches, the speculators must balance their position because they cannot produce, consume or store the commodity.

At this point, we may wonder why any rational person might want to engage in this type of speculation. If the markets are sufficiently competitive and all participants have access to enough information, the forward price should reflect the consensus expectation of the spot price. Hence buying low in the hope of selling high would seem more like gambling than a sound business strategy. To be successful as a speculator, therefore, one needs an advantage over other parties. This advantage is usually being less risk adverse than other market participants. Shareholders in some companies expect stable but not extraordinary returns. The management of these risk-averse companies will therefore try to limit its exposure to risks that might reduce profits significantly below expectations. On the other hand, shareholders in companies that engage in commodity speculation hope for very high returns but should not be surprised by occasional large losses. The management of these risk-loving companies will therefore feel free to take significant risks in order to secure larger profits. A risk-averse company will usually accept a price somewhat worse than it might be able to get later in exchange for the security of getting a fixed price now. A speculator, on the other hand, will demand a better price in exchange for accepting to shoulder the risk of future fluctuations. In essence, risk-averse companies remunerate speculators for their willingness to buy the risk.

As we discussed in the section on spot markets, producers and consumers of a commodity face other risks besides the price risk. They are therefore usually quite eager to pay another party to reduce their exposure to this additional risk. A speculator does not face other risks and has large financial resources that put it in a better position to offset losses against profits over a long period of time. In addition, most speculators do not limit themselves to one commodity. By diversifying into markets for different commodities, they further reduce their exposure to risk. Even though speculators make a profit from their trades, the market as a whole benefits from their activities because their presence increases the number and diversity of market participants. *Physical participants* (i.e. those who produce or consume the commodity) thus find counterparties for their trades more easily. This increased *liquidity* helps the market discover the price of a commodity.

2.4.4 Options

Futures and forwards contracts are *firm contracts* in the sense that delivery is unconditional. Any seller who is unable to deliver the quantity agreed must buy the missing amount on the spot market. Similarly, any buyer who cannot take full delivery must sell the excess on the spot market. In other words, imbalances are liquidated at the spot price on the date of delivery.

In some cases, participants may prefer contracts with a conditional delivery, which means contracts that are exercised only if the holder of the contract decides that it is in its interest to do so. Such contracts are called *options* and come in two varieties: *calls* and *puts*. A call option gives its holder the right to buy a given amount of a commodity at a price called the *exercise price*. A put option gives its holder the right to sell a given amount of a commodity at the exercise price. Whether the holder of an option decides to exercise its rights under the contract depends on the spot price for the commodity. A European option can be exercised only on its expiry date, while an American option can be exercised at any time before the expiry date. When an option contract is agreed, the seller of the option receives a nonrefundable option fee from the holder of the option.

2.4.4.1 Example 2.1

On June 1, the Pretty Good Breakfast Company purchased from farmer McDonald a European call option for 100 tons of wheat with an expiry date of 1 September and an exercise price of \$50 per ton. On 1 September, the spot price for wheat stands at \$60 per ton. Buying wheat on the spot market would cost the company \$10 per ton more than exercising the option. This call option therefore has a value of $100 \times 10 = \$1000$. The option thus gets exercised: farmer McDonald delivers 100 tons of wheat and the company pays $100 \times 50 = \$5000$.

On the other hand, if the spot price on 1 September is lower than the exercise price of the call, the option is worthless and lapses because it is cheaper for the company to buy wheat on the spot market.

2.4.4.2 Example 2.2

On 1 July, farmer McDonald bought a European put option for 100 tons of wheat from the Great Northern Wheat Trading Company. The exercise price of this contract is \$55 per ton and the expiry date is 1 September. If on 1 September the spot price for wheat is \$60, farmer McDonald does not exercise the option and sells his wheat on the spot market instead. On the other hand, if the spot price is \$50 per ton, the option has a value of $100 \times (55 - 50) = \$500$ and is obviously exercised.

Buying an option contract can therefore be viewed as a way for the holder of the contract to protect itself against the risk of having to trade the commodity at a price less favorable than the spot price. At the same time, it leaves the holder free to trade at a price that is better than the exercise price of the option. The seller of the option assumes the price risk in the place of the holder. In exchange for taking this risk, the seller receives the option fee when the contract is sold. This option fee represents a sunk cost for the buyer and does not affect whether the option is exercised or not.

It is worth noting at this point that option contracts for the delivery of electrical energy are not commonly traded. On the other hand, long-term contracts for the provision of reserve often include both an option fee and an exercise price and thus operate like option contracts.

2.4.5 Contracts for difference

Producers and consumers of some commodities are sometimes obliged to trade solely through a centralized market. Since they are not allowed to enter into bilateral agreements, they do not have the option to use forward, future or option contracts to reduce their exposure to price risks. In such situations, parties often resort to *contracts for difference* that operate in parallel with the centralized market. In a contract for difference, the parties agree on a *strike price* and an amount of the commodity. They then take part in the centralized market like all other participants. Once trading on the centralized market is complete, the contract for difference is settled as follows:

- If the strike price agreed in the contract is higher than the centralized market price, the buyer pays the seller the difference between these two prices times the amount agreed in the contract.
- If the strike price is lower than the market price, the seller pays the buyer the difference between these two prices times the agreed amount.

A contract for difference thus insulates the parties from the price on the centralized market while allowing them to take part in this market. A contract for difference can be described as a combination of a call option and a put option with the same exercise price. Unless the market price is exactly equal to the strike price, one of these options will necessarily be exercised.

2.4.5.1 Example 2.3

The Syldavia Steel Company is required to purchase its electrical energy from the Central Electricity Market of Syldavia. Because the price of energy in that market is highly volatile and it wants to limit its exposure to price risks, Syldavia Steel has signed a contract for difference with the Quality Electrons Generating Company. This contract specifies a uniform quantity of 500 MW and a uniform price of 20 \$/MWh at all hours of the day for a one-year period. Let us suppose that the market price for a specific one-hour trading period is 22 \$/MWh. Syldavia Steel pays to the Central Electricity Market $22 \times 500 = \$11\,000$ for the purchase of 500 MW during that hour. Quality Electrons gets paid $22 \times 500 = \$11\,000$ by the Central Electricity Market for the supply of 500 MW during the same hour. To settle their contract for difference, Quality Electrons pays Syldavia Steel $(22 - 20) \times 500 = \$1000$. Both firms have thus effectively traded 500 MW at 20 \$/MWh. If the market price had been less than 20 \$/MWh, Syldavia Steel would have paid Quality Electrons to settle the contract.

2.4.6 Managing the price risks

Firms that produce or consume large amounts of a commodity are exposed to other types of risk and will generally try to reduce their exposure to price risks by hedging their positions using a combination of forwards, futures, options and contracts for differences. Markets for these different types of contracts develop for all major commodities. Firms tend to use the spot market only for the residual volumes that result from unpredictable fluctuations in demand or production. The volume of trades in the spot market therefore represents only a small fraction of the volume traded on the other markets.

While the spot market volume may be relatively small, the spot price is the signal that drives all the other markets. Since the spot market is the market of last resort, the spot price represents the alternative against which other opportunities must be measured. A sustained increase in the spot price will therefore drive up the prices on the other markets, while a continuing reduction will force them lower.

2.4.7 Market efficiency

The theory that we developed at the beginning of this chapter suggests that if two parties put different values on the same good, a trade should take place. If such transactions are to happen quickly and easily, the market must be *liquid.* This means that there should always be enough participants willing to buy or sell goods. The mechanism through which the market price is discovered should also be reliable. Good mechanisms for disseminating widely comprehensive and unbiased information about the market conditions are indispensable to this price-discovery process. Participants will also have more confidence in the fairness of the market if its operation is as transparent as possible. Finally, the costs associated with trading (fees, administrative expenses and cost of gathering market information) should represent a small fraction of the value of each transaction. These transaction costs are considerably reduced if the commodity being traded is standardized in terms of quantity and quality. A market that satisfies these criteria is said to be *efficient*.

2.5 Markets with Imperfect Competition

2.5.1 Market power

So far, we have assumed that no market participant has the ability to influence the market price through its individual actions. This assumption is valid if the number of market participants is large and if none of them controls a large proportion of the production or consumption. Under these circumstances, any supplier who asks more than the market price and any consumer who offers less than the market price will simply be ignored because others can replace their contribution to the market. The price is thus set by the interactions of the buyers and the sellers, taken as groups. A market in which all participants act as price takers is said to have *perfect competition.* Achieving or approximating perfect competition is a very desirable goal from a global perspective because it ensures that the marginal cost of production is equal to the marginal value of the goods to the consumers. Such a situation encourages efficient behavior on both sides.

Markets for agricultural commodities are one of the best examples of perfect competition because the number of small producers and consumers of an undifferentiated commodity is very large. For many other goods, some producers and consumers control a share of the market that is large enough to enable them to exert *market power.* These market participants are called strategic players. As the following example shows, prices can be manipulated either by withholding quantity (physical withholding) or by raising the asking price (economic withholding).

2.5.1.1 Example 2.4

Suppose that a firm sells 10 widgets at a market price of \$1800 per widget. Its revenue from the sale of widgets is thus \$18 000. If this firm decides to offer only nine widgets for sale and as a consequence the market price for widgets rises, this firm has market power. If the price rises to \$2000, the firm achieves the same revenue even though it sells fewer widgets. Furthermore, its profit increases because it incurs the cost of producing only 9 widgets instead of 10.

Instead of withholding production, this firm could offer to sell nine widgets at \$1800 and one widget at a higher price in the hope that this last widget will sell and boost its profits.

2.5.2 Models of imperfect competition

In a perfectly competitive market, the market price is a parameter over which firms have no control. Equation (2.15) led us to the conclusion that under perfect competition, each firm should increase its production up to the point where its marginal cost is equal to the market price. When competition is not perfect, each firm must consider how the quantity it produces might affect the market price. Conversely, it should consider how the price it chooses might affect the quantity it sells. Imperfect competition can be modeled using either a Cournot model, in which firms decide how much they produce, or a Bertrand model, in which firms decide at what price they sell their production.

2.5.2.1 Cournot model

We will first consider the case of a *duopoly.* This is the case in which only two firms compete to sell a given good in a market. If both firms must decide simultaneously how much to produce, each of them will estimate the expected production of the other. Let us assume that firm 1 estimates that the production of firm 2 will be equal to y_2^e. Firm 1 then sets its production at a level y_1 that maximizes its expected profit:

$$\max_{y_1} \pi(y_1 + y_2^e)y_1 - c(y_1) \tag{2.19}$$

where $\pi(y_1 + y_2^e)$represents the market price that would result from the expected total output $y_1 + y_2^e$. The optimal production of firm 1 thus depends on its estimate of the production of firm 2. We can express this relation directly in the form of a *reaction function*:

$$y_1 = f_1(y_2^e) \tag{2.20}$$

Since firm 2 follows a similar process to optimize its production, we also have

$$y_2 = f_2(y_1^e) \tag{2.21}$$

At first, the estimates that each firm makes of the production of their competitor may be incorrect or inaccurate. However, as they observe the market and gather more information, they revise their estimates and adjust their production accordingly. Ultimately, their productions reach the *Cournot equilibrium*:

$$\begin{aligned} y_1^* &= f_1(y_2^*) \\ y_2^* &= f_2(y_1^*) \end{aligned} \tag{2.22}$$

Once this equilibrium is reached, neither firm would find it profitable to change its output.

Let us now consider the case in which there are n firms competing in the market. The total industry output is

$$Y = y_1 + \cdots + y_n \tag{2.23}$$

Firm i, like all the other firms, seeks to maximize its profit:

$$\max_{y_i}\{y_i \cdot \pi(Y) - c(y_i)\} \tag{2.24}$$

where the market price $\pi(Y)$is a function of the total industry output. This maximum is achieved when

$$\frac{\mathrm{d}}{\mathrm{d}y_i}\{y_i \cdot \pi(Y) - c(y_i)\} = 0 \tag{2.25}$$

or

$$\pi(Y) + y_i \frac{\mathrm{d}\pi(Y)}{\mathrm{d}y_i} = \frac{\mathrm{d}c(y_i)}{\mathrm{d}y_i} \tag{2.26}$$

Factoring out $\pi(Y)$ on the left-hand side and multiplying the second term by Y/Y, we get

$$\pi(Y)\left\{1 + \frac{y_i}{Y}\frac{Y}{\mathrm{d}y_i}\frac{\mathrm{d}\pi(Y)}{\pi(Y)}\right\} = \frac{\mathrm{d}c(y_i)}{\mathrm{d}y_i} \tag{2.27}$$

The right-hand side of this equation is equal to the marginal cost of production of firm i. If we define the *market share* of firm i as $s_i = y_i/Y$ and use the definition of elasticity given in Equation (2.3), we can write Equation (2.27) in the following form:

$$\pi(Y)\left\{1 - \frac{s_i}{|\varepsilon(Y)|}\right\} = \frac{\mathrm{d}c(y_i)}{\mathrm{d}y_i} \tag{2.28}$$

This expression shows that when the market share of a firm is not negligible, it maximizes its profit by setting its production at a level where its marginal cost is less than the market price. Equation (2.28) suggests that a low price elasticity of demand and a degree of market concentration facilitate the exercise of market power. It is interesting to note that one firm's ability to exert market power benefits all the firms in the market because it raises the price at which price-taking firms sell their products. Actions aimed at reducing market power therefore have to be initiated by regulatory authorities representing the interests of the customers. Such actions usually do not receive support from any of the producers.

Equation (2.28) is applicable to the extreme cases in which the firm has a monopoly ($s_i = 1$) and where its market share is negligible ($s_i \approx 0$). The biggest difference between price and marginal cost occurs in the case of a monopoly, where the monopolist's ability to raise prices is limited only by the elasticity of the demand. In the case of a firm with a very small market share, Equation (2.28) reduces to the same form as Equation (2.15), and the firm acts as a price taker.

2.5.2.2 Bertrand model

Let us consider the case of a market in which two firms produce identical products and have identical marginal cost curves. The Bertrand model assumes that these firms compete by setting their prices and letting the market decide how much each firm sells. Neither firm can set its price below its marginal cost of production because this would result in a loss. If firm 1 decides to set its price above the marginal cost, firm 2 can capture the entire market by setting its price just below the price set by firm 1. The products made by both firms being identical, all consumers would indeed choose the cheaper one. Since firm 1 could retaliate by setting its price below the price of firm 2, this model suggests that a sustainable equilibrium is reached only when the price matches the marginal cost of production.

This result is counterintuitive because one would expect duopoly competitors to be able to obtain a higher price than would result from a perfectly competitive market. On the other hand, the Bertrand model can be viewed as representing competitive bidding between firms that cannot collude. In such cases, the competitive price is the best price that firms can rationally expect to achieve.

In practice, however, collusion sometimes occurs. Firms can form a *cartel* to try to set prices and outputs in a way that maximizes profits for the industry as a whole. Since cartels are illegal, collusion often takes a subtler, tacit form in which firms that compete on a regular basis send each other signals through published prices. While a firm could achieve a bigger profit in the short term by undercutting its competitors' prices, it may realize that in the long run it is in all the firms' interest to maintain prices higher.

2.5.3 Monopoly

The *Minimum Efficient Size* (MES) of a firm in a particular industry provides a rough indication of the number of competitors that one is likely to find in the market for the product of this industry. This MES is equal to the level of output that minimizes the average cost for a typical firm in that industry. The shape of this curve is determined by the technology used to produce the goods. If, as illustrated in Figure 2.23(a), the MES is much smaller than the demand for the goods at this minimum average cost, the market should be able to support a large number of competitors. On the other hand, if, as shown in Figure 2.23(b), the MES is comparable to the demand, the market cannot support two profitable firms and a monopoly situation is likely to develop.

As Equation (2.28) suggests, given the opportunity, a monopolist will reduce its output and raise its price above its marginal cost of production to maximize its profit. From a global perspective, this is not satisfactory because consumers purchase less of the good than if it was sold on a competitive basis. One possible remedy to this problem is to establish a regulatory body whose function is to monitor the activities of the producer and set the price at an acceptable level. Ideally, the regulator should set the price at the marginal cost of the monopoly firm. Determining this marginal cost is not an easy task because the regulator does not have access to the same amount of information as the monopolist. Even when the regulator succeeds in determining accurately the marginal cost, setting the price at that level may not be acceptable because it could bankrupt the monopolist. For example, in the case shown in Figure 2.24, the intersection of the demand and marginal cost curves gives a price π_{MC} that is lower than the average cost of production. To avoid driving the monopolist out of business, the regulator should set the price at least at the value π_{AC} given by the intersection of the demand curve and the average cost curve AC. Such a situation is called a *natural monopoly*. It arises when producing the goods involves large fixed costs and relatively small variable costs. Transmission and distribution of electrical energy are very good examples of natural monopolies.

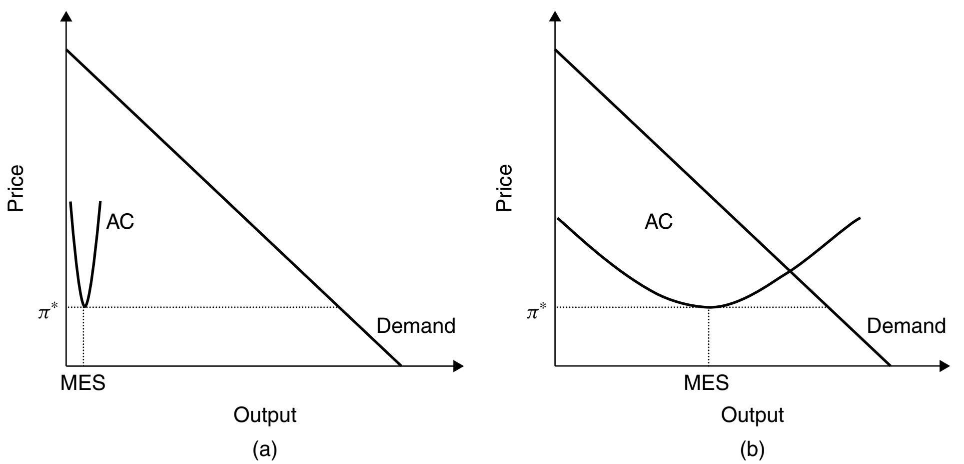


Figure 2.23 Concept of MES. (a) Competitive market, (b) Monopoly situation

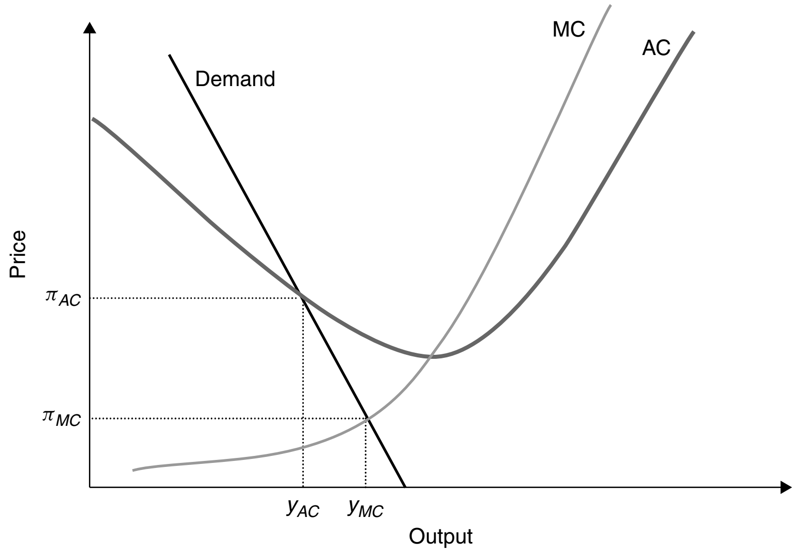


Figure 2.24 Unprofitable natural monopoly

2.6 Further Reading

The reader interested in studying microeconomics in more depth will find an abundance of textbooks on this subject. Most of these textbooks cover the same topics albeit at different levels. Engineers may find introductory texts somewhat frustrating because they assume that the reader does not know calculus. Explanations thus tend to be long and wordy. Intermediate level books, such as Varian (1999), are probably a better choice. An even more rigorous and mathematical treatment of the subject can be found in (Gravelle and Rees, 1992). Tirole (1988) analyzes in considerable depth the theory of the firm. Hunt and Shuttleworth (1996) provide a very readable introduction to the various types of contracts. A discussion of how the concept of elasticity can be applied in electricity markets can be found in (Kirschen *et al*., 2000). Borenstin (1999, 2001) has written very readable discussions of market power in electricity markets.

Borenstin S, *Understanding Competitive Pricing and Market Power in Wholesale Electricity Markets*, Working Paper PWP-067, August 1999, Program on Workable Energy Regulation, University of California Energy Institute. Available online from *www.ucei.org*.

Borenstin S, *The Trouble with Electricity Markets (and some solutions)*, Working Paper PWP-081, January 2001, Program on Workable Energy Regulation, University of California Energy Institute. Available online from *www.ucei.org*.

Gravelle H, Rees R, *Microeconomics*, Second Edition, London, Longman, 1992.

Hunt S, Shuttleworth G, *Competition and Choice in Electricity*, Wiley, Chichester, 1996.

Kirschen D S, Strbac G, Cumperayot P, Mendes D P, Factoring the elasticity of demand in electricity prices, *IEEE Transactions on Power Systems*, **15**(2), 2000, 612–617.

Tirole J, *The Theory of Industrial Organization*, MIT Press, Cambridge, MA, 1988.

Varian H R, *Intermediate Microeconomics: A Modern Approach*, Fifth Edition, W. W. Norton, New York, 1999.

2.7 Problems

2.1 A manufacturer estimates that its variable cost for manufacturing a given product is given by the following expression:
$C(q) = 25q^2 + 2000q$ [$] where C is the total cost and q is the quantity produced

a. Derive an expression for the marginal cost of production

b. Derive expressions for the revenue and the profit when the widgets are sold at marginal cost.

2.2 The inverse demand function of a group of consumers for a given type of widgets is given by the following expression:
$\pi = -10q + 2000$[$] where q is the demand and π is the unit price for this product

a. Determine the maximum consumption of these consumers

b. Determine the price that no consumer is prepared to pay for this product

c. Determine the maximum consumers' surplus. Explain why the consumers will not be able to realize this surplus

d. For a price π of 1000 $/unit, calculate the consumption, the consumers' gross surplus, the revenue collected by the producers and the consumers' net surplus.

e. If the price π increases by 20%, calculate the change in consumption and the change in the revenue collected by the producers.

f. What is the price elasticity of demand for this product and this group of consumers when the price π is 1000 $/unit

g. Derive an expression for the gross consumers' surplus and the net consumers' surplus as a function of the demand. Check these expressions using the results of part d.

h. Derive an expression for the net consumers' surplus and the gross consumers' surplus as a function of the price. Check these expressions using the results of part d.

2.3 Economists estimate that the supply function for the widget market is given by the following expression:

$$q = 0.2 \cdot \pi - 40$$

a. Calculate the demand and price at the market equilibrium if the demand is as defined in Problem 2.2.

b. For this equilibrium, calculate the consumers' gross surplus, the consumers' net surplus, the producers' revenue, the producers' profit and the global welfare.

2.4 Calculate the effect on the market equilibrium of Problem 2.3 of the following interventions:

a. A minimum price of $900 per widget

b. A maximum price of \$600 per widget

c. A sales tax of \$450 per widget.

In each case, calculate the market price, the quantity transacted, the consumers' net surplus, the producers' profit and the global welfare. Illustrate your calculations using diagrams.

2.5 The demand curve for a product is estimated to be given by the expression:

$$q = 200 - \pi$$

Calculate the price and the price elasticity of the demand for the following values of the demand: 0, 50, 100, 150 and 200.

Repeat these calculations for the case in which the demand curve is given by the expression:

$$q = \frac{10\,000}{\pi}$$

2.6 Vertically integrated utilities often offer two-part tariffs to encourage their consumers to shift demand from on-peak load periods to off-peak periods. Consumption of electrical energy during on-peak and off-peak periods can be viewed as substitute products. The table below summarizes the results of experiments that the Southern Antarctica Power and Light Company has conducted with its two-part tariff. Use these results to estimate the elasticities and cross-elasticities of the demand for electrical energy during peak and off-peak periods.

	On-peak price	**Off-peak price**	**Average on-peak demand**	**Average off-peak demand**
	π_1	π_2	D_1	D_2
	(\$/MWh)	**(\$/MWh)**	**(MWh)**	**(MWh)**
Base case	0.08	0.06	1000	500
Experiment 1	0.08	0.05	992	509
Experiment 2	0.09	0.06	985	510

2.7 Demonstrate that the marginal production cost is equal to the average production cost for the value of the output that minimizes the average production cost.

2.8 A firm's short-run cost function for the production of gizmos is given by the following expression:

$$C(y) = 10y^2 + 200y + 100\,000$$

a. Calculate the range of output over which it would be profitable for this firm to produce gizmos if it can sell each gizmo for \$2400. Calculate the value of the output that maximizes this profit.

b. Repeat these calculations and explain your results for the case in which the short-run cost function is given by

$$C(y) = 10y^2 + 200y + 200\,000$$

3

Markets for Electrical Energy

3.1 Introduction

As a first step in our study of electricity markets, we will assume that all generators and loads are connected to the same bus or that they are connected through a lossless network of infinite capacity. We will, therefore, ignore for now the complexities introduced by the transmission and distribution networks and concentrate on the trading of electrical energy.

Since it is currently not economical to store large quantities of electrical energy, this energy must be produced at pretty much the same time as it is consumed. Trade in electrical energy, therefore, always refers to a certain amount of megawatt-hours to be delivered *over a specified period of time*. The length of this period of time is typically set at an hour, half an hour or quarter of an hour depending on the country or region where the market is located. Since electrical energy delivered during one period is not the same commodity as electrical energy delivered during another period, the price will usually be different for each period. Demand, however, does not change neatly at the beginning of each period. Some adjustments in production must therefore be made on a much shorter basis to keep the system in balance. While such adjustments translate into trades of energy, they are best treated as services rather than commodities, and we will postpone their detailed analysis until Chapter 5.

3.2 What is the Difference Between a Megawatt-hour and a Barrel of Oil?

The development of electricity markets is based on the premise that electrical energy can be treated as a commodity. There are, however, important differences between electrical energy and other commodities such as bushels of wheat, barrels of oil or even cubic meters of gas. These differences have a profound effect on the organization and the rules of electricity markets.

The most fundamental difference is that electrical energy is inextricably linked with a physical system that functions much faster than any market. In this physical

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

power system, supply and demand – generation and load – must be balanced on a second-by-second basis. If this balance is not maintained, the system collapses with catastrophic consequences. Such a breakdown is intolerable because it is not only the trading system that stops working but also an entire region or country that may be without power for many hours. Restoring a power system to normal operation following a complete collapse is a very complex process that may take 24 h or more in large, industrialized countries. The social and economic consequences of such a systemwide blackout are so severe that no sensible government would agree to the implementation of a market mechanism that significantly increases the likelihood of such an event. Balancing the supply and the demand for electrical energy in the short run is thus a process that simply cannot be left to a relatively slow-moving and unaccountable entity such as a market. In the short run, this balance must be maintained, at practically any cost, through a mechanism that does not rely on a market to select and dispatch resources.

Another significant (but somewhat less fundamental) difference between electrical energy and other commodities is that the energy produced by one generator cannot be directed to a specific consumer. Conversely, a consumer cannot take energy from only one generator. Instead, the power produced by all generators is pooled on its way to the loads. This pooling is possible because units of electrical energy produced by different generators are indistinguishable. Pooling is desirable because it results in valuable economies of scale: the maximum generation capacity must be commensurate with the maximum aggregated demand rather than with the sum of the maximum individual demands. On the other hand, a breakdown in a system in which the commodity is pooled affects everybody, not just the parties to a particular transaction.

Finally, the demand for electrical energy exhibits predictable daily and weekly cyclical variations. However, it is by no means the only commodity for which the demand is cyclical. The consumption of coffee, to take a simple example, exhibits two or three rather sharp peaks every day, separated by periods of lower demand. Trading in coffee does not require special mechanisms because consumers can easily store it in solid or liquid form. On the other hand, electrical energy must be produced at the same time as it is consumed. Since its short-run price elasticity of demand is extremely small, matching supply and demand requires production facilities capable of following the large and rapid changes in consumption that take place over the course of a day. Not all of these generating units will be producing throughout the day. When the demand is low, only the most efficient units are likely to be competitive and the others will be shut down temporarily. These less efficient units are needed only to supply the peak demand. Since the marginal producer changes as the load increases and decreases, we should expect the marginal cost of producing electrical energy (and hence the spot price of this energy) to vary over the course of the day. Such rapid cyclical variations in the cost and price of a commodity are very unusual.

One could argue that trading in gas also takes place over a physical network in which the commodity is pooled and the demand is cyclical. However, the amount of energy stored in the gas pipelines is considerably larger than the amount of kinetic energy stored in electricity-generating units. An imbalance between production and consumption of gas would therefore have to last much longer before it would cause a collapse of the pipeline network. Unlike an imbalance in a power system, it can be corrected through a market mechanism.

3.3 The Need for a Managed Spot Market

As discussed in Chapter 2, a market is an environment designed to help buyers and sellers interact and agree on transactions. These interactions progressively lead to an equilibrium in which the price clears the market, that is, the supply is equal to the demand. If electrical energy is to be traded according to this free-market ideal, the equilibrium between the production and the consumption of electrical energy should be set through the direct interaction of buyers and sellers.

In this ideal market, large consumers and retailers purchase electrical energy from generating companies. Like all rational consumers, they have to estimate how much to purchase. To this end, they forecast their consumption or the consumption of their own customers for every market period (hour, half hour or quarter of an hour) before entering into contracts. For their part, generators schedule the production of their units to deliver at the agreed time the energy that they have sold. Each generator clearly tries to minimize the cost of producing this energy. In practice, however, things are not that simple. Neither party can reliably meet its contractual obligations with perfect accuracy. First, the actual demand of a group of consumers is never exactly equal to the value forecasted. Second, unpredictable problems often prevent generating units from delivering the contracted amount of energy. A sudden mechanical or electrical failure may force a unit to shut down or reduce its output. More mundane problems can delay the synchronization of a unit to the system and hence affect the timing of its production of energy.

These errors and unpredictable events introduce gaps between load and generation that must be bridged quickly and precisely to maintain the integrity of the power system. If these gaps between generation and load were to be treated as imbalances between supply and demand and corrected using an open market mechanism, producers and consumers would have to be kept informed of the state of the market (offer, demand, prices) on a second-by-second basis. A sufficiently large number of them would have to be willing to trade on this timescale. They would also have to be able to adjust their production or consumption at any time and at short notice to absorb any credible imbalance. In the current state of the technology, it is difficult to conceive a system capable of transmitting the vast amounts of data required and of recording the thousands of transactions involved. Even if such an information infrastructure could be put in place, it remains to be proven that such a system would be sufficiently fast and reliable to prevent imbalances that might lead to a collapse of the entire power system. Finally, the transaction costs associated with such a system would be prohibitive.

We can therefore conclude that, while a large proportion of the electrical energy can be traded through an unmanaged open market, such a market is unable to maintain the reliability of the power system. A managed spot market that provides a mechanism for balancing load and generation must therefore supersede the open energy market as the time of delivery approaches. Its function is to match residual load and generation by adjusting the production of flexible generators and curtailing the demand of willing consumers. It should also be able to respond to major disruptions caused by the sudden and unforeseen disconnection of large generating units because of unavoidable technical problems. Although the need for managing the spot market stems from technical considerations, this spot market must operate in an economically efficient manner. Being out of balance may be unavoidable for producers and consumers, but it should not be cost-free. To encourage efficient behavior, producers and consumers must pay the true cost of the electrical energy that is bought or sold in the spot market to restore the balance between load and generation.

Once a fair and efficient spot market is in place, electrical energy can be traded like other commodities. In the next section, we will discuss how this trade can be organized. We will then examine in more detail the design of a managed spot market and its interactions with the open electrical energy market.

3.4 Open Electrical Energy Markets

3.4.1 Bilateral trading

As its name implies, bilateral trading involves only two parties: a buyer and a seller. Participants thus enter into contracts without involvement, interference or facilitation from a third party. Depending on the amount of time available and the quantities to be traded, buyers and sellers will resort to different forms of bilateral trading:

Customized long-term contracts The terms of such contracts are flexible since they are negotiated privately to meet the needs and objectives of both parties. They usually involve the sale of large amounts of power (hundreds or thousands of MW) over long periods of time (several months to several years). The large transaction costs associated with the negotiation of such contracts make them worthwhile only when the parties want to buy or sell large amounts of energy.

Trading "over the counter" These transactions involve smaller amounts of energy to be delivered according to a standard profile, that is, a standardized definition of how much energy should be delivered during different periods of the day and week. This form of trading has much lower transaction costs and is used by producers and consumers to refine their position as delivery time approaches.

Electronic trading Participants can enter offers to buy energy and bids to sell energy directly in a computerized marketplace. All market participants can observe the quantities and prices submitted but do not know the identity of the party that submitted each bid or offer. When a party enters a new bid, the software that runs the exchange checks to see if there is a matching offer for the period of delivery of the bid. If it finds an offer whose price is greater than or equal to the price of the bid, a deal is automatically struck and the price and quantity are displayed for all participants to see. If no match is found, the new bid is added to the list of outstanding bids and will remain there until a matching offer is made or the bid is withdrawn or it lapses because the market closes for that period. A similar procedure is used each time a new offer is entered in the system. This form of trading is extremely fast and cheap. A flurry of trading activity often takes place in the minutes and seconds before the closing of the market as generators and retailers fine-tune their position ahead of the delivery period.

The essential characteristic of these three forms of bilateral trading is that the price of each transaction is set independently by the parties involved. There is thus no "official" price. While the details of negotiated long-term contracts are usually kept private, some independent reporting services usually gather information about over-the-counter trading and publish summary information about prices and quantities in a form that does not reveal the identity of the parties involved. This type of market reporting and the display of the last transaction arranged through electronic trading enhance the efficiency of the market by giving all participants a clearer idea of the state and the direction of the market.

3.4.1.1 Example 3.1

Borduria Power trades in the Bordurian electricity market that operates on a bilateral basis. It owns the three generating units whose characteristics are given in the table below. To keep things simple, we have assumed that the marginal cost of these units is constant over their range of operation. Because of their large start-up cost, Borduria Power tries to keep unit A synchronized to the system at all times and to produce as much as possible with unit B during the daytime. The start-up cost of unit C is assumed to be negligible.

Unit	Type	P_{min}	P_{max}	MC
		(MW)	(MW)	($/MWh)
A	Large coal	100	500	10.0
B	Medium coal	50	200	13.0
C	Gas turbine	0	50	17.0

Let us focus on the contractual position of Borduria Power for the period between 2:00 and 3:00 P.M. on 11 June. The table below summarizes the relevant bilateral contracts.

Type	Contract date	Identifier	Buyer	Seller	Amount	Price
					(MWh)	($/MWh)
Long term	10 January	LT1	Cheapo Energy	Borduria Power	200	12.5
Long term	7 February	LT2	Borduria Steel	Borduria Power	250	12.8
Future	3 March	FT1	Quality Electrons	Borduria Power	100	14.0
Future	7 April	FT2	Borduria Power	Perfect Power	30	13.5
Future	10 May	FT3	Cheapo Energy	Borduria Power	50	13.8

Note that Borduria Power has taken advantage of the price fluctuations in the forward market to buy back at a profit some of the energy that it had sold. Toward midmorning on 11 June, Fiona, the trader on duty at Borduria Power, must decide if she wants to adjust this position by trading on the screen-based Bordurian Power Exchange (BPeX).

On the one hand, Borduria Power has contracted to deliver 570 MWh and has a total production capacity of 750 MW available during that hour. On the other hand, her BPeX trading screen displays the following stacks of bids and offers:

11 June 14 : 00 to 15 : 00	**Identifier**	**Amount (MW)**	**Price ($/MWh)**
Bids to sell energy	B5	20	17.50
	B4	25	16.30
	B3	20	14.40
	B2	10	13.90
	B1	25	13.70
Offers to buy energy	O1	20	13.50
	O2	30	13.30
	O3	10	13.25
	O4	30	12.80
	O5	50	12.55

On the basis of her experience with this market, Fiona believes that it is unlikely that the offer prices will increase. Since she still has 130 MW of spare capacity on unit B, she decides to grab offers O1, O2 and O3 before one of her competitors does. These offers are indeed profitable because their price is higher than the marginal cost of unit B. After completing these transactions, Fiona sends revised production instructions for this hour to the power plants. Unit A is to generate at rated power (500 MW), while unit B is to set its output at 130 MW and unit C is to remain on standby.

Shortly before the BPeX closes trading for the period between 14 : 00 and 15 : 00, Fiona receives a phone call from the operator of plant B. He informs her that the plant has developed some unexpected mechanical problems. It will be able to remain on-line until the evening but will not be able to produce more than 80 MW. Fiona quickly realizes that this failure leaves Borduria Power exposed and that she has three options:

(1) Do nothing, leaving Borduria Power short by 50 MWh that would have to be paid for at the spot market price

(2) Make up this deficit by starting up unit C

(3) Try to buy some replacement power on the BPeX.

Since the spot market prices have been rather erratic lately, Fiona is not very keen on remaining unbalanced. She therefore decides to see if she can buy energy on the

BPeX for less than the marginal cost of unit C. Since she last traded on the BPeX, some bids have disappeared and new ones have been entered

11 June 14:00 to 15:00	Identifier	Amount (MW)	Price ($/MWh)
Bids to sell energy	B5	20	17.50
	B4	25	16.30
	B3	20	14.40
	B6	20	14.30
	B8	10	14.10
Offers to buy energy	O4	30	12.80
	O6	25	12.70
	O5	50	12.55

Fiona immediately selects bids B8, B6 and B3 because they allow her to restore the contractual balance of the company for this trading period at a cost that is less than the cost of covering the deficit with unit C. On balance, when trading closes for this hour, Borduria Power is committed to producing 580 MWh. Note that Fiona based all her decision on the incremental cost of producing energy. We will revisit this example below, once we have discussed the operation of the spot market.

3.4.2 Electricity pools

In the early days of the introduction of competition in electrical energy trading, bilateral trading was seen as too big a departure from the existing practice. Since electrical energy is pooled as it flows from the generators to the loads, it was felt that trading might as well be done in a centralized manner and involve all producers and consumers. Competitive electricity pools were thus created. Pools are a very unusual form of commodity trading but they have well-established roots in the operation of large power systems. In fact, some of the competitive electricity pools currently in operation were developed on the basis of collaborative pools created by monopoly utility companies with adjacent service territories.

Rather than relying on repeated interactions between suppliers and consumers to reach the market equilibrium, a pool provides a mechanism for determining this equilibrium in a systematic way. While there are many possible variations, a pool essentially operates as follows:

- Generating companies submit bids to supply a certain amount of electrical energy at a certain price for the period under consideration. These bids are ranked in order of increasing price. From this ranking, a curve showing the bid price as a function of the cumulative bid quantity can be built. This curve is deemed to be the supply curve of the market.

- Similarly, the demand curve of the market can be established by asking consumers to submit offers specifying quantity and price and ranking these offers in decreasing order of price. Since the demand for electricity is highly inelastic, this step is sometimes omitted and the demand is set at a value determined using a forecast of the load. In other words, the demand curve is assumed to be a vertical line at the value of the load forecast.
- The intersection of these "constructed" supply and demand curves represents the market equilibrium. All the bids submitted at a price lower than or equal to the market clearing price are accepted and generators are instructed to produce the amount of energy corresponding to their accepted bids. Similarly, all the offers submitted at a price greater than or equal to the market clearing price are accepted and the consumers are informed of the amount of energy that they are allowed to draw from the system.
- The market clearing price represents the price of one additional megawatt-hour of energy and is therefore called the system marginal price or SMP. Generators are paid this SMP for every megawatt-hour that they produce, whereas consumers pay the SMP for every megawatt-hour that they consume, irrespective of the bids and offers that they submitted.

Paying the SMP for all the generation that was accepted may appear surprising at first glance. Why shouldn't generators that were willing to produce for less be paid only their asking price? Wouldn't this approach reduce the average price of electricity? The main reason this pay-as-bid scheme is not adopted is that it would discourage generators from submitting bids that reflect their marginal cost of production. All generators would instead try to guess what the SMP is likely to be and would bid at that level to collect the maximum revenues. At best, the SMP would therefore remain unchanged. Inevitably, some low-cost generators would occasionally overestimate the value of SMP and bid too high. These generators would then be left out of the schedule and be replaced by generators with a higher marginal cost of production. The SMP would then be somewhat higher than it ought to be. Furthermore, this substitution is economically inefficient because optimal use is not made of the available resources. In addition, generators are likely to increase their prices slightly to compensate themselves for the additional risk of losing revenue because of the uncertainty on the SMP. An attempt to reduce the price of electricity would therefore result in a price increase!

3.4.2.1 Example 3.2

The electricity pool of Syldavia has received the bids and offers shown on the table below for the period between 9:00 and 10:00 A.M. on 11 June.

Bids	**Company**	**Quantity (MWh)**	**Price ($/MWh)**
	Red	200	12.00
	Red	50	15.00
	Red	50	20.00
	Green	150	16.00
	Green	50	17.00
	Blue	100	13.00
	Blue	50	18.00
Offers	Yellow	50	13.00
	Yellow	100	23.00
	Purple	50	11.00
	Purple	150	22.00
	Orange	50	10.0
	Orange	200	25.00

Figure 3.1 shows how these bids and offers stack up to form the supply and demand curves respectively. From the intersection of these two curves, we conclude that for this particular period the SMP will be set at 16.00 $/MWh and that 450 MWh will be traded through the Syldavian electricity pool. The table below shows how much energy each generator will be instructed to produce and how much energy each consumer will be allowed to draw. It also shows the revenues and expenses for each company

Company	**Production (MWh)**	**Consumption (MWh)**	**Revenue ($)**	**Expense ($)**
Red	250		4000	
Blue	100		1600	
Green	100		1600	
Orange		200		3200
Yellow		100		1600
Purple		150		2400
Total	450	450	7200	7200

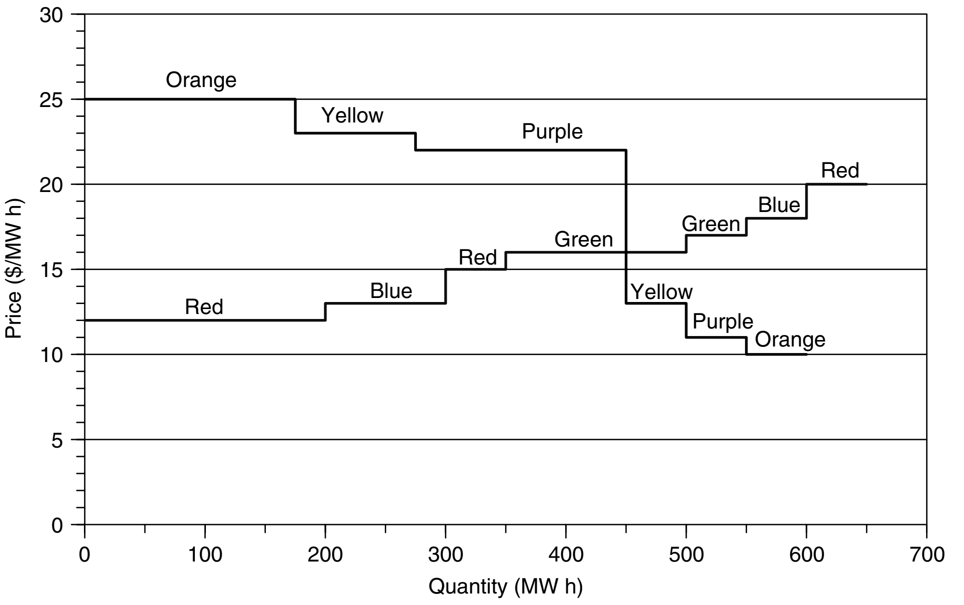


Figure 3.1 Stacks of bids and offers of Example 3.2

If, instead of asking consumers to submit offers, the Syldavian pool relied on a load forecast to represent the demand side and if the load for this period was forecasted to be 450 MWh, we would obtain exactly the same results.

In this example, generators submit simple bids consisting of price/quantity pairs. In some pools, generators submit complex bids for each of their generating units. These bids are supposed to reflect the cost characteristics of the unit (including the marginal, start-up and no-load costs) as well as some technical parameters (minimum and maximum output, flexibility). Rather than simply stacking the bids, the pool then performs a unit commitment calculation that determines the production schedule and the prices for an entire day divided in periods of half an hour or an hour. This approach to scheduling and pricing was used between 1990 and 2001 in the Electricity Pool of England and Wales.

3.4.3 Comparison of pool and bilateral trading

Since both the pool and the bilateral models of electrical energy trading have been adopted for electricity markets, it is worth reviewing the perceived advantages and disadvantages of both approaches.

As mentioned above, a competitive electricity pool is often created on the basis of an existing cooperation agreement between various utilities. Its conversion to operation on a competitive basis will therefore be less of a revolution than the creation of a completely new structure. Some of the concerns that accompany the introduction of competition may be alleviated by the somewhat less radical nature of the change. In particular, the public and the government are likely to have fewer concerns about the security of the electricity supply if the same organization remains in charge. A pool provides a much more centralized form of system management. Not only does it handle all the physical electrical energy transactions but it usually also assumes the responsibility for operating the transmission system. This combination of roles avoids the multiplication of organizations but makes it more difficult to distinguish between the various functions that need to be performed in an electricity market.

Most small and medium electricity consumers have very little incentive to take an active part in an electricity market. Even when they are aggregated, the retailer that represents them has no direct means of adjusting consumption in response to changes in prices. One might therefore argue that the transaction costs could be reduced significantly if the demand is deemed to be passive and is represented by a load forecast in an electricity pool. Many economists are unhappy with this approach because they feel that direct negotiations between consumers and producers are essential if efficient prices are to be reached. Some economists dislike pools simply because they are only administered approximations of a market and not true markets.

Pools also provide a mechanism for reducing the scheduling risk faced by generators and hence, hopefully, the cost of electrical energy. When a generator sells energy on the basis of simple bids, for each market period separately, it runs the risk that for some periods it may not have sold enough energy to keep the plant on-line. At that point, it must decide whether to sell energy at a loss to keep the unit running or to shut it down and face the expense of another start-up at a later time. Either option increases the cost of producing energy with this unit and forces the generator to raise its average bid price. If this generator trades in a pool that operates on the basis of complex bids, the rules of this pool probably ensure that it recovers the start-up and no-load components of its bid. Moreover, the scheduling algorithm implemented by the pool usually tries to avoid unnecessary shutdowns. Since these factors reduce the risks faced by the generators, one would expect that they should foster lower average prices. This reduction in risk, however, comes at the price of an increase in the complexity of the pool rules. More complex rules reduce the transparency of the price setting process and increase opportunities for price manipulations. In practice, it is not clear whether complex bids and pool-based scheduling actually lower electricity prices.

3.5 The Managed Spot Market

For every commodity, imbalances almost always arise between the amount that a party has contracted to buy or sell and the amount that it actually needs or can produce. Spot markets provide a mechanism for handling these imbalances. If electrical energy is to be treated as a commodity, a spot market must therefore be organized. Unfortunately, as we discussed above, imbalances between generation and load must be corrected so quickly that a conventional spot market mechanism is not feasible. Instead, the system operator (SO) is given the responsibility to maintain the system in balance using what one might call a "managed spot market". This mechanism is a market because the energy that is used to achieve this balance is freely offered by the participants at a price of their own choosing. It is a spot market because it determines the price at which imbalances are settled. However, it is a managed market because the bids and offers are selected by a third party (the SO) rather than through bilateral deals.

In the next paragraphs, we discuss the functionality of a generic managed spot market for electricity. Actual implementations can differ substantially from this blueprint.

There is also no consensus on the name to be given to this function. In addition to the term "spot market", names such as "reserve market", "balancing mechanism" and others are used.

3.5.1 Obtaining balancing resources

If market participants were able to predict with enough lead time and with perfect accuracy the amount of energy that they will consume or produce, the SO would not have to take balancing actions. The participants themselves could trade to cover their deficits and absorb their surpluses. In practice, there are always small imbalances and the SO must obtain adjustments in generation or load. Integrated over time, these adjustments translate into purchases and sales of electrical energy that can be settled at a spot price reflecting the market's willingness to provide these adjustments. In keeping with the free market philosophy, any party that is willing to adjust its production or consumption should be allowed to do so on a competitive basis. This should give the SO the widest possible choice of balancing options and should therefore help reduce the cost of balancing. These balancing resources can be offered either for a specific period or on a long-term basis. Balancing services for a specific period are normally offered by market participants to the SO after the open energy market for that period has closed. Generating units that are not fully loaded can submit bids to increase their output. A generating unit can also offer to pay to reduce its output. This is a profitable proposition if the incremental price of this offer is less than the incremental cost of producing energy with this unit. A generating unit that submits such an offer is, in effect, trying to replace its own generation by cheaper power purchased on the spot market.

The demand side can also provide balancing resources. A consumer would offer to reduce its consumption if the price is greater than the value it places on consuming electricity during that period. Such demand reductions have the advantage that they can be implemented very quickly. It is also conceivable that consumers might offer to increase their demand if the price is sufficiently low.

Since these offers of balancing resources are submitted shortly before real time, the SO may be concerned about the amount or the price of balancing resources that will be offered. To guard against such problems, the SO can purchase balancing resources on a long-term basis. Under such contracts, the supplier is paid a fixed price (often called the option fee) to keep available some generation capacity. The contract also specifies the price or exercise fee to be paid for each MWh produced at the request of the SO using this capacity. The SO would call upon this contract only if the exercise fee is lower than what it would have to pay for a similar balancing resource offered on a short-term basis. As the terminology suggests, these contracts are equivalent to the option contracts used in financial and commodity markets. Their purpose is the same: to protect the buyer (in this case the SO) against price increases while guaranteeing some revenues to the supplier.

Imbalances due to forecasting errors by the participants are relatively small, evolve gradually and can be predicted to a certain extent. On the other hand, the imbalances that are caused by failures are often large, unpredictable and sudden. Many generating units can adjust their output at a rate that is sufficient to cope with the first type of imbalances. Handling the second type of imbalances requires generating units that can increase their output rapidly and sustain this increased output for a certain time. We will explore the issue of reserve generation capacity in more detail when we discuss system security in Chapter 5. In the meantime, it is important to realize that all the units of energy that are traded to keep the system in balance do not have the same value. A megawatt obtained by slightly increasing the output of a large thermal plant costs considerably less than a megawatt of load that must be shed to prevent the system from collapsing. To be able to keep the system in balance at minimum cost, the SO should therefore have access to a variety of balancing resources. When producers and consumers bid to supply balancing resources, their bids must specify not only a quantity and a price but also how quickly the change in power injection can be delivered.

3.5.2 Gate closure

As we argued above, energy trading must stop at some point before real time to give the SO time to balance the system. How much time should elapse between this gate closure and real time is a hotly debated issue. System operators prefer longer intervals because this gives them more time to develop their plans and more flexibility in their selection of balancing resources. For example, if the gate closes half an hour before real time, there is not enough time to bring on-line a large coal-fired plant to make up a deficit in generation. Participants in the energy market, on the other hand, usually prefer a shorter gate closure because it reduces their exposure to risk. A load forecast calculated one hour ahead of real time is usually much more accurate than a forecast calculated four hours ahead. A retailer would therefore like to trade electronically up to the last minute to match its purchases with its anticipated load. This is considered preferable to relying on the managed spot market in which it is exposed to prices over which it has no control. Generators too prefer shorter gate closures because of the risk of sudden unit outage. If a unit fails after gate closure, there is nothing that the generator can do except hope that the spot market price will not be too high. On the other hand, if the unit fails before gate closure, the generator can try to make up the deficit in generation by purchasing at the best possible price on the electronic exchange. In general, traders prefer a true spot market that is driven solely by market forces to a managed spot market that is heavily influenced by complex technical considerations.

3.5.3 Operation of the managed spot market

Figure 3.2 summarizes the operation of a managed spot market. At gate closure, the producers and the consumers must inform the SO of their contractual positions, that is, how much power they intend to produce or consume during the period under consideration. The SO combines that information with its own forecast of the total load to determine by how much the system is likely to be in imbalance. If generation exceeds the load, the system is said to be long. If the opposite holds, the system is short. The SO must then decide which balancing bids and offers it will use to cover the imbalances.

When an electricity market operates on the basis of a pool run by the system operator, the balancing function will often be so closely integrated with the energy market function that they may be difficult to separate.

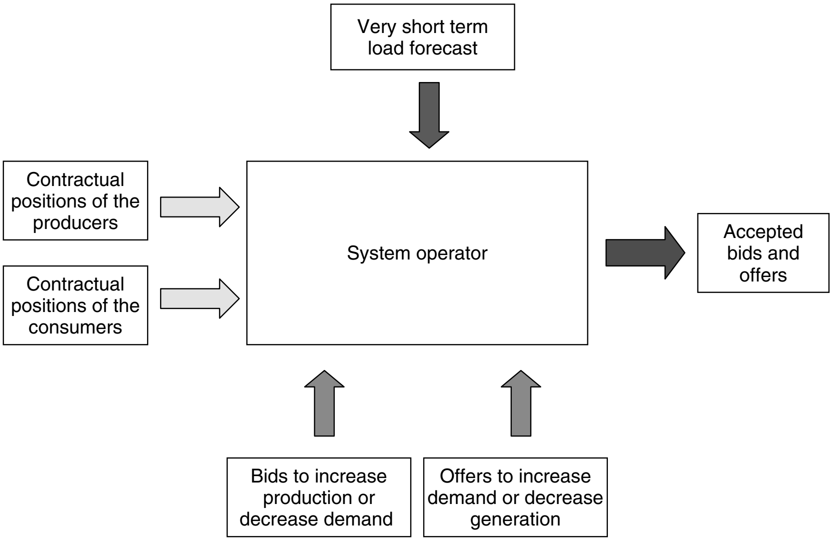


Figure 3.2 Schematic diagram of the operation of a managed spot market for electricity

3.5.3.1 Example 3.3

Let us revisit our trader for Borduria Power from Example 3.1 and see how she handles the spot market. When the gate closes for bilateral trading, Fiona has contracted to produce a net amount of 580 MWh for the period under consideration. She informs the system operator that her company intends to produce this amount as follows:

Unit	Scheduled production (MW)
A	500
B	80
C	0

Fiona then has to decide what bids and offers she wants to make in the managed spot market. To help her in this decision, she considers the scheduled output and the characteristics of Borduria Power's generating units:

Unit	P^{sched} (MW)	P^{min} (MW)	P^{max} (MW)	MC ($/MWh)
A	500	100	500	10.0
B	80	50	80	13.0
C	0	0	50	17.0

The only bid to increase generation that she can offer involves unit C because units A and B are scheduled to produce their maximum output. Such a bid would be for a maximum of 50 MWh and would have to be priced at 17.00 $/MWh or above to be profitable, if we assume that the start-up cost of unit C is negligible.

Fiona also considers the possibility of reducing the production of units A and B. She would be willing to pay up to 10 $/MWh to reduce the output of unit A and up to 13 $/MWh to reduce the output of unit B because that is the marginal cost of producing power with these units. The output of these units can be reduced by 400 MW and 30 MW respectively without affecting the plans for the following periods, if we assume that there are no restrictions on the ramp rate of the units. Further reductions would require their shutdown and might then preclude Borduria Power from meeting its commitments for later hours. Furthermore, the cost of restarting these units would reduce their profitability.

In Chapter 2, we argued that, in a perfectly competitive market, the optimal strategy of each participant is to bid its marginal cost or to offer its marginal value. As we will discuss in a subsequent section, electricity markets are usually not perfectly competitive. Some participants can increase their profits by bidding above their marginal costs or offering below their marginal value. On the basis of her experience, Fiona decides that the following bids and offers are likely to maximize Borduria Power's profit:

Type	Identifier	Price ($/MWh)	Amount (MW)
Bid	SMB-1	17.50	50
Offer	SMO-1	12.50	30
Offer	SMO-2	9.50	400

We will revisit this example one more time after discussing the settlement process.

3.5.4 Interactions between the managed spot market and the other markets

Since the managed spot market is the market of last resort for electrical energy, it has a strong influence on the other markets. If the spot price tends to be low, purchasers of energy will not be unduly concerned about being short because they can make up their deficits in the spot market at a reasonable price. They might therefore buy slightly less than they need in the forward market and will thus push down the price of energy in this market. On the other hand, if the spot price tends to be high, these same purchasers will push up the price in the forward market as they buy more to make sure that they cover all their needs at the better price. If electricity was a simple commodity, such discrepancies should even themselves out over time and the forward prices should reflect the expected value of the spot price.

Electricity is certainly not the only commodity whose spot price is highly volatile. A weather forecast predicting frost in the coffee-producing regions of Brazil will send the spot price of coffee soaring. This price may very well come crashing down the next day if the forecast turns out to be inaccurate or the damage to the crop is more limited than was feared. The difference between coffee and electrical energy is that coffee traded in the spot market is produced exactly the same way as coffee sold under a long-term contract. On the other hand, an MWh sold through the managed spot market often will have been produced by a plant that is much more flexible than the plants that generate the bulk of the energy consumed at that time.

Ensuring the smooth and secure operation of a power system requires the performance of ancillary services such as load following and frequency response as well as the provision of spinning reserve capacity. Providing these services is another way for generators and consumers to increase their revenues. A few generating plants that are very flexible but find it hard to compete in the energy market usually occupy this market niche. The provision of these services translates into the delivery of a relatively small amount of electrical energy. If the providers of these services are remunerated on the basis of the megawatt-hour produced, they will have to charge a fairly high price per megawatt-hour to collect enough revenue to stay in business. Including the cost of this energy in the calculation of the spot price will often result in sharp price spikes. These price spikes do not reflect a sudden deficit of electrical energy in the market. They are a consequence of an acute but temporary lack of liquidity. Price spikes occur because, for a short time, the number of participants able to provide this "service energy" is very small and because consumers are not able or not willing to reduce their demand at short notice. Price spikes represent a risk for companies that are forced to buy from the managed spot market and will encourage them to purchase more in the forward markets and hence drive up the forward prices. These forward prices are thus artificially inflated by the need to generate a small fraction of the total energy demand at short notice. Chapter 5 discusses in more detail the provision of ancillary services and alternative methods of remunerating the participants that provide them.

3.6 The Settlement Process

Commercial transactions are normally settled directly between the two parties involved: following the delivery of the goods by the seller to the buyer, the buyer pays the seller the agreed price. If the amount delivered is less than the amount contracted, the buyer is entitled to withhold part of the payment. Similarly, if the buyer consumes more than the agreed amount, the seller is entitled to an additional payment. This process is more complex for electricity markets because the energy is pooled during its transmission from the producers to the consumers. This is why a centralized settlement system is needed.

For bilateral transactions in electrical energy, the buyer pays the seller the agreed price as if the agreed quantity had been delivered exactly. Similarly, the anonymous transactions arranged through screen-based trading are settled through the intermediary of the power exchange as if they had been executed perfectly. However, there will always be inaccuracies in the completion of the contracts. If a generator fails to produce the amount of energy that it has contracted to sell, the deficit cannot simply be withheld from this generator's customers. Instead, to maintain the stability of the system, the system operator buys replacement energy on the managed spot market. Similarly, if a large user or retailer consumes less than it has bought, the system operator sells the excess on the managed spot market. These balancing activities make all bilateral contracts look as if they have been fulfilled perfectly. They also carry a cost. In most cases, the amount of money paid by the system operator to purchase replacement energy is not equal to the amount of money earned when selling excess energy. The parties that are responsible for the imbalances should pay the cost of these balancing activities.

The first step in the settlement process consists, therefore, in determining the net position of every market participant. To this end, each generator must report to the settlement system the net amount of energy that it had contracted to sell for each period, including the energy traded through the managed spot market. This amount is subtracted from the amount of energy that it actually produced. If the result is positive, the generator is deemed to have sold this excess energy to the system. On the other hand, if the result is negative, the generator is treated as if it had bought the difference from the system.

Similarly, all large consumers and retailers must report the net amount of energy that they had contracted to buy for each period, including the energy traded through the managed spot market. This amount is subtracted from the amount of energy actually consumed. Depending on the sign of the result, the consumer or the retailer is deemed to have sold energy to the system or bought energy from the system.

These imbalances are charged at the spot market price. If this market is suitably competitive, this price should reflect the incremental cost of balancing energy. As we discussed in the previous section and as we will investigate further in Chapter 5, it is debatable whether the cost of the energy supplied by participants providing ancillary services should be included in this price.

Settlement in a pool-based electricity market is more straightforward because all transactions take place through the pool.

3.6.1.1 Example 3.4

In Examples 3.1 and 3.3, we looked at the trading activities of Borduria Power for the period from 2:00 to 3:00 P.M. on 11 June in the bilateral market and the managed spot market. Let us assume that the following events took place after gate closure:

- Faced with a deficit of generation, the System Operator called 40 MWh of Borduria Power's spot market bid at 17.50 $/MWh (SMB-1).
- The troubles with Borduria Power's unit B turned out to be worse than anticipated, forcing its complete shutdown soon after the beginning of the period. It was only able to produce 10 of the 80 MWh that it was scheduled to produce, leaving Borduria Power with a deficit of 70 MWh.
- The spot price of electrical energy was 18.25 $/MWh for this period.

The following table shows the details of the flows of money in and out of Borduria Power's trading account.

Market	Identifier	Amount (MWh)	Price ($/MWh)	Income ($)	Expense ($)
Futures and forwards	Cheapo Energy	200	12.50	2500.00	
	Borduria Steel	250	12.80	3200.00	
	Quality Electrons	100	14.00	1400.00	
	Perfect Power	−30	13.50		405.00
	Cheapo Energy	50	13.80	690.00	
Power exchange	O1	20	13.50	270.00	
	O2	30	13.30	399.00	
	O3	10	13.25	132.50	
	B3	−20	14.40		288.00
	B6	−20	14.30		286.00
	B8	−10	14.10		141.00
Spot market	SMB-1	40	17.50	700.00	
	Imbalance	−70	18.25		1277.50
Total		550		9291.50	2397.50

Bilateral trades are settled directly between Borduria Power and its counterparties. Since trades on the power exchange are anonymous, they are settled through BPeX (the company running the power exchange). Finally, activity on the managed spot market (both voluntary and compulsory) is settled through the system operator or its settlement agent. The bottom line of this table indicates that the trading revenue of Borduria Power for this period amount to $6894.00. To determine whether this trading period was profitable, we would have to compute the cost of producing the energy that Borduria Power delivered. Carrying out this computation for a single trading period is rather difficult because there is no simple way to allocate the start-up and no-load costs of the generating units.

3.7 Further Reading

The theory of spot pricing was first applied to electrical energy by Schweppe *et al.* (1988) and is described in detail in Schweppe's book that is often credited with providing the theoretical underpinning for the introduction of competition in power systems. The book by Stoft (2002) discusses the design of electricity markets in much more detail than here. A considerable amount of material on the organization of electricity markets is available from regulatory bodies such as the Federal Energy Regulatory Commission (FERC) in the United States and the Office of Gas and Electricity Markets (OFGEM) in the United Kingdom or from market operators such as PJM.

FERC, *http://www.ferc.gov*.

OFGEM, *http://www.ofgem.gov.uk/public/adownloads.htm#retabm*.
PJM, *http://www.pjm.com*.
Schweppe F C, Caramanis M C, Tabors R D, Bohn R E, *Spot Pricing of Electricity*, Kluwer Academic Publishers, Boston, MA 1988.
Stoft S, *Power System Economics*, John Wiley & Sons, 2002.

3.8 Problems

3.1 Choose an electricity market about which you have access to sufficient information, preferably the same market that you studied for the problems of Chapter 1. Describe the implementation of this market. In particular, determine the aspects that are based on bilateral trading and those that are centrally operated. Discuss the mechanism used to set prices in the managed spot market.

3.2 The rules of the Syldavian electricity market stipulate that all participants must trade energy exclusively through the Power Pool. However, the Syldavia Aluminum Company (SALCo) and the Northern Syldavia Power Company (NSPCo) have signed a contract for difference for the delivery of 200 MW on a continuous basis at a strike price of 16 $/MWh.

a. Trace the flow of power and money between these companies when the pool price takes the following values: 16 $/MWh, 18 $/MWh and 13 $/MWh.

b. What happens if during one hour the Northern Syldavia Power Company is able to deliver only 50 MWh and the pool price is 18 $/MWh?

c. What happens if during one hour the Syldavia Aluminum Company consumes only 100 MWh and the pool price is 13 $/MWh?

3.3 The following six companies participate, along with others, in the Southern Antarctica electrical energy market:

- **Red:** A generating company owning a portfolio of plants with a maximum capacity of 1000 MW.
- **Green:** Another generating company with a portfolio of plants with a maximum capacity of 800 MW.
- **Blue:** A retailer of electrical energy.
- **Yellow:** Another retailer of electrical energy.
- **Magenta:** A trading company with no generating assets and no demand.
- **Purple:** Another trading company with no physical assets.

The following information pertains to the operation of this market for Monday 29 February 2016 between 1:00 and 2:00 P.M.

Load forecasts Blue and Yellow forecast that their customers will consume 1200 MW and 900 MW respectively during that hour.

Long-term contracts June 2015: Red signs a contract for the supply of 600 MW at 15 $/MWh for all hours between 1 January 2015 and 31 December 2020

July 2015: Blue signs a contract for the purchase of 700 MW for all hours between 1 February 2016 and 31 December 2016. The price is set at 12 $/MWh for off-peak hours and at 15.50 $/MWh for peak hours.

August 2015: Green signs a contract for the supply of 500 MW at 16 $/MWh for peak hours in February 2016.

September 2015: Yellow signs a contract for the purchase of electrical energy. The contract specifies a profile of daily and weekly volumes and a profile for daily and weekly prices. In particular, on weekdays between 1 : 00 and 2 : 00 P.M., the volume purchased is 550 MW at 16.25 $/MWh.

Futures contracts All contracts are for delivery on 29 February 2016 between 1 : 00 and 2 : 00 P.M.

Date	**Company**	**Type**	**Amount**	**Price**
10/9/15	Magenta	Buy	50	14.50
20/9/15	Purple	Sell	100	14.75
30/9/15	Yellow	Buy	200	15.00
10/10/15	Magenta	Buy	100	15.00
20/10/15	Red	Sell	200	14.75
30/10/15	Green	Sell	250	15.75
30/10/15	Blue	Buy	250	15.75
10/11/15	Purple	Buy	50	15.00
15/11/15	Magenta	Sell	100	15.25
20/11/15	Yellow	Buy	200	14.75
30/11/15	Blue	Buy	300	15.00
10/12/15	Red	Sell	200	16.00
15/12/15	Red	Sell	200	15.50
20/12/15	Blue	Sell	50	15.50
15/1/16	Purple	Sell	200	14.50
20/1/16	Magenta	Buy	50	14.25
10/2/16	Yellow	Buy	50	14.50
20/2/16	Red	Buy	200	16.00

25/2/16	Magenta	Sell	100	17.00
28/2/16	Purple	Buy	250	14.00
28/2/16	Yellow	Sell	100	14.00

Options contracts In November 2015, Red bought a put option for 200 MWh at 14.75 $/MWh. The option fee was $50.

In December 2015, Yellow bought a call option for 100 MWh at 15.50 $/MWh. The option fee was $25.

Outcome

- The spot price on the Southern Antarctica electricity market was set at 15.75 $/MWh for 29 February 2016 between 1 : 00 and 2 : 00 P.M.
- Owing to the difficulties at one of its major plants, Red was able to generate only 800 MW. Its average cost of production was 14.00 $/MWh.
- Green generated 770 MW at an average cost of 14.25 $/MWh.
- Blue's demand turned out to be 1250 MW. Its average retail price was 16.50 $/MWh.
- Yellow's demand turned out to be 850 MW. Its average retail price was 16.40 $/MWh.

Assuming that all imbalances are settled at the spot market price, calculate the profit or loss made by each of these participants.

3.4 The operator of a centralized market for electrical energy has received the bids shown in the table below for the supply of electrical energy during a given period.

Company	**Amount (MWh)**	**Price ($/MWh)**
Red	100	12.5
Red	100	14.0
Red	50	18.0
Blue	200	10.5
Blue	200	13.0
Blue	100	15.0
Green	50	13.5
Green	50	14.5
Green	50	15.5

a. Build the supply curve

b. Assume that this market operates unilaterally, that is, that the demand does not bid and is represented by a forecast. Calculate the market price, the quantity produced by each company and the revenue of each company for each of the following loads: 400 MW, 600 MW, 875 MW.

c. Suppose that instead of being treated as constant, the load is represented by its inverse demand curve, which is assumed to have the following form:

$$D = L - 4.0 \cdot \pi$$

where D is the demand, L is the forecasted load and π is the price. Calculate the effect that this price sensitivity of demand has on the market price and the quantity traded.

3.5 The Syldavian Power and Light Company owns one generating plant and serves some load. It has been actively trading in the electricity market and has established the following position for 11 June between 10 : 00 and 11 : 00 A.M.:

- Long-term contract for the purchase of 600 MW during peak hours at a price of 20.00 $/MWh
- Long-term contract for the purchase of 400 MW during off-peak hours at a price of 16.00 $/MWh
- Long-term contract with a major industrial user for the sale of 50 MW at a flat rate of 19.00 $/MWh
- The remaining customers purchase their electricity at a tariff of 21.75 $/MWh
- Future contract for the sale of 200 MWh at 21.00 $/MWh
- Future contract for the purchase of 100 MWh at 22.00 $/MWh
- Call option for 150 MWh at an exercise price of 20.50 $/MWh
- Put option for 200 MWh at an exercise price of 23.50 $/MWh.
- Call option for 300 MWh at an exercise price of 24.00 $/MWh.

The option fee for all the options is 1.00 $/MWh. The peak hours are defined as being the hours between 8 : 00 A.M. and 8 : 00 P.M.
The outcome for 11 June between 10 : 00 and 11 : 00 is as follows:

- The spot price is set at 21.50 $/MWh.
- The total load of the Syldavian Power and Light Company is 1200 MW, including the large industrial customer.
- The power plant produces 300 MWh at an average cost of 21.25 $/MWh.

a. Assuming that all imbalances are settled at the spot market price, calculate the profit or loss made by the company during that hour.

b. What value of the spot market would reduce the profit or loss of the company to zero? Would this change in spot price affect any of the option contracts?

3.6 A company called Borduria Energy owns a nuclear power plant and a gas-fired power plant. Its trading division has entered into the following contracts for 25 January:

T-1. A forward contract for the sale of 50 MW at a price of 21.00 $/MWh. This contract applies to all hours.
T-2. A long-term contract for the sale of 300 MW during off-peak hours at a price of 14.00 $/MWh
T-3. A long-term contract for the sale of 350 MW at 20 $/MWh during peak hours.

In addition, for the trading period from 2 : 00 to 3 : 00 P.M. on that day, it has entered into the following transactions:

T-4. A future contract for the purchase of 600 MWh at 20.00 $/MWh
T-5. A future contract for the sale of 100 MWh at 22.00 $/MWh
T-6. A put option for 250 MWh at an exercise price of 23.50 $/MWh
T-7. A call option for 200 MWh at an exercise price of 22.50 $/MWh
T-8. A put option for 100 MWh at an exercise price of 18.75 $/MWh
T-9. A bid in the spot market to produce 50 MW using its gas-fired plant at 19.00 $/MWh
T-10. A bid in the spot market to produce 100 MW using its gas-fired plant at 22.00 $/MWh

The option fee for all call and put options is $2.00/MWh. The peak hours are defined as being the hours between 8 : 00 A.M. and 8 : 00 P.M.

Borduria Energy also sells electrical energy directly to small consumers through its retail division. Residential customers pay a tariff of 25.50 $/MWh

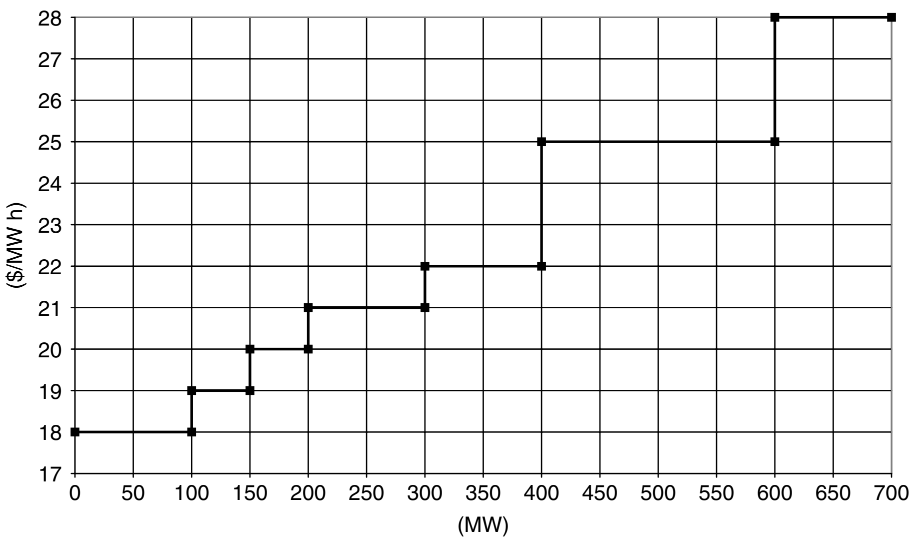


Figure 3.3 Stack of bids for Problem 3.6

and commercial consumers pay a tariff of 25.00 \$/MWh. Borduria Energy does not sell electricity to industrial consumers.

The graph on Figure 3.3 shows the stack of bids that the spot market operator has received for the trading period from 2 : 00 to 3 : 00 P.M. on 25 January. In order to balance load and generation, it accepted bids for 225 MW in increasing order of price for that hour. The spot price was set at the price of the last accepted bid.

During that hour, the residential customers served by Borduria Energy consumed 300 MW, while its commercial customers consumed 200 MW. The nuclear power plant produced 400 MWh at an average cost of 16.00 \$/MWh. Its gas-fired plant produced 200 MWh at an average cost of 18.00 \$/MWh. All imbalances are settled at the spot market price.

a. Calculate the profit or loss made by Borduria Energy during that hour.

b. Calculate the effect that the sudden outage of the nuclear generating plant at 2 : 00 P.M. on 25 January would have on the profit (or loss) of Borduria Energy for that hour.

4 Participating in Markets for Electrical Energy

4.1 Introduction

In the previous chapter, we discussed the basic principles of markets for electrical energy and illustrated through some examples how market participants interact with such markets. In this chapter, we discuss in more detail the decisions that generators, consumers and others take to optimize the benefits that they derive from these markets.

We will first discuss why consumers have a much more passive role than producers do in electricity markets and how retailers serve as their intermediaries in the market for electrical energy.

We will then adopt the perspective of a generating company and consider the case in which this company faces a perfectly competitive market. In such a market, since the company's actions do not affect the prices, it can optimize its activities independently of what other producers or consumers might do. Such a scenario is unrealistic in the context of electricity markets because the short-term elasticity of the demand for electricity is very low and because in most markets the bulk of the electrical energy is produced by a small number of producers. We will therefore discuss some of the techniques that have been proposed to analyze the operation of imperfectly competitive markets and to maximize a producer's profit in such markets.

Finally, we will explore how storage facilities and other hybrid participants can make a profit from trading electrical energy.

4.2 The Consumer's Perspective

Microeconomic theory suggests that consumers of electricity, like consumers of all other commodities, increase their demand up to the point at which the marginal benefit they derive from the electricity is equal to the price they have to pay. For example, a manufacturer will not produce widgets if the cost of the electrical energy required to produce them makes their sale unprofitable. Similarly, the owner of a fashion boutique will increase the lighting level only up to the point at which it attracts more customers. Finally, at home during a cold winter evening, there comes a point at which most of us

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

will put on some extra clothes rather than turning up the thermostat and face a very large electricity bill. Since this chapter deals only with the short-term behavior of consumers, we do not consider the possibility that they will purchase new appliances, machinery or other facilities that would allow them to change their pattern of consumption.

If these industrial, commercial and residential customers pay a flat rate for each kilowatt-hour that they consume, they are insulated from the spot price of electricity and their demand is affected only by the cycle of their activities. Averaged over a few weeks or months, their demand reflects only their willingness to pay this flat rate. But what happens when the price of electrical energy fluctuates more rapidly? Empirical evidence suggests that demand does decrease in response to a short-term price increase, but that this effect is relatively small. In other words, the price elasticity of the demand for electricity is small. On a price versus quantity diagram, the slope of the demand curve is therefore very steep. Determining the shape of the demand curve with any kind of accuracy is practically impossible for a commodity like electrical energy. It is, however, interesting to compare the range of prices for electrical energy sold on a competitive market such as the Electricity Pool of England and Wales (see Table 4.1) with a measure of the value that consumers place on the availability of electrical energy. One such measure is the value of lost load (VOLL), which is obtained through surveys of consumers and represents the average price per megawatt-hour that consumers would be willing to pay to avoid being disconnected without notice. For the same period as the data shown in Table 4.1, the VOLL used in England and Wales was set at 2768 £/MWh.

Two economic and social factors explain this weak elasticity. First, the cost of electrical energy makes up only a small portion of the total cost of producing most industrial goods and represents only a small fraction of the cost of living for most households. At the same time, electricity is indispensable in manufacturing and most individuals in the industrialized world regard it as essential to their quality of life. Most industrial consumers therefore will not reduce their production drastically to avoid a small increase in their electricity costs. In the short term, the savings might be more than offset by the loss of revenue. Similarly, most residential consumers will probably not reduce their comfort and convenience to cut their electricity bill by a few percents. The second factor explaining this weak elasticity is historical. Since the early days of commercial electricity generation over a century ago, electricity has been marketed as a commodity that is easy to use and always available. This convenience has become so ingrained that it is fair to say that very few people carry out a cost/benefit analysis each time they turn on the light!

Rather than simply reducing their demand in response to a sudden increase in the price of electrical energy, consumers instead may decide to delay this demand until a time when prices are lower. For example, a manufacturer may decide to delay the

Table 4.1 Pool selling price on the electricity pool of England and Wales (in £/MWh)

	Minimum	Maximum	Average
January 2001	0.00	168.49	21.58
February 2001	10.00	58.84	18.96
March 2001	8.00	96.99	20.00

(Source: The Electricity Pool of England and Wales).

completion of a particularly energy-intensive step of a production process until the night shift if the price of electrical energy is expected to be lower at that time. Similarly, residential consumers in some countries take advantage of lower nighttime tariffs by waiting until later in the evening to wash and dry clothes or heat water. Shifting demand is possible only if the consumer is able to store intermediate products, heat, electrical energy, or dirty clothes. Such storage facilities and the associated control equipment carry a significant investment cost. The savings that can be achieved by shifting demand for electricity from periods of high prices to periods of low prices might not justify these costs. Furthermore, managing demand requires more flexibility or more willingness to accept a loss of convenience than may be available.

Most small residential and commercial consumers therefore will not be very interested in reacting to hourly or half-hourly price changes. Even if they were, the cost of the communication infrastructure needed to inform them of these prices and to register their consumption during each period would absorb most, if not all, of the potential benefits. For the foreseeable future, these consumers probably will continue purchasing electrical energy on the basis of a tariff. Such tariffs insulate them from daily fluctuations in prices and therefore reduce to zero their contribution to the overall short-term elasticity of demand.

This very low elasticity of the demand has undesirable effects on the operation of markets for electrical energy. In particular, when we discuss imperfect markets, we will see that it facilitates the exercise of market power by the producers.

4.2.1 Retailers of electrical energy

Consumers whose peak demand is at least a few hundreds kilowatts may be able to save significant amounts of money by employing specialized personnel to forecast their demand and trade in the electricity markets to obtain lower prices. Such consumers can be expected to participate directly and actively in the markets. On the other hand, such active trading is not worthwhile for smaller consumers. These smaller consumers usually prefer purchasing on a tariff, that is, at a constant price per kilowatt-hour that is adjusted at most a few times per year. Electricity retailers are in business to bridge the gap between the wholesale market and these smaller consumers.

The challenge for them is that they have to buy energy at a variable price on the wholesale market and sell it at a fixed price at the retail level. A retailer will typically lose money during periods of high prices because the price it has to pay for energy is higher than the price at which it resells this energy. On the other hand, during periods of low prices it makes a profit because its selling price is higher than its purchase price. To stay in business, the quantity-weighted average price at which a retailer purchases electrical energy should therefore be lower than the rate it charges its customers. This is not always easy to achieve because the retailer does not have direct control over the amount of energy that its customers consume. Each retailer is deemed to have sold to its customers the amount of energy that went through their meters. If for any period the aggregate amount over all its customers exceeds the amount that it has contracted to buy, the retailer has to purchase the difference on the spot market at whatever value the spot price reached for that period. Similarly, if the amount contracted exceeds the amount consumed by its customers, the retailer is deemed to have sold the difference on the spot market.

To reduce its exposure to the risk associated with the unpredictability of the spot market prices, a retailer therefore tries to forecast as accurately as possible the demand of its customers. It then purchases energy on the various markets to match this forecast. A retailer thus has a strong incentive to understand the consumption patterns of its customers. It will often encourage its customers to install meters that record the energy consumed during each period so it can offer them more attractive tariffs if they reduce their energy consumption during peak price hours. By taking into account all the meteorological, astronomical, economic, cultural and special factors that influence the consumption of electricity and using the most sophisticated forecasting techniques available, it is possible to predict the value of the demand at any hour with an average accuracy of about 1.5 to 2%. However, such accuracy is possible only with large groups of consumers in which the aggregation effects reduce the relative importance of the random fluctuations. A retailer who does not have a monopoly on the supply of electricity in a given region can therefore forecast the demand of its customers with considerably less accuracy than what a monopoly utility can achieve. This problem is exacerbated if, as one would expect, customers have the opportunity to change retailer to get a better tariff. An unstable customer base makes it much harder for the retailer to gather the reliable statistical data it needs to refine its demand forecast.

4.2.1.1 Example 4.1

Table 4.2 illustrates the daily operation of a retailer. Figures 4.1, 4.2 and 4.3 give a graphical representation of the data contained in this table. As shown on the second and third lines of Table 4.2, our retailer has forecasted the demand of its customers for a 12-h period and has purchased energy to meet this anticipated demand. The amount purchased for each hour results from a combination of contracts of various types (long-term bilateral, forwards, futures, screen-based transactions). The fourth and fifth lines of the table show respectively the average and total costs of the energy purchased for each period. The average cost tends to be higher during hours of peak demand.

As one might expect, the actual demand does not match the forecast and there are positive and negative imbalances at each hour. These imbalances are compulsorily settled at the spot prices shown in the eighth line of the table and result in additional balancing costs (or revenue if the imbalance is negative) for our retailer. Adding the balancing and contract costs gives the total cost of energy for each hour. We will assume that our retailer has opted for a very simple tariff structure and charges a flat rate of 38.50 \$/MWh to all of its customers. The "Total Revenues" and "Profits" lines of the table show the amounts that accrue for each hour. Our retailer makes an operational profit during the hours of low prices and a loss during the hours of high prices. Overall, for this 12-h period, the bottom line shows a loss of \$1154. Our retailer has to hope that this is not a typical period and that the average cost of purchasing electricity will be lower on other days. If this turns out to be a typical day, the retail rate will have to be raised to above the average purchasing cost of electrical energy (including spot purchases), which is \$39.23 for this period. The relatively high balancing costs suggest that our retailer could also increase its profitability by

Table 4.2 Data for Example 4.1

Period	Units	1	2	3	4	5	6	7	8	9	10	11	12	Average	Total
Load forecast	(MWh)	221	219	254	318	358	370	390	410	382	345	305	256	325	3828
Contract purchases	(MWh)	221	219	254	318	358	370	390	410	382	345	305	256	325	3828
Average costs	($/MWh)	24.70	24.5	27.50	35.20	40.70	42.40	45.50	48.60	44.20	38.80	33.40	27.70	36.10	
Contract costs	($)	5459	5366	6985	11 194	14 571	15 688	17 745	19 926	16 884	13 386	10 187	7091	12 040	144 482
Actual loads	(MWh)	203	203	287	328	361	401	415	407	397	381	331	240	330	3954
Imbalances	(MWh)	−18	−16	33	10	3	31	25	−3	15	36	26	−16	10.5	
Spot prices	($/MWh)	13.20	12.50	17.40	33.30	69.70	75.40	70.10	102.30	81.40	63.70	46.90	18.30	50.35	
Balancing costs	($)	−238	−200	574	333	209	2337	1753	−307	1221	2293	1219	−293	742	8901
Total costs	($)	5221	5166	7559	11 527	14 780	18 025	19 498	19 619	18 105	15 679	11 406	6798	12 782	153 383
Total revenues	($)	7815.5	7815.5	11 050	12 628	13 899	15 439	15 978	15 670	15 285	14 669	12 744	9240	12 686	152 229
Profits	($)	2595	2650	3491	1101	−882	−2587	−3521	−3950	−2821	−1011	1338	2442	−96	−1154
Profits w/o error	($)	3050	3066	2794	1049	−788	−1443	−2730	−4141	−2177	−104	1556	2765	241	2896

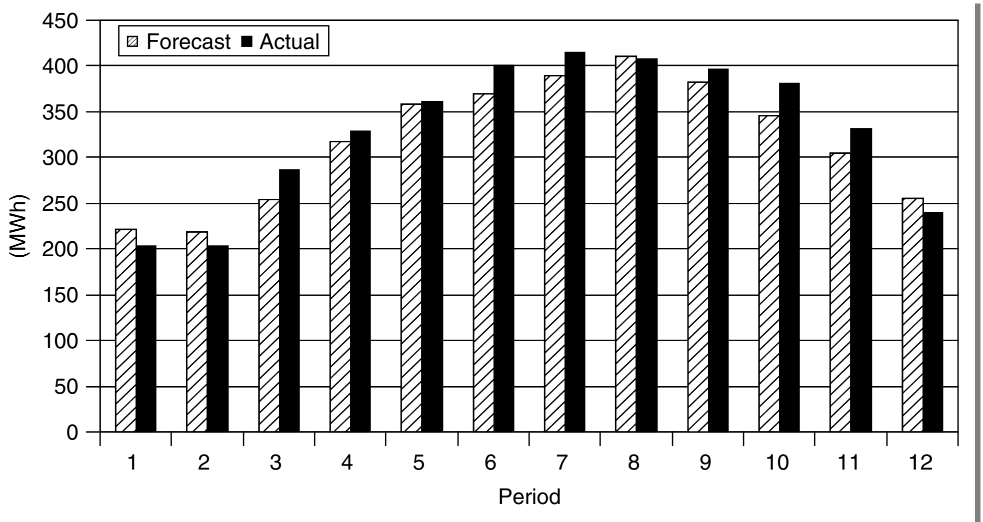


Figure 4.1 Forecast and actual demand for Example 4.1

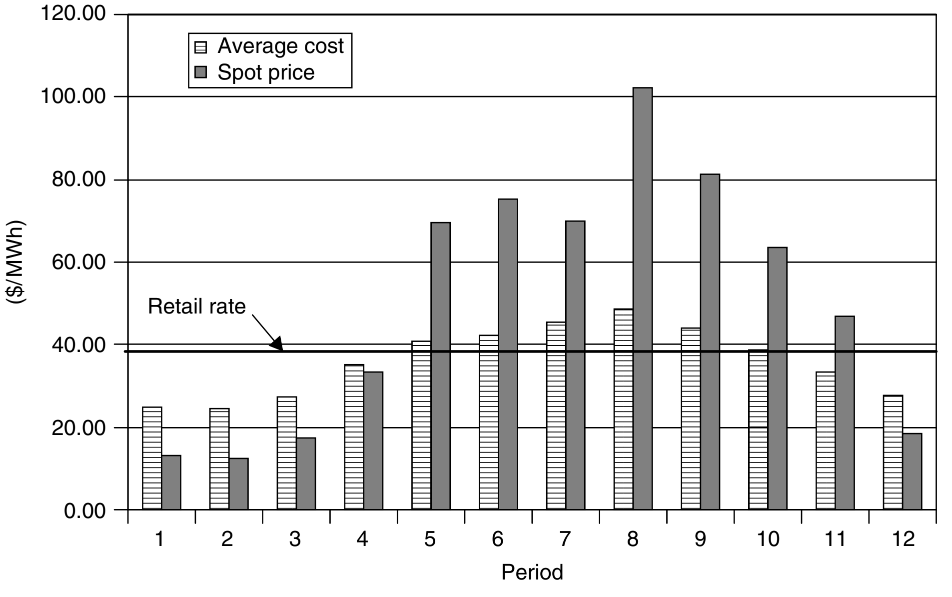


Figure 4.2 Costs and prices for Example 4.1

improving the accuracy of its forecast. To illustrate this point, the last line of the table shows what the profits would be if the demand turned out to be exactly equal to the forecast and the retailer was not exposed to the spot prices. If this perfect forecast had been achieved during this period, our retailer would have made a profit of $2896.

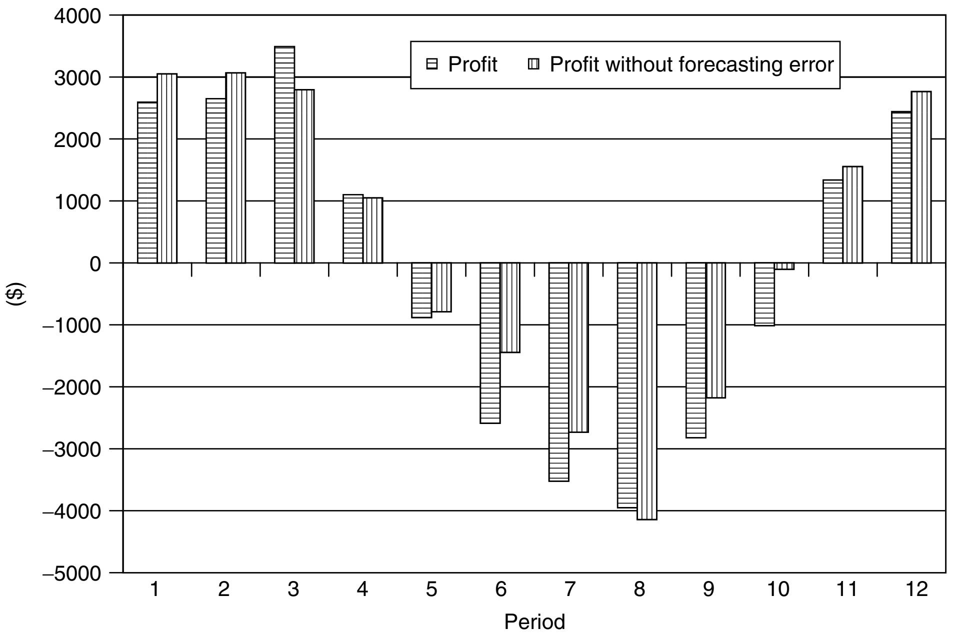


Figure 4.3 Profits and losses for Example 4.1

4.3 The Producer's Perspective

In this section, we will take the perspective of a generating company that tries to maximize the profits it derives from the sale of electrical energy produced by a single generating unit called unit i. For the sake of simplicity, we will consider a period of one hour and we will assume that all quantities remain constant during that period. The maximization of the profit from this unit during this hour can be expressed as the difference between the revenue resulting from the sale of the energy it produces and the cost of producing this energy:

$$\max \Omega_i = \max[\pi \cdot P_i - C_i(P_i)] \tag{4.1}$$

where P_i is the power produced by unit i during that hour, π is the price at which this energy is sold and $C_i(P_i)$ is the cost of producing this energy. If we assume that the only variable over which the company has direct control is the power produced by this unit, the necessary condition for optimality corresponding to Equation (4.1) is

$$\frac{\mathrm{d}\Omega_i}{\mathrm{d}P_i} = \frac{\mathrm{d}(\pi \cdot P_i)}{\mathrm{d}P_i} - \frac{\mathrm{d}C_i(P_i)}{\mathrm{d}P_i} = 0 \tag{4.2}$$

The first term in this expression represents the marginal revenue of unit i, that is, the revenue the company would get for producing an extra megawatt during this hour. The second term represents the cost of producing this extra megawatt, that is, the marginal cost. To maximize profits, the production of unit i must therefore be adjusted up to the level at which its marginal revenue is equal to its marginal cost:

$$MR_i = MC_i \tag{4.3}$$

4.3.1 Perfect competition

4.3.1.1 Basic dispatch

If competition is perfect (or if the potential output of the unit is very small compared to the size of the market), the price π is not affected by changes in P_i. The marginal revenue of unit i is thus

$$MR_i = \frac{\mathrm{d}(\pi \cdot P_i)}{\mathrm{d}P_i} = \pi \tag{4.4}$$

which expresses the fact that a price-taking generator collects the market price for each megawatt-hour that it sells. Under these conditions, if the marginal cost is a monotonically increasing function of the power produced, the generating unit should increase its output up to the point at which the marginal cost of production is equal to the market price:

$$\frac{\mathrm{d}C_i(P_i)}{\mathrm{d}P_i} = \pi \tag{4.5}$$

The marginal cost includes the costs of fuel, maintenance and all other items that vary with the power produced by the unit. Costs that are not a function of the amount of power produced in the period under consideration (e.g. the amortized cost of building the plant or the fixed maintenance and personnel costs) are not factored in the marginal cost and are thus irrelevant when making short-term production decisions.

As long as competition is perfect, the output of each generating unit should be determined using Equation (4.5). Since the price is considered as given, this implies that all generating units can be dispatched independently, even if a generating company owns more than one unit. We will consider in a later section the much more complicated case in which the total capacity of the generating units owned by a single company is large enough to influence the price of energy.

4.3.1.2 Example 4.2

Generating units fired with fossil fuels are characterized by input–output curves that specify the amount of fuel (usually expressed in MJ/h or MBTU/h) required to produce a given and constant electrical power output for one hour.

Consider a coal-fired steam unit whose minimum stable generation is 100 MW (i.e. the minimum amount of power that it can produce continuously) and whose maximum output is 500 MW. On the basis of measurements taken at the plant, the input–output curve of this unit is estimated as

$$H_1(P_1) = 110 + 8.2\,P_1 + 0.002\,P_1^2 \text{ MJ/h}$$

The hourly cost of operating this unit is obtained by multiplying the input–output curve by the cost of fuel F in \$/MJ:

$$C_1(P_1) = 110\,F + 8.2\,F\,P_1 + 0.002\,F\,P_1^2 \ \$/\text{h}$$

If we assume that the cost of coal is 1.3 \$/MJ, the cost curve of this unit is

$$C_1(P_1) = 143 + 10.66\,P_1 + 0.0026\,P_1^2 \ \$/\text{h}$$

If the price at which electrical energy can be sold is 12 \$/MWh, the output that this unit should produce is given by

$$\frac{\mathrm{d}C_1(P_1)}{\mathrm{d}P_1} = 10.66 + 0.0052\,P_1 = 12\ \$/\text{MWh or } P_1 = 257.7\ \text{MW}$$

In practice, optimally dispatching even a single generating unit is more complex than Equation (4.5) suggests. In the following subsections, we will examine how the cost and technical characteristics of the generating units affect the basic dispatch.

4.3.1.3 Unit limits

Suppose that the maximum power $P_i^{\max}$ that can be produced by generating unit i is such that

$$\left.\frac{\mathrm{d}C_i(P_i)}{\mathrm{d}P_i}\right|_{P_i^{\max}} \leq \pi \tag{4.6}$$

This generating unit should therefore produce $P_i^{\max}$. On the other hand, if the minimum stable generation of unit i is such that

$$\left.\frac{\mathrm{d}C_i(P_i)}{\mathrm{d}P_i}\right|_{P_i^{\min}} > \pi \tag{4.7}$$

this unit cannot generate profitably at that price and the only way to avoid losing money on its operation is to shut it down.

4.3.1.4 Example 4.3

The generating unit of the previous example should operate at its maximum output whenever the price of electrical energy is greater than or equal to

$$\left.\frac{\mathrm{d}C_i(P_i)}{\mathrm{d}P_i}\right|_{500\,\text{MW}} = 10.66 + 0.0052 \cdot 500 = 13.26\ \$/\text{MWh}$$

On the other hand, this unit cannot operate profitably if the price drops below

$$\left.\frac{\mathrm{d}C_i(P_i)}{\mathrm{d}P_i}\right|_{100\,\text{MW}} = 10.66 + 0.0052 \cdot 100 = 11.18\ \$/\text{MWh}$$

4.3.1.5 Piecewise linear cost curves

Input–output curves are drawn on the basis of measurements taken while the generating unit is operating at various levels of output. Even if every effort is made to make these measurements as accurate as possible, the data points usually do not line up along a smooth curve. A piecewise linear interpolation of this data is therefore just as acceptable as a quadratic function.

Figure 4.4 shows a piecewise linear cost curve and its associated marginal cost curve. Since each segment of the cost curve is linear, each segment of the marginal cost curve is constant. This makes the process of dispatching the unit in response to

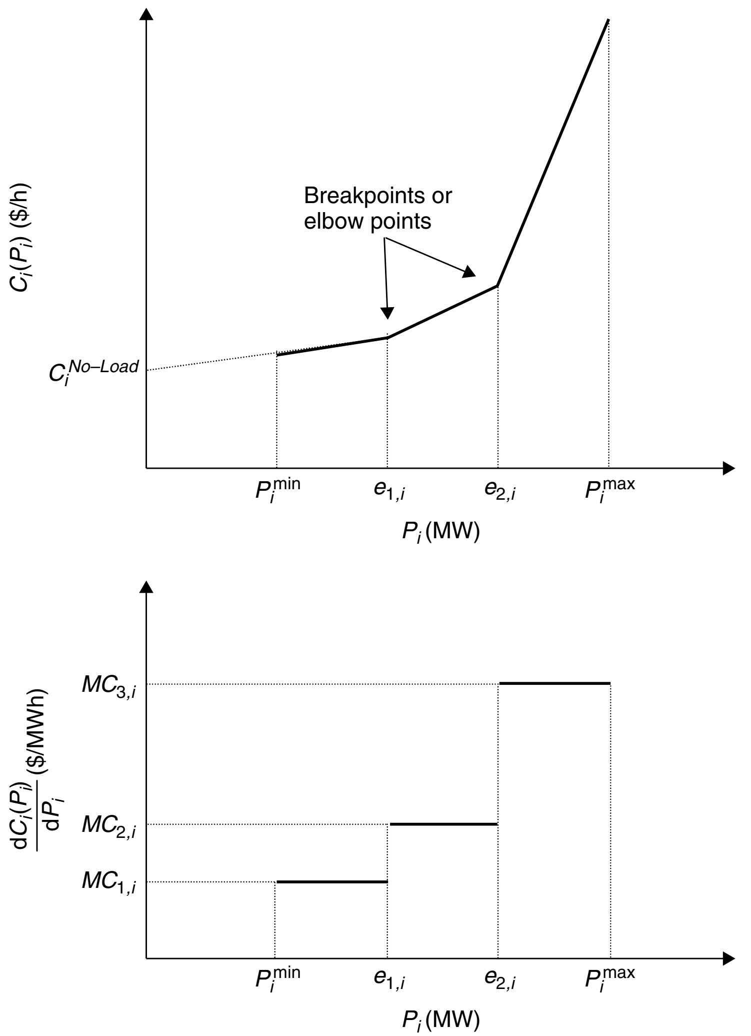


Figure 4.4 Piecewise linear cost curve and its associated piecewise constant incremental cost curve

electrical energy prices very simple.

$$\begin{aligned} \pi < MC_{1,i} &\Rightarrow P_i = P_i^{\min} \\ MC_{1,i} < \pi < MC_{2,i} &\Rightarrow P_i = e_{1,i} \\ MC_{2,i} < \pi < MC_{3,i} &\Rightarrow P_i = e_{2,i} \\ MC_{3,i} < \pi &\Rightarrow P_i = P_i^{\max} \end{aligned} \tag{4.8}$$

If the price is exactly equal to the value of one of the segments of the marginal cost curve, the generation can take any value within that segment. The marginal cost at a breakpoint is equal to the slope of the next segment because the marginal cost is traditionally defined as the cost of the next megawatt, not the cost of the previous megawatt.

4.3.1.6 Example 4.4

The quadratic cost curve of Example 4.2 can be approximated by the following three-segment piecewise linear cost curve:

$$\begin{aligned} 100 \le P_1 \le 250 &: C_1(P_1) = 11.57\, P_1 + 78.0\,\$/\text{h} \\ 250 \le P_1 \le 400 &: C_1(P_1) = 12.35\, P_1 - 117.0\,\$/\text{h} \\ 400 \le P_1 \le 500 &: C_1(P_1) = 13.00\, P_1 - 377.0\,\$/\text{h} \end{aligned}$$

Figure 4.5 shows how this unit should be dispatched as the price paid for electrical energy varies.

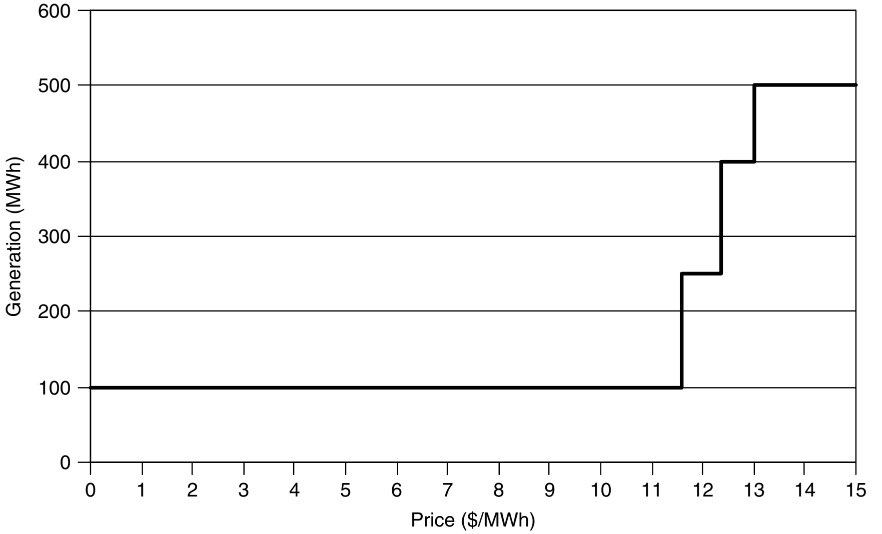


Figure 4.5 Dispatch of the generating unit of Example 4.4 as a function of the price of electrical energy

4.3.1.7 No-load cost

Producers do not decide the amount of electrical energy that they sell solely on the basis of a comparison between the market price and the marginal cost of production. Running a unit in such a way that its marginal cost is equal to the market price does not guarantee that a profit will be made. Producers must also consider the quasi-fixed costs associated with the operation of a generating unit, that is, the costs that are incurred only if the unit is generating but are independent of the amount of power generated. The first type of quasi-fixed cost is the no-load cost. If it is possible for the unit to remain connected to the system while supplying no electrical power, the no-load cost represents the cost of the fuel required to keep the unit running. Such a mode of operation is not possible for most thermal generating units. The no-load cost is then simply the constant term in the cost curve and does not have a physical meaning.

As we discussed in Chapter 2, selling at marginal cost is profitable only if this marginal cost is greater than the *average* cost of production.

4.3.1.8 Example 4.5

Let us assume that the unit of Example 4.4 is always dispatched optimally as the market price for electrical energy varies. This means that it is dispatched according to Figure 4.5. Figure 4.6 shows that its profits increase in a piecewise linear fashion with the price of electrical energy. Because of the no-load cost, the unit becomes profitable only when the price reaches 11.882 \$/MWh.

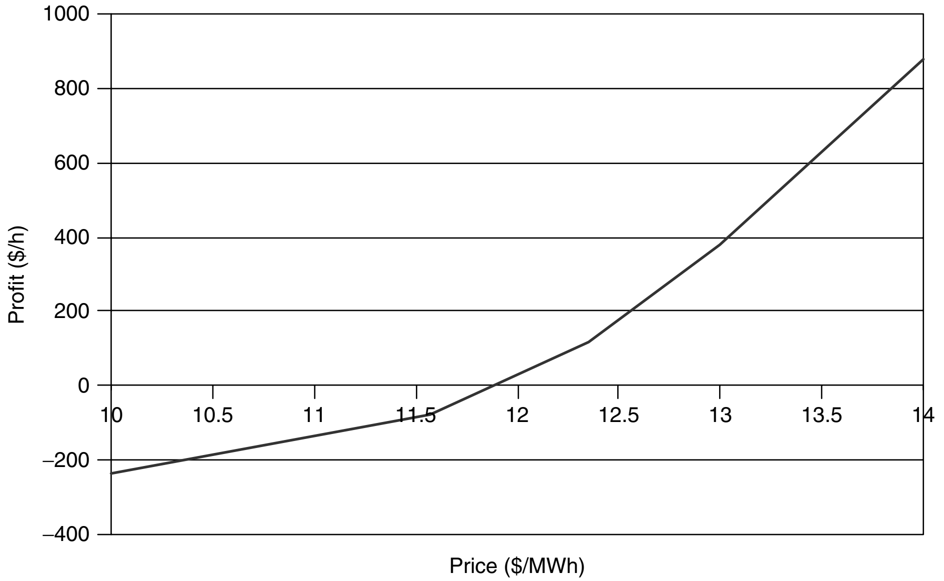


Figure 4.6 Profit accrued by the operator of the generating unit of Example 4.4 if this unit is dispatched optimally as the price of electrical energy varies

4.3.1.9 Scheduling

Since the demand for electrical energy changes over time, the price that a generator gets for its production varies. As we saw in the previous chapter, the price for electrical energy is usually constant for a period of time whose duration ranges from a few minutes to an hour depending on the market. Given a profile of prices extending over a day or more, the optimization described above could be repeated for each market period taken separately. Unfortunately, the resulting production schedule would not be optimal because it neglects the cost of starting up generating units. It would also often be technically infeasible because this approach ignores the constraints on the transitions that generating units can make between operating states. Other economic opportunities and environmental constraints may also affect the optimization of the sale of electrical energy. These different types of constraints are discussed below.

Generating units that have large start-up costs or must obey restrictive constraints will therefore not maximize their profits if their output is optimized over each period individually. Instead, their operation must be scheduled over a horizon ranging from one day to a week or more. This problem has some similarities with the unit commitment problem that monopoly utilities solve to determine how to meet a given load schedule at minimum cost with a given set of generating units. The essence of both problems is to balance the quasi-fixed and variable elements of the cost while satisfying the constraints. In the unit commitment problem, the production of all units is optimized together because their total output must be equal to the total load. On the other hand, if we assume that a generator is a price taker, its production can be optimized independently of the production of other generators. Even when this price-taking approximation holds, scheduling generation to maximize profits is computationally complex. The on/off nature of some of the decision variables makes the problem nonconvex and a rigorous treatment of the constraints significantly increases the dimensionality of the problem. Techniques such as dynamic or mixed-integer programming have been successfully used to solve this problem.

Optimally scheduling the production of a generation unit over a time horizon requires a forecast of the price of electrical energy at each period. The error inherent in such a forecast has an effect on the a posteriori optimality of the schedule. Forecasting prices accurately is extremely complex because of the number of influential factors and the lack of information on some of these factors. Since the price of electrical energy depends on the market equilibrium, it is influenced by both load and generation factors. On the load side, all the temporal, meteorological, economic and special factors that are used in load forecasting must also be taken into account when forecasting prices. The generation side is considerably more troublesome because some events occur at random (e.g. failures of generating units) and others are not always publicly announced in advance (e.g. planned outages for maintenance).

4.3.1.10 Start-up cost

The start-up cost of a generating unit represents the cost of getting this unit running and ready to produce from a shutdown state. It is thus another type of quasi-fixed cost. Diesel generators and open cycle gas turbines have low start-up costs because units of this type start quickly. On the other hand, large thermal units require a considerable amount of heat energy before the steam is at a temperature and pressure that are sufficient to sustain the generation of electric power. These units therefore have a large start-up cost. To maximize the profitability of a thermal unit, these start-up costs must be amortized over a long period. This may even involve operating the unit at a loss for a few hours rather than shutting it down and having to re-incur the start-up cost when prices increase again.

4.3.1.11 Example 4.6

Let us examine how the coal-fired plant of Example 4.2 should be scheduled over a period of several hours. We will assume that the price at which electrical energy can be sold is set on an hourly basis and that the prices for the next few hours are shown in Figure 4.7. Suppose also that the generating unit is started up at hour 1 and that the cost of bringing it online is $600. The table below summarizes the results.

Hour	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Price ($/MWh)	12.0	13.0	13.5	10.5	12.5	13.5	11.5
Generation (MW)	257.7	450.0	500.0	100.0	353.8	500.0	161.5
Revenue ($)	3092	5850	6750	1050	4423	6750	1858
Running cost ($)	3063	5467	6123	1235	4240	6123	1933
Start-up cost ($)	600	0	0	0	0	0	0
Total cost ($)	3663	5467	6123	1235	4240	6123	1933
Profit ($)	−571	383	627	−185	183	627	−75
Cumulative profit ($)	−571	−188	439	254	437	1064	989

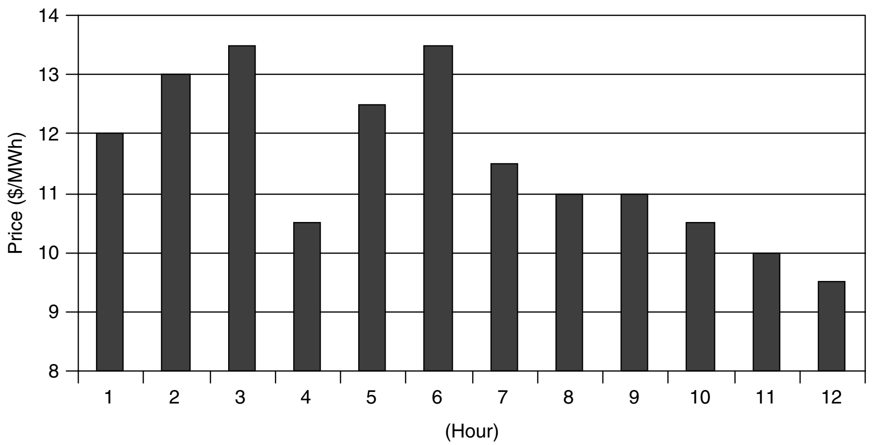


Figure 4.7 Price of electrical energy for Example 4.6

The first thing to notice is that the optimal generation varies substantially as the price of electrical energy fluctuates. The unit generates at maximum capacity at hours 3 and 6 and at minimum capacity at hour 4. The operation of this unit records a deficit at hour 1 because of the cost of starting up the unit. By hour 3, this start-up cost has been recovered and the unit begins making a profit. The price at hour 4 is so low that the unit shows a loss even though it operates at its minimum capacity. Not shutting down the unit at this hour is, however, the best decision because it avoids incurring the start-up cost again at hour 5. At hour 7, the unit records a deficit even though it is not running at its minimum capacity. This is because the unit does not generate enough to recover its no-load cost. If the price continues to decline over the next few hours, the best strategy would be to shut the unit down at the end of hour 6 and to wait until prices are higher before starting it up again.

4.3.1.12 Dynamic constraints

Starting up or shutting down a thermal generating unit or even increasing or decreasing its output by more than a small amount causes considerable mechanical stress in the prime mover. Excessive stress damages the plant and shortens its life. Limits are therefore often placed on such changes to protect these expensive assets. These safeguards have long-term benefits but short-term costs. In particular, placing a limit on the rate at which a unit can increase or decrease its output may prevent it from achieving its economically optimal output in successive periods. Minimizing the cost of these ramp-rate limits requires that the operation of the unit be optimized over at least several hours.

To limit the damage caused by frequent start-ups and shutdowns, a minimum is often placed on the number of hours that a thermal unit must remain connected to the system once it has been started. A similar limit is usually placed on the number of hours that a unit must remain idle once it has been shut down. These limits ensure that there is enough time for the temperature gradients in the turbine to subside. These minimum uptime and minimum downtime constraints reduce the opportunities to change the status of the unit and can have a profound impact on the optimal schedule. For example, the minimum downtime constraint could force a unit to continue generating at a loss during a short period of low prices because shutting it down would prevent it from reaping larger profits later on.

4.3.1.13 Environmental constraints

Generating plants must abide by environmental regulations that may affect their ability to operate at their economic optimum. Emissions of certain pollutants by fossil-fuel-fired power plants are increasingly regulated. In some cases, the rate at which a certain pollutant is released in the atmosphere is limited, thereby reducing the maximum power output of the plant. In other cases, it is the total amount of a pollutant released over a year that is capped, putting a complex integral constraint on the operation of the plant.

While hydroelectric plants do not emit pollutants and are more flexible than thermal plants, there may be constraints on their use of water. These constraints are designed to ensure the availability of water for recreation or to help the reproduction of endangered species of fish. Water must also be made available for irrigation and for other hydroelectric plants. Optimizing the operation of hydroelectric plants is a very complex problem, particularly in river basins with multiple interconnected plants.

4.3.1.14 Other economic opportunities

The amount of electric power produced by cogeneration or combined heat and power plants is often determined by the needs of the associated industrial process. The ability of such plants to take advantage of opportunities to sell energy on the electricity market therefore may be limited.

Besides electrical energy, generators can provide other services such as reserve capacity, load following, frequency regulation and voltage regulation. These other services, which are usually called ancillary or system services, constitute a source of revenue that is distinct from the sale of electrical energy. We will discuss the issues related to the provision of these services in Chapter 5. At this point, we simply need to note that a producer's ability to trade electrical energy may be affected by contracts that it has entered into for the provision of ancillary services. Similarly, the production of electrical energy may hamper a generator's ability to provide ancillary services.

4.3.2 The production versus purchase decision

Consider the case of a generating company that has signed a contract for the supply of a given load L during a single hour. Let us first assume that this company decides to meet its contractual obligation to supply this load using its portfolio of N generating plants. It will obviously try to produce the energy required at minimum cost to itself. Mathematically, this can be formulated as the following optimization problem:

$$\text{Minimize} \sum_{i=1}^{N} C_i(P_i) \text{ subject to } \sum_{i=1}^{N} P_i = L \tag{4.9}$$

where P_i represents the production of unit i of the portfolio and $C_i(P_i)$ the cost of producing this amount of power with this unit. From calculus, we know that forming a Lagrangian function ℓ that combines the objective function and the constraint is the easiest way to solve such an optimization problem:

$$\ell(P_1, P_2, \ldots, P_N, \lambda) = \sum_{i=1}^{N} C_i(P_i) + \lambda \left(L - \sum_{i=1}^{N} P_i \right) \tag{4.10}$$

where λ is a new variable called a Lagrange multiplier.

Setting the partial derivatives of this Lagrangian function to zero gives the necessary conditions for optimality and solving these equations gives the optimal solution as follows:

$$\frac{\partial \ell}{\partial P_i} \equiv \frac{\mathrm{d}C_i}{\mathrm{d}P_i} - \lambda = 0 \quad \forall i = 1, \ldots N$$
$$\frac{\partial \ell}{\partial \lambda} \equiv \left(L - \sum_{i=1}^{N} P_i \right) = 0 \tag{4.11}$$

From these optimality conditions, we conclude that all the generating units in the portfolio should be operated at the same marginal cost and that this marginal cost is equal to the value of the Lagrange multiplier λ:

$$\frac{\mathrm{d}C_1}{\mathrm{d}P_1} = \frac{\mathrm{d}C_2}{\mathrm{d}P_2} = \cdots \frac{\mathrm{d}C_N}{\mathrm{d}P_N} = \lambda \tag{4.12}$$

The value of the Lagrange multiplier is thus equal to the cost of producing one additional megawatt-hour with any of the generating units. This Lagrange multiplier is therefore often called the *shadow price* of electrical energy.

Let us now suppose that this generating company can participate in a spot market for electricity. If the market price π is lower than the shadow price λ at which it can produce energy, our generating company should purchase energy on the market and reduce its own production up to the point at which

$$\frac{\mathrm{d}C_1}{\mathrm{d}P_1} = \frac{\mathrm{d}C_2}{\mathrm{d}P_2} = \cdots \frac{\mathrm{d}C_N}{\mathrm{d}P_N} = \pi \tag{4.13}$$

If the amount of energy involved is significant, the market may not have enough liquidity to handle this transaction without an increase in the price π. This issue will be discussed in more details in the next section.

4.3.2.1 Example 4.7

The 300-MW load of a small power system must be supplied at minimum cost by two thermal generating units and a small run-of-the-river hydro plant. The hydro plant generates a constant 40 MW and the cost functions of the thermal plants are given by the following expressions:

$$\text{Unit A: } C_A = 20 + 1.7\, P_A + 0.04\, P_A^2 \$/\text{h}$$
$$\text{Unit B: } C_B = 16 + 1.8\, P_B + 0.03\, P_B^2 \$/\text{h}$$

Since the variable operational cost of the hydro unit is negligible, the Lagrangian function of this optimization problem can be written as follows:

$$\ell = C_A(P_A) + C_B(P_B) + \lambda(L - P_A - P_B)$$

where L represents the 260 MW load that the thermal units must provide.

Setting the partial derivatives of the Lagrangian equal to zero, we obtain the necessary conditions for optimality:

$$\frac{\partial \ell}{\partial P_A} \equiv 1.7 + 0.08\,P_A - \lambda = 0$$

$$\frac{\partial \ell}{\partial P_B} \equiv 1.8 + 0.06\,P_B - \lambda = 0$$

$$\frac{\partial \ell}{\partial \lambda} \equiv L - P_A - P_B = 0$$

Solving this system of equations for λ , we get the marginal cost of electrical energy in this system for this loading condition:

$$\lambda = 10.67\,\$/\text{MWh}$$

We can then calculate the optimal outputs of the thermal units:

$$P_A = 112.13\,\text{MW}$$

$$P_B = 147.87\,\text{MW}$$

Replacing these values in the cost functions, we find the total cost of supplying this load:

$$C = C_A(P_A) + C_B(P_B) = 1{,}651.63\,\$/\text{hour}$$

4.3.3 Imperfect competition

When competition is less than perfect, some firms (the strategic players) are able to influence the market price through their actions. It is quite common for an electricity market to consist of a few strategic players and a number of price takers. A company that owns more than one generating unit is likely to have a greater influence on the market price if it optimizes the combined output of its entire portfolio of units. Optimizing the output of each unit separately would thus not maximize the firm's profit. The total profit of a firm that owns multiple generating units is thus

$$\Omega_f = \pi \cdot P_f - C_f(P_f) \tag{4.14}$$

where P_f represents the combined output of all the units controlled by that firm, while $C_f(P_f)$ represents the minimum cost at which this firm is able to produce this power. We will now assume that the market price π is no longer a variable that is beyond the control of any single market participant. Similarly, the power sold by firm f depends not only on its own decisions but also on those of its competitors. We will therefore rewrite Equation (4.14) as follows to summarize these dependencies:

$$\Omega_f = \Omega_f(X_f, X_{-f}) \tag{4.15}$$

where X_f represents the actions of firm f and X_{-f} those of its competitors.

Equation (4.15) shows that firm f cannot optimize its profits in isolation. It must consider what the other firms will do. At first sight, this may seem very difficult because these firms are competitors and exchanging information would be illegal. However, it is reasonable to make the assumption that all firms are behaving in a rational manner, that is, that they are all trying to maximize their profits. In other words, we have to find for each firm f the actions X_f^* such that

$$\Omega_f(X_f^*, X_{-f}^*) \geq \Omega_f(X_f, X_{-f}^*) \ \forall f \tag{4.16}$$

where X_{-f}^* represents the optimal action of the other firms.

Such interacting optimization problems form what is called in game theory a *noncooperative game*. The solution of such a game, if it exists, is called a *Nash equilibrium* and represents a market equilibrium under imperfect competition.

While representing the possible actions or decisions of a firm by the generic variable X_f allowed us to formulate the problem elegantly, it hides the fact that the solution of Equation (4.16) depends on how we model the strategic interactions between the firms. In the following subsections, we discuss three models that have been proposed in the literature.

4.3.3.1 Bertrand interaction or game in prices

If we assume that the participants interact according to the Bertrand model, the price at which each firm offers its electrical energy is the only decision variable:

$$X_f = \pi_f \ \forall f \tag{4.17}$$

The amount of energy sold by firm f is thus a function of its own offer price and the offer prices of its competitors. Firm f's revenue is thus given by

$$\pi \cdot P_f = \pi \cdot P_f(\pi_f, \pi_{-f}^*) \tag{4.18}$$

Firm f acts as if its competitors do not change their offer prices in response to its own decisions. For an undifferentiated commodity such as electrical energy, f can sell as much as it wants as long as its price is lower than the prices of its competitors:

$$\begin{aligned} P_f(\pi_f, \pi_{-f}^*) &= P_f \text{ if } \pi_f \leq \pi_{-f}^* \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{4.19}$$

As we discussed in Chapter 2, the assumption that the competition will not adjust its prices is unrealistic. Under this model, the market price will be equal to the marginal cost of production of the most efficient firm. On the one hand, no firm can offer a lower price without making a loss. On the other hand, a higher price is not sustainable because it will be undercut by the most efficient firm.

If firms can differentiate their products (e.g. if firm f sells "green" electricity while the others do not), the relation between quantity sold and price is more complex than Equation (4.19) suggests and higher prices might be sustainable.

4.3.3.2 Cournot interaction or game in quantities

In a Cournot model, each firm decides the quantity that it wants to produce

$$X_f = P_f \quad \forall f \tag{4.20}$$

The price is then determined by the inverse market demand function, which expresses the market price as a function of the total amount of energy traded

$$\pi = \pi(P_f + P_{-f}) = \pi(P) \tag{4.21}$$

If firm f assumes that its competitors will not adjust the amount of energy they produce, its revenue is given by

$$\pi \cdot P_f = \pi(P_f + P_{-f}^{*}) \cdot P_f \tag{4.22}$$

Its marginal revenue is therefore

$$MR_f = \frac{\partial(\pi(P) \cdot P_f)}{\partial P_f} = \pi + \frac{\partial \pi}{\partial P} \cdot P_f \tag{4.23}$$

The Cournot model suggests that firms should be able to sustain prices that are higher than the marginal cost of production, with the difference being determined by the price elasticity of the demand. Numerical results obtained with Cournot models are very sensitive to this elasticity. In particular, for a commodity like electrical energy that has a very low elasticity, the equilibrium price calculated using a Cournot model tends to be higher than the prices that are observed in the actual market.

4.3.3.3 Example 4.8

Let us consider the case of a market where two firms (A and B) compete for the supply of electrical energy. We will assume that empirical studies have shown that the inverse demand curve at a particular hour is given by

$$\pi = 100 - D \text{ \$/MWh} \tag{4.24}$$

where D is the demand for electrical energy at this hour. Let us also suppose that Firm A can produce energy more cheaply than Firm B:

$$\begin{aligned} C_A &= 35 \cdot P_A \text{ \$/h} \\ C_B &= 45 \cdot P_B \text{ \$/h} \end{aligned} \tag{4.25}$$

If we assume a Bertrand model for the competition in this market, Firm A would set its price at slightly less than the marginal cost of production of Firm B, (i.e. 45 \$/MWh) and would capture the whole market. At that price, the demand would be 55 MWh and Firm A would achieve a profit of \$550. At that price Firm B would lose money on any megawatt-hour that it sold and would therefore decide not to produce anything. It would then obviously not make any profit.

On the other hand, if we assume a Cournot model of competition, the state of the market is determined by the production decisions made by each firm. Let us suppose that Firms A and B have both decided to produce 5 MWh. According to the Cournot model, the market price must be such that the demand equals the total production. The total demand will be 10 MWh and, according to Equation (4.24), the market price will be 90 \$/MWh. Given the market price and the productions, we can easily find that Firms A and B make a profit of \$275 and \$225 respectively. The following cell summarizes this state of the market:

10	275
225	90

Similar cells can be generated for other combinations of productions and arranged as shown in Table 4.3. This table illustrates the interactions of the two firms under a Cournot model of competition. Toward the top left corner of the table, generators are driving the price up by limiting production. As production increases (i.e. as we move right or down through the table), the price decreases and the demand increases. Toward the bottom right corner of the table, the market is flooded and the price drops below Firm B's marginal cost of production, causing it to lose money. Among the possibilities shown in Table 4.3, Firm A would prefer the situation in which it produces 30 MWh and B produces 5 MWh because this would maximize its profit. Similarly, Firm B would like A to produce only 5 MWh so that it could produce 25 MWh and maximize its own profit. The market will not settle in either of these situations because they are not in the best interests of the other firm. Instead, the market will settle at the Nash equilibrium where neither firm can increase its profit through its own actions. The highlighted cell in Table 4.3 corresponds to this equilibrium. The profit of Firm A (\$625) is the largest that it can achieve in that row, that is, by adjusting its own production. Similarly, the profit of Firm B (\$225) is the largest in this column. Therefore, neither firm has an incentive to produce any other amount. While Firm A captures a larger share of the market because its marginal cost of production is lower, it does not freeze Firm B completely out of the market. These firms manage to maintain a price that is much higher than the marginal cost of production. This price is also higher than the value predicted by the Bertrand model.

Rather than constructing a table showing every possible pair of productions, we can formulate and solve this problem mathematically. Since each firm uses the quantity it produces as its decision variable, the profits earned by each firm are given by the following expressions:

$$\Omega_A(P_A, P_B) = \pi(D) \cdot P_A - C_A(P_A) \tag{4.26}$$

$$\Omega_B(P_A, P_B) = \pi(D) \cdot P_B - C_B(P_B) \tag{4.27}$$

Table 4.3 Illustration of Cournot competition in the two-firm market of Example 4.8. The numbers in each cell represent the following quantities:

Demand	Profit of A
Profit of B	Price

Production of firm B ↓ / Production of firm A →	5		10		15		20		25		30		35	
5	10	275	15	500	20	675	25	800	30	875	35	900	40	875
	225	90	200	85	175	80	150	75	125	70	100	65	75	60
10	15	250	20	450	25	600	30	700	35	750	40	750	45	700
	400	85	350	80	300	75	250	70	200	65	150	60	100	55
15	20	225	25	400	30	525	35	600	**40**	**625**	45	600	50	525
	525	80	450	75	375	70	300	65	**225**	**60**	150	55	75	50
20	25	200	30	350	35	450	40	500	45	500	50	450	55	350
	600	75	500	70	400	65	300	60	200	55	100	50	0	45
25	30	175	35	300	40	375	45	400	50	375	55	300	60	175
	625	70	500	65	375	60	250	55	125	50	0	45	−125	40
30	35	150	40	250	45	300	50	300	55	250	60	150	65	0
	600	65	450	60	300	55	150	50	0	45	−150	40	−300	35
35	40	125	45	200	50	225	55	200	60	125	65	0	70	−175
	525	60	350	55	175	50	0	45	−175	40	−350	35	−525	30

where $\pi(D)$ represents the inverse demand curve. If each firm tries to maximize its profit, we have two separate optimization problems. These two optimization problems cannot be solved independently because both firms compete in the same market and the supply must be equal to the demand. Therefore we must also have

$$D = P_A + P_B \tag{4.28}$$

For each of these problems we can write a condition for optimality:

$$\frac{\partial \Omega_A}{\partial P_A} = \pi(D) - \frac{dC_A}{dP_A} + P_A \cdot \frac{d\pi}{dD} \cdot \frac{dD}{dP_A} = 0 \tag{4.29}$$

$$\frac{\partial \Omega_B}{\partial P_B} = \pi(D) - \frac{dC_B}{dP_B} + P_B \cdot \frac{d\pi}{dD} \cdot \frac{dD}{dP_B} = 0 \tag{4.30}$$

Inserting the values given by Equations (4.24) and (4.25) in Equations (4.29), (4.30), and (4.28), we get the following reaction curves:

$$P_A = \tfrac{1}{2}(65 - P_B) \tag{4.31}$$

$$P_B = \tfrac{1}{2}(55 - P_A) \tag{4.32}$$

Solving these two equations simultaneously gives the same equilibrium as what we found by building Table 4.3:

$$P_A = 25\,\text{MWh},\ P_B = 15\,\text{MWh},\ D = 40\,\text{MWh},\ \pi = 60\,\$/\text{MWh}.$$

4.3.3.4 Example 4.9

The data from the previous example also gives us an opportunity to explore what happens when the number of firms competing in a market increases. For the sake of simplicity, we will consider the case in which Firm A competes against an increasing number of firms idetical to Firm B. An optimality condition similar to Equation (4.29) or Equation (4.30) can be written for each of these firms, and this system of equations can be solved together with the inverse demand relation (4.24) and the equation expressing that all these firms compete in the same market:

$$D = P_A + P_B + \cdots + P_N \tag{4.33}$$

where N represents the number of firms competing in this market. In this particular case, these equations are easy to solve for an arbitrary number of firms because firms B to N are identical and thus produce the same amount of energy. Since Firm A produces electrical energy at a lower cost than the other firms, it has a competitive advantage in this market. Figure 4.8 shows that it always produces more than any other firm does. While its share of the market decreases as the number of competing firms increases, it does not tend to zero like the individual share of the other firms. We can see from Figure 4.9 that an increase in the number of competing firms depresses the market price, even if the new firms have the same marginal cost of production as the existing

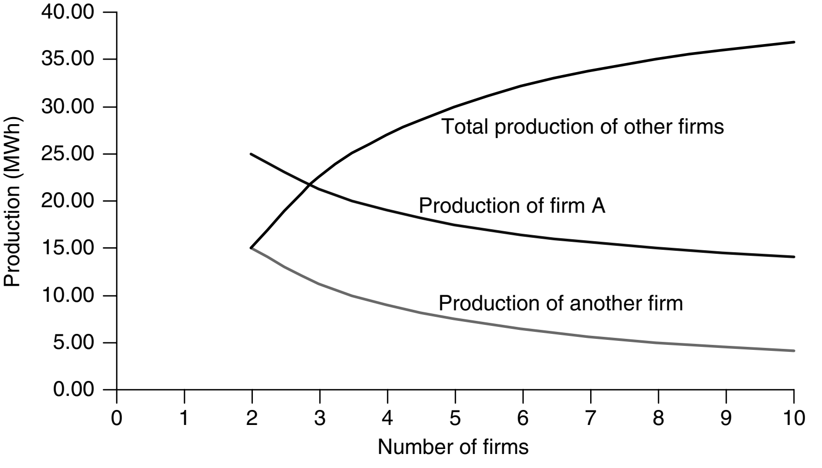


Figure 4.8 Evolution of the production of each firm as the number of competitors increases in the Cournot model of Example 4.9

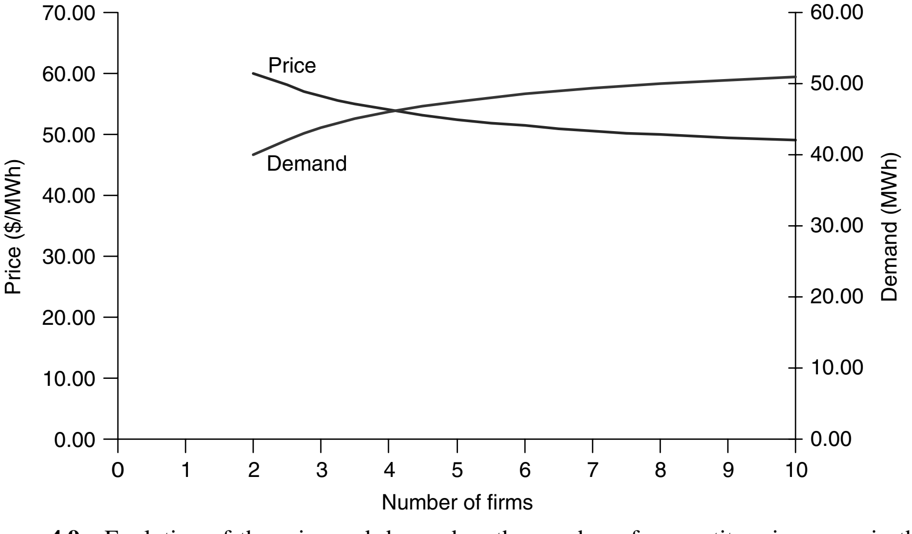


Figure 4.9 Evolution of the price and demand as the number of competitors increases in the Cournot model of Example 4.9

ones. In this case, however, the price asymptotically tends toward 40 $/MWh, which is the marginal cost of production of firms B to N. This heightened competition induces an increase in demand and therefore benefits the consumers. Finally, as Figure 4.10 shows, this increased competition also reduces the profits made by each firm. Because of its cost advantage, the profits of Firm A are larger than the combined profits of all the other firms and, unlike the profits of the firms in the competitive fringe, it does not tend to zero as the number of competitors increases.

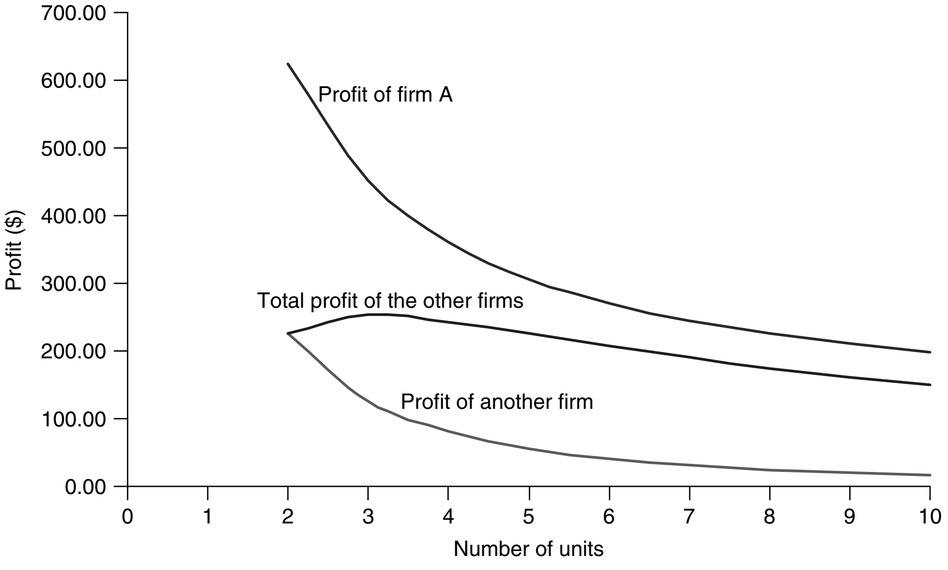


Figure 4.10 Evolution of the profits of each firm as the number of competitors increases in the Cournot model of Example 4.9

4.3.3.5 Supply functions equilibria

While the Cournot model provides interesting insights into the operation of a market with imperfect competition, its application to electricity markets produces unreasonably high forecasts for the market price. More complex representations of the strategic behavior of generating companies have therefore been developed to obtain more realistic market models. In these models, it is assumed that the amount of energy that a firm is willing to deliver is related to the market price through a *supply function*:

$$P_f = P_f(\pi)\ \forall f \tag{4.34}$$

In this case, the decision variables of each firm are thus neither the price nor the quantity but the parameters of its supply function.

At equilibrium, the total demand is equal to the sum of the quantities produced by all the firms:

$$D(\pi) = \sum_f P_f(\pi) \tag{4.35}$$

The profit of each firm can be expressed as follows:

$$\begin{aligned}\Omega_f &= \pi \cdot P_f - C_f(P_f)\\ &= \pi \cdot \left[D(\pi) - \sum_{-f} P_{-f}(\pi)\right] - C_f\left(D(\pi) - \sum_{-f} P_{-f}(\pi)\right)\ \forall f\end{aligned} \tag{4.36}$$

These profit functions can be differentiated with respect to the price to get the necessary conditions for optimality, which after some manipulations can be expressed in the following form:

$$P_f(\pi) = \left(\pi - \frac{\mathrm{d}C_f(P_f)}{\mathrm{d}P_f}\right) \cdot \left(-\frac{\mathrm{d}D}{\mathrm{d}\pi} + \sum_{-f} \frac{\mathrm{d}P_{-f}(\pi)}{\mathrm{d}\pi}\right) \ \forall f \tag{4.37}$$

The solution of this system of equation is an equilibrium point at which all firms simultaneously maximize their profits. These optimality conditions are differential equations because the parameters of the supply functions are unknown. In order to find a unique solution to this set of differential equations, the supply and cost functions are usually assumed to have respectively linear and quadratic forms:

$$P_f(\pi) = \beta_f(\pi - \alpha_f) \ \forall f \tag{4.38}$$

$$C_f(P_f) = \tfrac{1}{2}a_f P_f^2 + b_f P_f \ \forall f \tag{4.39}$$

The decision variables are thus

$$X_f = \{\alpha_f, \beta_f\} \ \forall f \tag{4.40}$$

The optimal values of these variables can be computed by inserting Equations (4.38) and (4.39) as well as the inverse demand function into Equation (4.37). Once these optimal values have been computed using an iterative process, it is then possible to calculate the market price, the demand, as well as the production of each firm. It is interesting to note that if the inverse demand function is affine (i.e. it includes a linear term plus an offset), the supply functions do not depend on the actual level of the demand.

4.3.3.6 Limitations of these models

Published applications to electricity markets of the models described in the previous sections have dealt so far mostly with predictions of market shares over a period of years. These models work on the aggregated capacity of each generating company and are probably not yet sophisticated enough to be useful in the daily optimization of individual generating units. In particular, they do not take into account nonlinearities such as no-load and start-up costs and dynamic constraints on the output of each unit.

Furthermore, formulating the problem as a short-run profit maximization is probably an oversimplification. In some cases, a generating company that has market power may decide to limit or even drive down the market price. Such a course of action could be justified by a desire to increase or maintain market share, by a strategy to discourage entry in the market by new participants or by a fear of attracting regulatory intervention.

4.4 Perspective of Plants with very Low Marginal Costs

Some types of plants (nuclear, hydroelectric, renewable) have negligible or almost negligible marginal costs. The challenge for the owners of such plants is to generate enough revenue to cover their very large investment costs. This causes different problems for the different types of units. Nuclear units must be operated at an almost constant generation level because adjusting their output is usually difficult. Ideally, these plants should be shut down only for refueling because of their extremely high start-up costs. The owners of nuclear power stations must therefore sell the nominal power of their units at every hour and almost at any price. On the other hand, hydroelectric plants (at least those that have a substantial reservoir) can adjust their production at will. However, the amount of energy they have available is determined by the amount of rain or snow that falls in their hydrological basin. To maximize revenues, they must therefore forecast the periods when the price paid for electrical energy will be the highest and sell energy during these periods. Finally, renewable generation depends on the availability of energy sources such as wind and sun that are not only uncontrollable but also unpredictable. Owners of such generating plants will often have to sell their production at rather unfavorable prices.

4.5 The Hybrid Participant's Perspective

A small but growing number of market participants can choose to behave like producers or consumers depending on the circumstances. Pumped hydro plants are the most common type of hybrid participant. In a traditional environment, such plants would consume energy by pumping water during periods of light load. On the other hand, during periods of high load they would produce energy by releasing this water through turbines. These cycles of consumption and production reduce the difference between the peaks and the troughs in the demand curve and hence the total cost of producing energy with thermal plants. In a competitive environment, the operation of such a plant can be profitable if the revenue generated by selling energy during periods of high prices is larger than the cost of the energy consumed during periods of low prices. This calculation must take into account the fact that, because of the losses, only about 75% of the energy consumed for pumping can be sold back.

4.5.1.1 Example 4.10

Consider a pumped hydro plant with an energy storage capacity of 1000 MWh and an efficiency of 75%. Assume that it takes four hours to completely empty or fill the upper reservoir of this plant if it operates at rated power. Suppose that the operator of this plant has decided to go through a full cycle during the 12-h period shown in Figure 4.11. A very simple strategy has been adopted: water will be pumped to the upper reservoir during the four hours with the lowest energy prices (hours 1 to 4) and will be released during the four hours with the highest energy prices (hours 7 to 10). Table 4.4 summarizes the results of this cycle. Only 750 MWh are produced

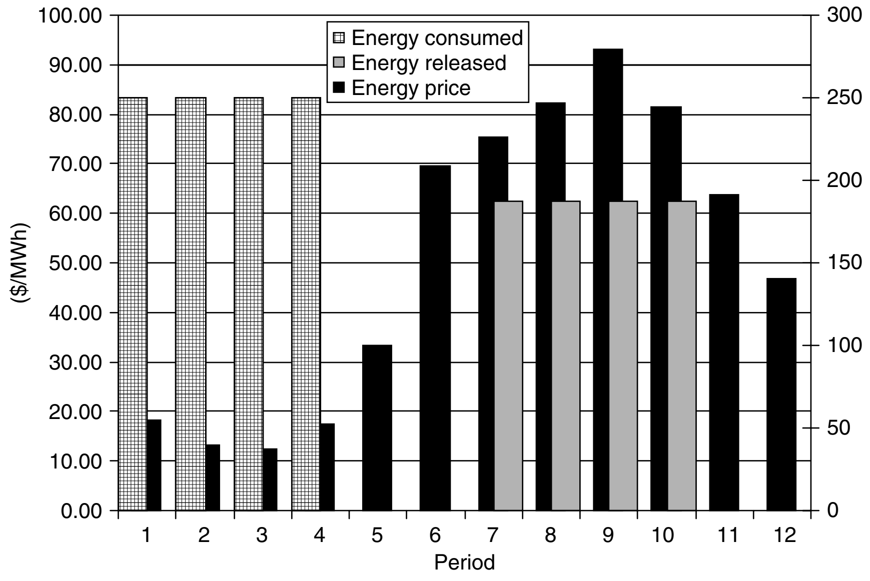


Figure 4.11 Energy prices, energy consumed and energy released for the storage plant of Example 4.10

Table 4.4 Data for Example 4.10

Period Units	Energy prices ($/MWh)	Energy Consumed (MWh)	Energy released (MWh)	Revenue ($)
1	18.30	250	0	−4575
2	13.20	250	0	−3300
3	12.50	250	0	−3125
4	17.40	250	0	−4350
5	33.30	0	0	0
6	69.70	0	0	0
7	75.40	0	187.5	14 137.5
8	82.40	0	187.5	15 450
9	93.20	0	187.5	17 475
10	81.40	0	187.5	15 262.5
11	63.70	0	0	0
12	46.90	0	0	0
Total		1000	750	46 975

and resold because the plant is 75% efficient. In this case, since there are substantial differences between the periods of low prices and the periods of high prices, this cycle of pumping and turbining results in a profit of $46 975. If the differences in price were smaller, this profit would be substantially reduced and might even become negative.

In this example, the price of electrical energy has been taken as given. In practice, a pumped-storage plant would represent a nonnegligible portion of the load, particularly during off-peak hours. The operating strategy of the plant would therefore have to consider the effect that it might have on prices. This type of arbitrage operations using energy storage plants is usually not profitable because the very high cost of amortizing the plant must be subtracted from the operational profits. Since pumped storage hydro plants are highly flexible, they also have the opportunity to take part in the market for ancillary services.

An increasing number of industrial consumers operate processes that cannot be shut down because of interruptions in the electricity supply without causing significant financial losses. Such consumers often install emergency generators capable of supplying at least part of their load during period of outages. When the power system operates normally, but prices are high, these consumers may find that even though the marginal cost of operating these emergency generators may be high, it is lower than the spot price of electrical energy. Under these conditions, they might want to start up their emergency generators to reduce their demand and possibly sell the surplus on the market.

Some power systems in which a competitive electricity market has been introduced are interconnected with neighboring systems that are operated by vertically integrated utilities. These utilities often take part in the competitive market. If the price paid for electrical energy is higher than their marginal cost of production, they will behave like producers in this market. On the other hand, if the price is lower than their marginal cost of production, it is in their best interest to reduce the output of their own generators and purchase power in the competitive market.

4.6 Further Reading

Bunn (2000) surveys recent developments in techniques for forecasting loads and prices, whereas Bushnell and Mansur (2001) provide some data about how consumers respond when faced with market prices for electricity. The complexity of the generation scheduling problem is explained in considerable detail (albeit in the context of a traditional utility) in (Wood and Wollenberg, 1996). Techniques for maximizing the profits of a price-taking generator are discussed in (Arroyo and Conejo, 2000 and Chan, 2000). The reader interested in supply function equilibria will find theoretical and practical discussions of this subject in (Day *et al*. (2002), Green (1996) and Klemperer (1989)). Kirschen (2003) discusses in more details demand-side participation in electricity markets.

Arroyo J M, Conejo A J, Optimal response of a thermal unit to an electricity spot market, *IEEE Transactions on Power Systems*, **15**(3), 2000, 1098–1104.

Bunn D W, Forecasting loads and prices in competitive power markets, *Proceedings of the IEEE*, **88**(2), 2000, 163–169.

Bushnell J B, Mansur E T, *The Impact of Retail Rate Deregulation on Electricity Consumption in San Diego*, Working Paper PWP-082, Program on Workable Energy Regulation, University of Californian Energy Institute, April 2001, *www.ucei.org*.

Chan C J S, *Development of a Profit Maximisation Unit Commitment Program*, MSc Dissertation, UMIST, 2000.

Day C J, Hobbs B F, Pang J-S, Oligopolistic competition in power networks: a conjectured supply function approach, *IEEE Transactions on Power Systems*, **17**(3), 2002, 597–607.

Green R, Increasing competition in the British electricity spot market, *The Journal of Industrial Economics*, **XLIV**(2), 1996, 205–216.

Kirschen D S, A demand-side view of electricity markets, *IEEE Transactions on Power Systems*, **18**(2), 2003, 520–527.

Klemperer P D, Meyer M A, Supply function equilibria in oligopoly under uncertainty, *Econometrica*, **57**(6), 1989, 1243–1277.

Wood A J, Wollenberg B F, *Power Generation, Operation and Control*, Second Edition, John Wiley & Sons, New York, 1996.

4.7 Problems

4.1 Cheapo Electrons is an electricity retailer. The table below shows the load that it forecast its consumers would use over a 6-h period. Cheapo Electrons purchased in the forward market and the power exchange exactly enough energy to cover this forecast. The table shows the average price that it paid for this energy for each hour. As one might expect, the actual consumption of its customers did not exactly match the load forecast and it had to purchase or sell the difference on the spot market at the prices indicated. Assuming that Cheapo Electrons sells energy to its customers at a flat rate of 24.00 \$/MWh, calculate the profit or loss that it made during this 6-h period. What would be the rate that it should have charged its customers to break even?

Period	**1**	**2**	**3**	**4**	**5**	**6**
Load Forecast (MWh)	120	230	310	240	135	110
Average cost (\$/MWh)	22.5	24.5	29.3	25.2	23.1	21.9
Actual load (MWh)	110	225	330	250	125	105
Spot price (\$/MWh)	21.6	25.1	32	25.9	22.5	21.5

4.2 The input–output curve of a gas-fired generating unit is approximated by the following function:

$$H(P) = 120 + 9.3\,P + 0.0025\,P^2\ \text{MJ/h}$$

This unit has a minimum stable generation of 200 MW and a maximum output of 500 MW. The cost of gas is 1.20 \$/MJ. Over a 6-h period, the output of this unit is sold in a market for electrical energy at the prices shown in the table below.

Period	**1**	**2**	**3**	**4**	**5**	**6**
Price (\$/MWh)	12.5	10	13	13.5	15	11

Assuming that this unit is optimally dispatched, is initially on-line and cannot be shut down, calculate its operational profit or loss for this period.

4.3 Repeat the calculation of Problem 4.2 assuming that the cost curve is replaced by a three-segment piecewise linear approximation whose values correspond with those given by the quadratic function for 200 MW, 300 MW, 400 MW and 500 MW.

4.4 Assume that the unit of Problem 4.2 has a start-up cost of \$500 and that it is initially shut down. Given the same prices as in Problem 4.2, when should this unit be brought on-line and when should it be shut down to maximize its operational profit? Assume that dynamic constraints do not affect the optimal dispatch of this generating unit.

4.5 Repeat Problem 4.4, taking into account that the minimum uptime of this unit is four hours.

4.6 Borduria Generation owns three generating units that have the following cost functions:

$$\text{Unit A: } 15 + 1.4\,P_A + 0.04\,P_A^2 \text{ \$/h}$$

$$\text{Unit B: } 25 + 1.6\,P_B + 0.05\,P_B^2 \text{ \$/h}$$

$$\text{Unit C: } 20 + 1.8\,P_C + 0.02\,P_C^2 \text{ \$/h}$$

How should these units be dispatched if Borduria Generation must supply a load of 350 MW at minimum cost?

4.7 How would the dispatch of Problem 4.6 change if Borduria Generation had the opportunity to buy some of this energy on the spot market at a price of 8.20 \$/MWh?

4.8 If, in addition to supplying a 350-MW load, Borduria Generation had the opportunity to sell energy on the electricity market at a price of 10.20 \$/MWh, what is the optimal amount of power that it should sell? What profit would it derive from this sale?

4.9 Repeat Problem 4.8 if the outputs of the generating units are limited as follows:

$$P_A^{MAX} = 100 \text{ MW}$$

$$P_B^{MAX} = 80 \text{ MW}$$

$$P_C^{MAX} = 250 \text{ MW}$$

4.10 Consider a market for electrical energy that is supplied by two generating companies whose cost functions are

$$C_A = 36 \cdot P_A \text{ \$/h}$$

$$C_B = 31 \cdot P_B \text{ \$/h}$$

The inverse demand curve for this market is estimated to be

$$\pi = 120 - D \text{ \$/MWh}$$

Assuming a Cournot model of competition, use a table similar to the one used in Example 4.8 to calculate the equilibrium point of this market (price, quantity, production and profit of each firm).
(Hint: Use a spreadsheet. A resolution of 5 MW is acceptable)

4.11 Write and solve the optimality conditions for Problem 4.10.

4.12 Consider the pumped hydro plant of Example 4.10 and the price profile shown in the table below. Assuming that the operator uses the same strategy as in the example (reservoir initially empty, pumping during four hours of lowest prices and turbining during four hours of highest prices), calculate the profit or loss that this plant would make during this cycle of operation. Determine the value of the plant efficiency that would make the profit or loss equal to zero.

Period	1	2	3	4	5	6
Price ($/MWh)	40.92	39.39	39.18	40.65	45.42	56.34
Period	7	8	9	10	11	12
Price ($/MWh)	58.05	60.15	63.39	59.85	54.54	49.50

5

System Security and Ancillary Services

5.1 Introduction

Markets for electrical energy can function only if they are supported by the infrastructure of a power system. One of the differences with other commodities is that the market participants have no choice: they must use the service provided by the existing system to buy or sell energy. As we saw in Chapter 4, being deprived of electrical energy is extremely inconvenient and costly to consumers. Service interruptions hurt producers to a lesser extent by depriving them of the ability to sell the output of their plants. Users of the system therefore have the right to expect a certain level of continuity in the service provided by the power system. On the other hand, the cost of providing this security of supply should match the value it provides to users.

On a basic level, security means that the power system should be kept in a state in which it can continue operating indefinitely if external conditions do not change. This implies that no component should function outside its safe operating range. For example, no transmission line should be loaded to such an extent that the temperature rise in the conductors due to ohmic losses causes the line to sag low enough to create a fault. Assuming that external conditions will not change is unfortunately very optimistic. In a system that consists of tens of thousands of components, the failure of a single component is not a rare event. This is particularly true if some of these components (such as the transmission lines) are exposed to inclement weather conditions and others (such as the generating plants) are subjected to repeated changes in operating temperature. The cost to society of customer outages is so high that it is universally agreed that power systems should be able to ride through all common disturbances without extensive load disconnections. By this, we mean that the power system should remain stable following any of these common disturbances and that it should be able to continue operating in this new state long enough to give the operator time to restore the system to a normal state. Operators must therefore consider not only the expected evolution of the system but also the consequences of a predefined set of credible contingencies. Typically, the set of credible contingencies contains the outage of all system components (branches, generators and shunt elements) taken separately. The probability of two nearly simultaneous independent

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

faults or failures is generally assumed to be so small that such events do not deserve to be considered.

When preparing to deal with possible contingencies, operators consider both corrective and preventive actions. Preventive measures are designed to put the system in a state such that the occurrence of a credible disturbance does not cause it to become unstable. In practice, this means operating the system at less than its full capacity. From a market perspective, this implies that some transactions will not be possible.

5.1.1.1 Example 5.1

Let us consider the two-generator system shown in Figure 5.1. If both generating units have a capacity of 100 MW, the maximum load that can be handled securely by this system is typically taken to be 100 MW, and not 200 MW as one might have expected. The spare capacity is indeed needed in the event when one of the generating units suddenly fails. A system with more generating units would obviously be able to operate with a much smaller security margin.

Corrective actions are intended to limit the consequences of a disturbance and are taken only if this disturbance occurs. In a traditional environment, all of the resources required to implement corrective actions are under the control of the vertically integrated utility. On the other hand, in a competitive environment, some of these resources belong to other industry participants. They are therefore no longer automatically and freely available to the system operator and must be treated as services that must be purchased on a commercial basis. We shall call these services *ancillary* because they support trading in the main commodity, that is, electrical energy. While some ancillary services result in the delivery of electrical energy, their importance proceeds mostly from the potential to deliver energy or another resource upon request. Consequently, their value should be quantified in terms of their ability to respond when needed. Ancillary services should therefore not be remunerated in terms of energy and cannot be handled as an extension to the energy market. Separate mechanisms must therefore be developed to ensure the provision and the remuneration of these essential services.

In the rest of this chapter, we will first analyze the different types of perturbations that affect power systems and the impact that these perturbations have on system security. On the basis of this analysis, we will describe the types of ancillary services that are needed. We will then discuss how to determine the amount of each service that is required and explore the mechanisms that can be set up to procure them. Finally, we will take the perspective of a provider of ancillary services and investigate how they can be integrated with transactions for electrical energy to maximize operating profits.

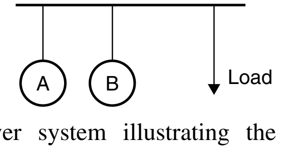


Figure 5.1 Two-generator power system illustrating the limitations that security places on operation

5.2 Describing the Needs

Let us first consider the security issues caused by a global imbalance between load and generation. We will then discuss security problems that arise from the transmission network. This distinction is far from perfect, and on several occasions, we will have to highlight interactions between balancing and network issues.

5.2.1 Balancing issues

When discussing the global balance between load and generation, we can assume that all the loads and generators are connected to the same busbar. In an interconnected system, this busbar is also the terminal of all the tie lines with other regions or countries. At this level of abstraction, the only system variables that remain are the generation, the load, the frequency and the interchanges. As long as the production is equal to the consumption, the frequency and the interchanges remain constant. However, the balance between load and generation is constantly perturbed by fluctuations in the load, by imprecise control of the output of generators and occasionally by the sudden outage of a generating unit or of an interconnection. In an isolated system, a surplus of generation boosts the frequency while a deficit depresses it. The rate at which the frequency changes because of an imbalance is determined by the inertia of all the generators and the rotating loads connected to the system. A local imbalance in an interconnected system affects the flows in the tie lines between the affected region and the rest of the system. Frequency deviations are much less of a problem in an interconnected system because the total inertia increases with the size of the system.

Large frequency deviations can lead to a system collapse. Generating units are indeed designed to operate within a relatively narrow range of frequencies. If the frequency drops too low, protection devices disconnect the generating units from the rest of the system to protect them from damage. Such disconnections exacerbate the imbalance between generation and load, causing a further drop in frequency and additional disconnections. There have also been instances where a system collapsed because protection relays tripped generating units that were exceeding their safe operating speed. The loss of these units caused a deficit of generation that led to a frequency collapse. A large and sudden regional imbalance between load and generation in an interconnected system can cause the disconnection of the tie lines or affect the stability of the neighboring networks. The system operator must therefore take preventive measures to ensure that it can start correcting large imbalances as soon as they arise.

Minor imbalances between load and generation do not represent an immediate security threat because the resulting frequency deviations and inadvertent interchanges are small. However, these imbalances should be eliminated quickly because they weaken the system. A system that is operating below its nominal frequency or in which the tie lines are inadvertently overloaded is indeed less able to withstand a possible further major incident.

The following example illustrates the imbalances that might be observed in an isolated power system.

5.2.1.1 Example 5.2

Figure 5.2(a) shows the load variation observed in the Bordurian power system over five trading periods. This load exhibits random fluctuations superimposed on a slower cyclical evolution. Like all other electricity markets, the Bordurian market makes the simplifying assumption that the demand is constant over each period. Figure 5.2(a) displays a staircase function that illustrates the energy that was traded in the market for each period. This staircase function differs from the actual load in two ways. First, it obviously cannot track the random and cyclical changes in load within each period. Second, if the market were able to predict the load fluctuations with perfect accuracy, the energy traded for each period would be equal to the integral over the period of the instantaneous power demand. In practice, since the market operates on the basis of forecasts that are always inaccurate, the amount traded in the energy market is not an exact average of the actual load. The staircase function also represents the expected total output of the generators. In practice, generators are not able to meet this profile with perfect accuracy. The dashed line in Figure 5.2(a) represents the actual output of the generating units that sold through the energy market. In addition to some minor discrepancies during each period, there are also differences at the transitions between periods. Because of limits on the rate at which units can adjust their output, generators are unable to achieve the idealized production profile that results from market trading. In our example, a much more severe imbalance between scheduled generation and load develops in the middle of period 4. This imbalance is due to the sudden outage of a large generating unit.

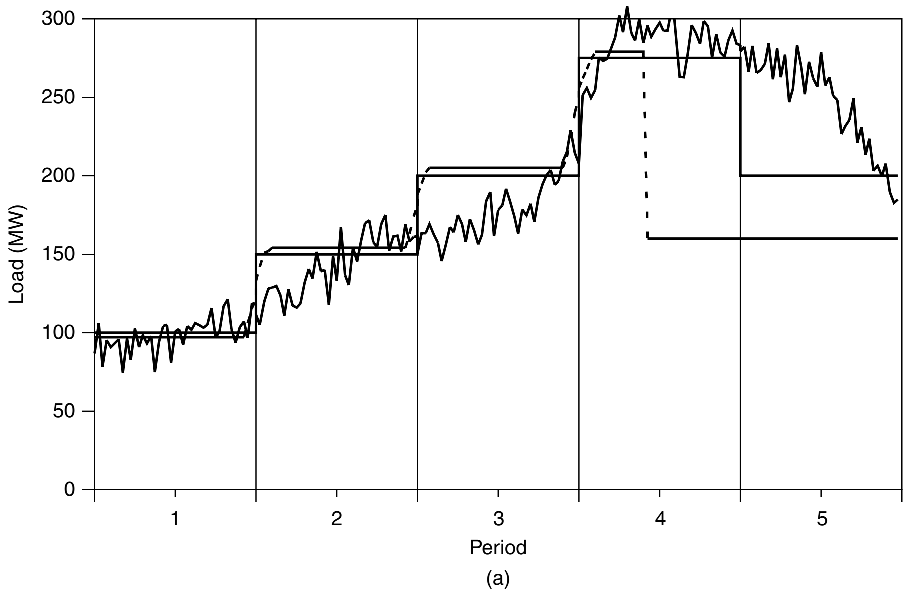
Figure 5.2(b) shows the difference between the actual production of the units scheduled through the energy market and the load. The shape of this curve illustrates that imbalances between load and generation have three components with different time signatures: rapid random fluctuations, slower cyclical fluctuations and occasional large deficits. A smoothed version of the load fluctuation has been added to the figure to highlight the slower cyclical variations.

As the previous example shows, several phenomena create imbalances between load and generation in a competitive electricity market. Since each of these phenomena causes a component of imbalance with a different "time signature", it is better to treat them separately. The system operator can then tailor the different ancillary services it needs to cope with a specific component of the total imbalance.

The *regulation* service is designed to handle rapid fluctuations in loads and small unintended changes in generation. This service helps maintain the frequency of the system at or close to its nominal value and reduce inadvertent interchanges with other power systems. Generating units that can increase or decrease their output quickly will typically provide this service. These units must be connected to the grid and must be equipped with a governor. They will usually be operating under automatic generation control.

Generating units providing the *load-following* service handle the slower fluctuations, in particular, the intraperiod changes that the energy market does not take into account. These units obviously must be connected to the system and should have the ability to respond to these changes in load.

Regulation and load-following services require more or less continuous action from the generators providing these services. However, regulation actions are relatively small



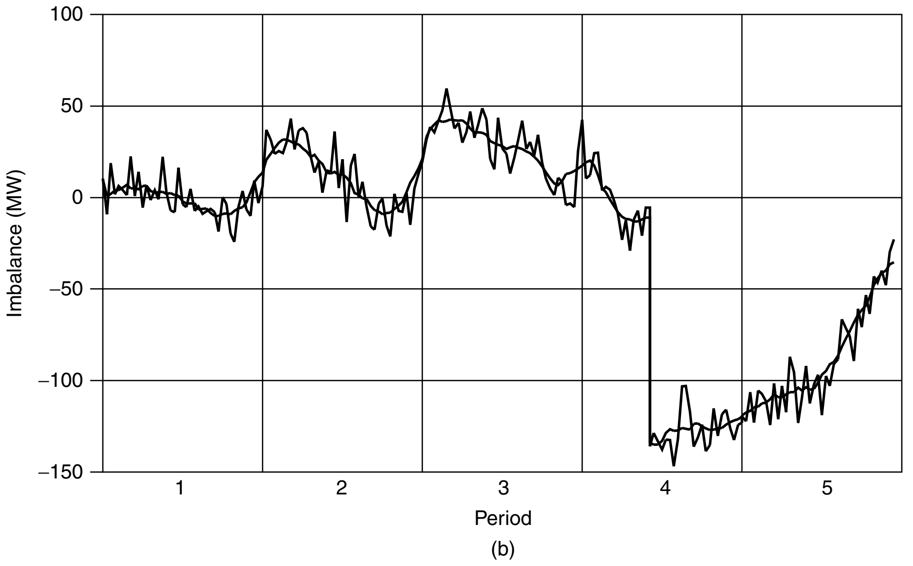


Figure 5.2 (a) Typical load and generation fluctuations over five market periods, (b) Imbalances resulting from these fluctuations

and load-following actions are fairly predictable. By keeping the imbalance close to zero and the frequency close to its nominal value, these services are used as preventive security measures. On the other hand, *reserve* services are designed to handle the large and unpredictable power deficits that could threaten the stability of the system. Reserve services are used to provide corrective actions. Obtaining reserve services, however, can be considered as a form of preventive security action.

Reserve services are usually classified into two categories. Units that provide *spinning reserve* must start responding immediately to a change in frequency, and the full amount of reserve capacity that they are supposed to contribute must be available very quickly. On the other hand, generating units providing *supplemental reserve services* do not have to start responding immediately. Depending on local rules, some forms of supplemental reserve services may be provided by units that are not synchronized to the grid but can be brought online quickly. In some cases, customers who agree to have their load disconnected during emergencies can also provide reserve services. Besides the speed and rate of response, the definition of reserve services must also specify the amount of time during which generating units must be able to sustain this response. All these parameters vary considerably depending on the reliability criteria and on the size of the system. For example, preventing unacceptable frequency deviations in a small isolated system requires faster acting reserves than in a large interconnected system.

It would be nice if we could draw a clear distinction between balancing ancillary services and balancing energy, which is traded in the spot energy market. Unfortunately, the wide variety of designs among electricity markets makes an unambiguous classification impossible. In general, if the time that elapses between the closure of the open market and the real time is short, the system operator is able to buy a substantial portion of its balancing needs in the spot energy market. On the other hand, if the market operates on a day-ahead basis, a complex mechanism is likely to be needed for the procurement of balancing services.

The rate at which the output of a generating unit can be adjusted is obviously the most important factor in determining its ability to provide balancing services. In some cases, however, its location in the network may affect its ability to provide these services. A generating plant that is connected to the "main" part of the system through a transmission corridor that is often congested would not be a suitable candidate to provide these services. Its ability to increase its output could indeed be limited by these transmission constraints.

5.2.1.2 Example 5.3

Figure 5.3 illustrates the frequency response of a power system following a major generation outage and the response of the reserve services. This example is based on an actual incident. On 15 August 1995 at 12 : 25 : 30, 1220 MW of generation was suddenly disconnected from the power system of Great Britain. This system has a total installed capacity of about 65 GW but does not have ac interconnections with any other system. It is therefore prone to significant frequency fluctuations. The two main categories of reserve ancillary services that have been defined for the operation of this system reflect this characteristic. *Primary response* must be fully available within 10 s and sustainable for a further 20 s. *Secondary response* must be fully available within 30 s of the incident and must be sustainable for a further 30 min. As can be seen from the figure, primary response succeeded in arresting the frequency drop before it reached the statutory limit of 49.5 Hz. Secondary response then helped bring the system frequency closer to its nominal value. However, in this case, gas turbines, which were started at 12 : 29 : 20, produced the increase in frequency that we observe at the extreme right of the graph.

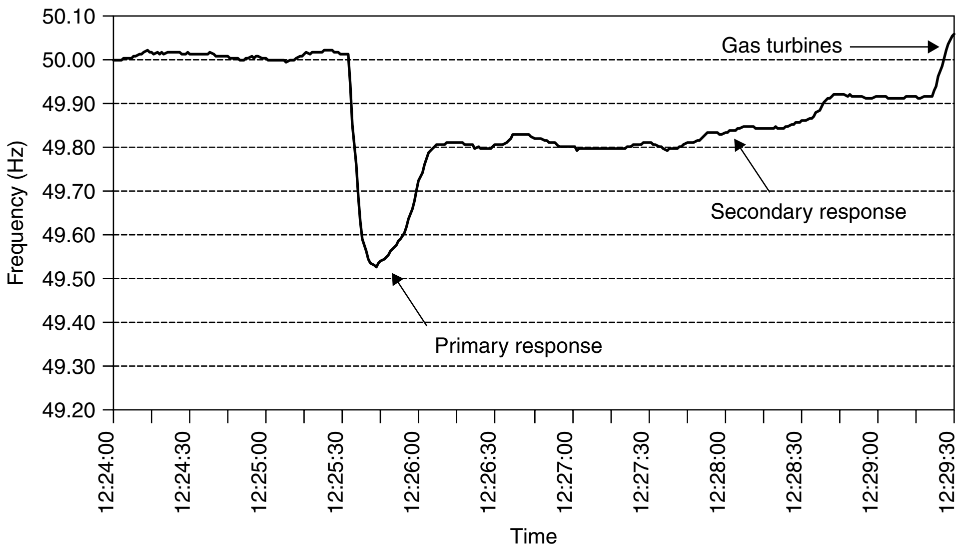


Figure 5.3 Example of frequency and reserve response following a major generation outage

5.2.2 Network issues

5.2.2.1 Limits on power transfers

In a real power system, consumers and producers are distributed over a wide geographical area and connected by a network. As loads and generations vary, the flows in the branches and the voltages at the nodes of the network fluctuate. The system operator must therefore consider the effect of these changes on security. Besides continuously checking that no equipment is being operated outside its safe operating range, the operator periodically performs a computerized contingency analysis. This analysis takes as its starting point the current state of the power system and checks that no credible contingency would destabilize the system. Depending on the nature of the power system, this destabilization can take several forms:

- Following the outage of a branch, the power that was carried by that branch reroutes itself through the network. In this postcontingency state, one or more other branches may be loaded beyond their thermal capacity. Unless the system operator can correct this situation quickly, overloaded lines will sag, cause a fault and be disconnected. Similarly, overloaded transformers may be taken out of service to prevent heat-related damage. These additional outages further weaken the network and may lead to a system collapse as more and more branches become overloaded.
- The sudden outage of a generating unit or of a reactive compensation device can deprive the system of essential reactive support. Similarly, the outage of an important branch can increase the reactive losses in the network beyond what the system can provide. The voltage in a region or even in the entire network may then collapse.

- A fault in a heavily loaded line may cause the rotor angle of some generators to increase so much that a portion of the network dynamically separates from the rest, causing one or both regions to collapse because generation and load are no longer balanced.

When the state of the system is such that a credible contingency would trigger any one of these types of instabilities, operators must act by taking preventive actions.

Implementing some types of preventive actions involves a cost that is either very small or negligible. For example, operators can increase the margin to voltage collapse by adjusting the transformer taps and the voltage set point of generators or by switching in or out banks of capacitors and reactors. They can also reduce the potential for postcontingency overloads by rerouting active power flows using phase-shifting transformers. While these low-cost preventive measures can be very effective, there is a limit to the contribution they can make to the security of the system. As the loading of the system increases, there comes a point where security can be maintained only by placing restrictions on the flow of active power on some branches. These restrictions constrain the amount of power that can be produced by generating units located upstream from the critical branches and prevent them from producing all the energy that they could sell in the market. Limitations on active power flows thus carry a very real and often very significant cost.

5.2.2.2 Example 5.4

Let us consider the two-bus power system shown in Figure 5.4, and quantify the amount of power that the generating unit located at bus A is able to sell to the load connected at bus B. Limitations imposed by the thermal capacity of the lines are the easiest ones to calculate. If each line is designed to be able to carry 200 MW continuously without overheating, the maximum amount of power that the load at bus B can obtain from unit A is limited to 200 MW. The spare 200 MW of transmission capacity must be kept in reserve in case a fault occurs and one of the lines must be disconnected. This very substantial security margin can be reduced if we consider the possibility of postcontingency corrective action. Let us suppose that either line can withstand a 10% overload for 20 min without sagging and causing another fault and without damage to the conductors. If the system operator can obtain from the generating unit at bus B that it will increase its output by 20 MW in 20 min if necessary, the maximum amount of power that can be transmitted from bus A to bus B can be raised up to 220 MW.

In order to calculate the effect of transient stability on the maximum power that can be transmitted from A to B, we need more information about the system. To avoid unnecessary complications, we will assume that bus B behaves like an infinite bus and

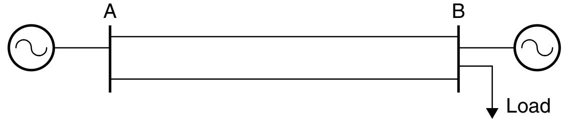


Figure 5.4 Two-bus power system used to illustrate the limitations that network security places on the operation of the system

that the generator at bus A has an inertia constant H of 2 s, and can be modeled as a constant voltage behind a transient reactance X' of 0.9 p.u. The reactance of each line is equal to 0.3 p.u. The voltage at both buses is kept constant at 1.0 p.u. The worst contingency in this system is a fault in one of the lines close to bus A. We will assume that such a fault would be cleared in 100 ms by tripping of the faulted line. Using a transient stability program, it is easy to compute that, under these conditions, the maximum power that can be transmitted from A to B without endangering the transient stability of the system is 108 MW.

Let us now consider how voltage instability might limit the power transfer from A to B. Again, to avoid unnecessary complications, we will adopt a very simple system model and assume that the point of voltage collapse is reached when the power flow stops converging. This assumption gives us a good first approximation of the maximum flow that the system can handle. In case a more accurate measure of voltage stability is needed, more complex analysis techniques have been developed.

The amount of reactive support at bus B has a strong influence on the transfer capacity. Let us first consider the case in which no voltage support is available because the generator at bus B has reached its upper MVAr limit. Using a power flow program, we can calculate that when both lines are in service, 198 MW can be transmitted from A to B before the voltage at B drops below the usual 0.95 p.u. limit. However, if the power transfer exceeds 166 MW and one of the lines is disconnected, the voltage collapses. On the other hand, if 25 MVAr of reactive support is available at bus B, the power transfer can be increased up to 190 MW before a line outage would cause a voltage collapse.

In this example, transient stability places the most severe restriction on the maximum power transfer between A and B. In practical systems, the maximum allowable power transfer would be determined using considerably more sophisticated models and would require a significant amount of computing resources. Such limits have a profound effect on the structure and operation of markets for electrical energy. We will consider this issue in detail in the next chapter.

5.2.2.3 Voltage control and reactive support services

The previous example also shows how the operator can use reactive power resources to increase the amount of power that can be transferred from one part of the network to another. Some of these reactive resources and voltage control devices (e.g. mechanically switched capacitors and reactors, static VAR compensators, tap-changing transformers) are typically under the direct control of the operator and can be used at will. Generating units, however, provide the best way to control voltage. A *voltage control service* therefore needs to be defined to specify the conditions under which the system operator can make use of the resources owned by the generating companies. Generators providing this service produce or absorb reactive power in conjunction with their active power production. It is also conceivable that businesses might be setup for the sole purpose of selling reactive support or voltage control.

The definition of a voltage control service must consider not only the operation of the system under normal conditions but also the possibility of unpredictable outages. Under normal operating conditions, operators use reactive power resources to maintain the voltage at all buses within a relatively narrow range around the nominal voltage. Typically, this range is

$$0.95 \text{ p.u.} \leq V \leq 1.05 \text{ p.u.} \tag{5.1}$$

Keeping transmission voltages within this range is partially justified by the need to facilitate voltage regulation in the distribution network. It also makes the operation of the transmission system more secure. Maintaining the voltage at or below the upper limit reduces the likelihood of insulation failures. The lower limit is more arbitrary. In general, keeping voltages high under normal condition makes it more likely that the system would avoid a voltage collapse if an unpredictable outage does occur. A good voltage profile, however, does not guarantee the voltage security of the system. The outage of a heavily loaded transmission line increases the reactive losses in the remaining lines. If these losses cannot be supplied, the voltage collapses. The amount of reactive power needed following an outage is therefore much larger than what is required during normal operation. Voltage control services should therefore be defined not only in terms of the ability to regulate the voltage during normal operation but also to provide reactive power in case of emergency. The voltage control service is in fact often called *reactive support service*.

5.2.2.4 Example 5.5

Using again a power flow program, we can explore the nature of the voltage control or reactive support service using a two-bus example similar to the one shown in Figure 5.4. Each of the transmission lines in this system is modeled using the π-equivalent circuit shown in Figure 5.5. The load at bus B has a power factor of unity. Let us first examine how the operator might control the voltage at bus B using the reactive capability of the generator at this bus. We assume that the voltage at bus A is kept constant at its nominal value. Figure 5.6 shows that when the amount of power transferred from bus A to bus B is small, the reactive power produced by the equivalent shunt capacitances of the lines exceeds the reactive power consumed in the equivalent series reactances. The generator at bus B must absorb this excess to keep the voltage at the upper limit of the acceptable range. When the amount of power transferred is between 100 and 145 MW, the reactive power balance is such that the voltage stays naturally within the acceptable limits. A reactive injection is not needed at bus B for these conditions. When the power transfer exceeds 145 MW, the reactive losses in the lines must be compensated by a reactive injection at bus B to keep the voltage from dropping below the lower limit.

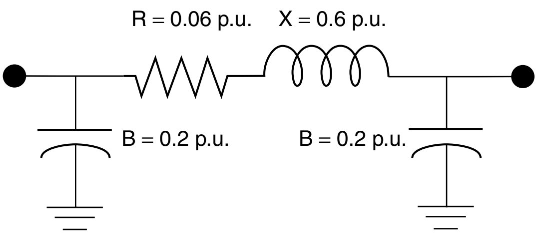


Figure 5.5 π-model of the transmission lines in Example 5.5

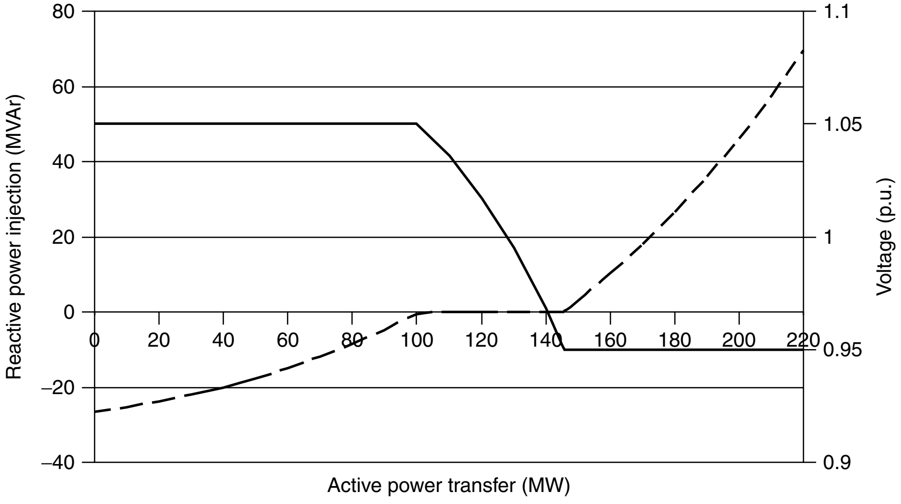


Figure 5.6 Variation in the reactive power injection and the voltage at bus B of the two-bus system in Example 5.5. The voltage (solid line) and the reactive injection needed to keep this voltage within normal limits (dashed line) are plotted as a function of the power transferred from bus A

If the generator connected to bus B is disconnected or asks too high a price for regulating the voltage at bus B, the system operator could attempt to control it by adjusting the voltage set point of the generator at bus A. When the amount of power transferred is small, the voltage at bus B is high. To keep it below its upper limit, the voltage set point of the generator at bus A must be lowered. This implies that reactive power must be absorbed by this generator. Table 5.1 shows that when 49 MW is transferred, the voltage at B is at its upper limit and the voltage at A is at its lower limit. A lower power transfer could therefore not be accommodated. On the other hand, when the power transfer is high, the voltage set point of generator A must be increased to keep the voltage at B above its lower limit. Table 5.1 shows that when this power transfer reaches 172.5 MW, the voltage at A is at its upper limit and the voltage at B is at its lower limit. A power transfer smaller than 49.0 MW or larger than 172.5 would therefore cause a violation of a voltage limit at either bus A or bus B. Further reactive power injections at bus A are pointless outside of this range of power transfers. We can therefore conclude that local control of the voltage is much more effective than remote control, even under normal operating conditions.

As we have already mentioned above, the true value of reactive support services, however, does not reside in the actual production of VARs but in the ability to supply

Table 5.1 Limits on the control of the voltage at bus B using the voltage set point of the generator at bus A

Power transfer (MW)	V_B (p.u.)	V_A (p.u.)	Q_A (MVAr)
49.0	1.05	0.95	−68.3
172.5	0.95	1.05	21.7

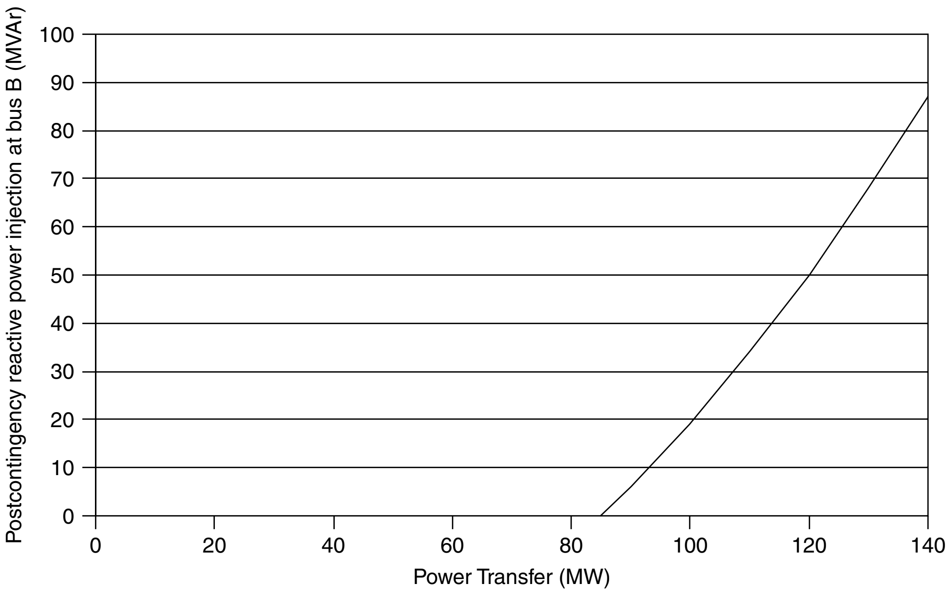


Figure 5.7 Postcontingency reactive support requirement at bus B following the outage of one of the lines connecting buses A and B

reactive power and prevent a voltage collapse following an outage. A power flow program can provide a rough estimate of the amount of reactive power that must be injected after an outage to prevent a voltage collapse. A more precise calculation of the need for reactive power reserves requires the consideration of dynamic effects. Figure 5.7 shows how much reactive power must be injected at bus B to prevent a voltage collapse following the outage of one of the two lines of our two-bus system. Before the contingency, the voltage at bus A is kept at its nominal value by the generator connected to that bus. This graph shows that the system can withstand a line outage without reactive support at B when the power transfer is smaller than 85 MW. However, the postcontingency reactive support requirement increases rapidly beyond that threshold.

Figure 5.8 illustrates the precontingency and postcontingency reactive power balances for the case in which 130 MW is transferred from A to B. The generator at B maintains the voltage at its nominal value before and after the outage. Under precontingency conditions, the lines produce about 25 MVAr, which must be absorbed by the generator at bus A. The active power losses are about 3 MW. After the contingency, both generators must inject reactive power in the remaining line to prevent a voltage collapse. Instead of producing reactive power, the line now consumes 107 MVAr. On the other hand, the active power losses increase only to 15 MW.

5.2.2.5 Stability services

Some system operators may also need to obtain other network security services from generators. For example, *intertrip schemes* can mitigate transient stability problems.

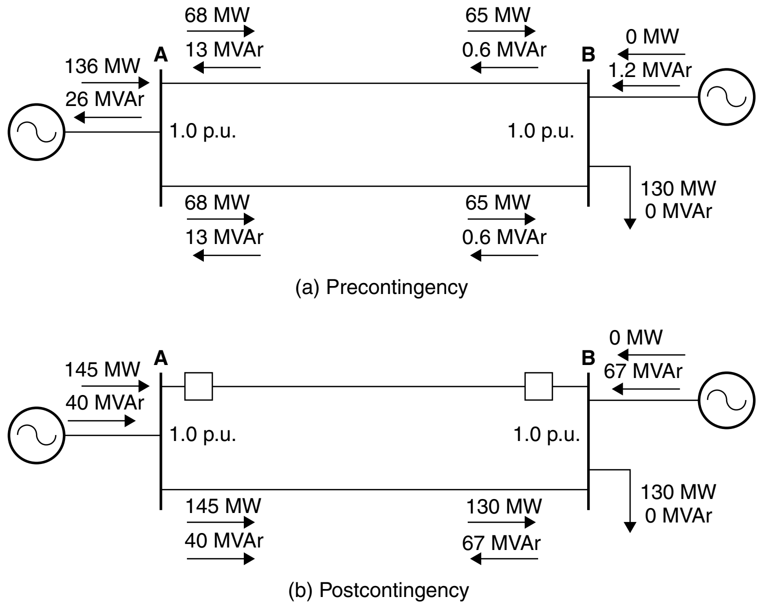


Figure 5.8 Precontingency and postcontingency active and reactive power flows in the two-bus system

These schemes have no effect on the current state of the power system but in the event of a fault, they automatically disconnect some generation and/or some load to maintain the stability of the system. Similarly, *power system stabilizers* make minute adjustments to the output of generators to dampen oscillations that might develop in the network. The action of these stabilizers increases the amount of power that can be transmitted.

5.2.3 System restoration

Despite the best efforts of the system operator, a disturbance occasionally spirals out of control and the entire power system collapses. It is then the responsibility of the system operator to restore the system to its normal operating state as soon as possible. However, restarting large thermal generating plants requires a significant amount of electric power, which is not available if the entire system has collapsed. Fortunately, some types of generators (e.g. hydroplants, and small diesel generators) are able to restart either manually or using energy stored in batteries. The system operator must ensure that enough of these restoration resources are available to guarantee a prompt restoration of service at any time. This ancillary service is usually called *black-start capability*.

5.3 Obtaining Ancillary Services

In the previous section, we saw that the system operator needs some resources to maintain the security of the system and that some of these resources must be obtained from other industry participants in the form of ancillary services. At this point, we need to examine the two mechanisms that can be used to ensure that the system operator obtains the amount of ancillary services that are required. The first approach consists in making the provision of some ancillary service compulsory. The second entails the creation of a market for ancillary services. As we will see, both approaches have advantages and disadvantages. The choice of one mechanism over the other is influenced not only by the type of ancillary service but also by the nature of the power system and historical circumstances.

5.3.1 Compulsory provision of ancillary services

In this approach, as a condition for being allowed to connect to the power system, a category of industry participants is required to provide a certain type of ancillary service. For example, connection rules may require all generating units to

- be equipped with a governor with a 4% droop coefficient. This requirement ensures that all units contribute equally to frequency regulation;
- be capable of operating at a power factor ranging from 0.85 lead to 0.9 lag, and be equipped with an automatic voltage regulator. This forces all units to participate in voltage regulation and contribute to voltage stability.

This approach represents the minimum deviation from the practice of vertically integrated utilities. It also guarantees that enough resources will be available to maintain the security of the system. While compulsion is apparently simple, it is not necessarily a good economic policy, and presents certain implementation difficulties:

- These mandates may cause unnecessary investments and produce more resources than what is actually needed. For example, not all generating units need to take part in frequency control to maintain the security of the system. Similarly, not all generating units need to be equipped with a power system stabilizer to dampen system oscillations.
- This approach does not leave room for technological or commercial innovation. New and more efficient ways of providing a service are unlikely to be developed by industry participants or sought by the system operator if traditional providers are compelled to offer this service.
- Compulsion tends to be unpopular among providers because they feel that they are forced to supply a service that adds to their costs without being remunerated. For example, generators complain that producing reactive power increases the losses in the synchronous machine and sometimes reduces the amount of active power that they are able to produce and sell.
- Some participants may be unable to provide some services or may be unable to provide them cost effectively. Nuclear units, for example, are unable to provide services that demand rapid changes in active power output. Highly efficient units should not be forced to operate at part load so they can provide reserve. It is

considerably cheaper to determine centrally how much reserve is needed and to schedule a few marginal or extra-marginal units to provide this reserve. Compulsion is, therefore, not applicable for all services and even for those services where it seems appropriate, some participants may need exemptions. Such exemptions may be seen as distorting competition.

5.3.2 Market for ancillary services

Given the economic disadvantages and the practical difficulties of compelling participants to provide ancillary services, it is usually considered desirable to set up a market mechanism for the procurement of at least some ancillary services. The preferred form of this mechanism depends on the nature of the service. Long-term contracts are preferable for services in which the amount needed does not change or changes very little over time, and for services in which the availability is determined mostly by equipment characteristics. Black-start capability, intertrip schemes, power-system stabilizers and frequency regulation are typically procured under long-term contracts. On the other hand, a spot market is needed for services in which the needs vary substantially over the course of the day, and the offers change because of interactions with the energy market. For example, at least part of the necessary reserve services is often procured through a short-term market mechanism. However, the system operator will often seek to reduce the risk of not having enough reserve capacity or of having to pay too much for this capacity by arranging some long-term contracts for the provision of reserve. In a mature market, providers of reserve services should also find desirable a mixture of short- and long-term contracts.

Markets provide a more flexible and hopefully more economically efficient mechanism for the procurement of ancillary services than compulsion. However, it is not clear at this point if a market-based approach can be applied to all ancillary services. In some cases, the number of participants that are actually able to provide a certain ancillary service is so small that the potential for abuse of market power precludes procurement on a competitive basis. For example, in some remote parts of a transmission network, there may be only one generating unit that can effectively support the voltage by providing reactive power in case of emergencies. A reactive power market would, therefore, need to be strictly controlled to avoid possible abuses.

5.3.3 Demand-side provision of ancillary services

Before the introduction of competition in the supply of electricity, generating units owned by the vertically integrated utilities provided virtually all the ancillary services. Unfortunately, the definitions of ancillary services in many electricity markets still reflect this practice. In a truly competitive environment, the system operator should have no obligation or incentive to favor generators in the procurement of ancillary services as long as other providers are able to deliver services of the same quality. Encouraging consumers to offer ancillary services has several advantages. First, a larger number of providers should increase competition in the markets for ancillary services. Second, from a global economic perspective, the provision of ancillary services by the demand side improves the utilization of the resources. For example, if interruptible loads provide some of the reserve requirements, some of the generation capacity does not have to be held back. Generating units can then be used for producing electrical energy, which is what they were designed for. If the mix of generation technologies continues to evolve toward a combination of large inflexible units and renewable generation, resources for system control may have to come from the demand side. Finally, the demand side may be a more reliable supplier of some ancillary services than large generating units. The probability that the demand side may fail to deliver a critical service on time is indeed smaller. This service would be provided by the combination of a large number of relatively small loads, all of which are much less likely to fail at the same time than a large generating unit.

The demand side is probably most competitive in the provision of the different types of reserve services. Some consumers (for example, those who have large water pumping loads equipped with variable speed drives) might also be able to compete for the provision of regulation.

5.4 Buying Ancillary Services

We argued at the beginning of this chapter that the purpose of ancillary services is to maintain the security of the system in the face of unpredictable events. Security is a "system" concept that must be centrally managed. The system operator is thus responsible for purchasing security on behalf of the users of the system. If we assume that a market mechanism has been adopted for the procurement of ancillary services, then this system operator will have to pay the providers of these services. It will then have to recover this cost from the users. Since the amount of money involved is not negligible, these users are likely to scrutinize this purchasing process. They need to be convinced that the optimal amount of services is purchased, that the right price is paid and that each user pays its fair share of the cost of ancillary services.

5.4.1 Quantifying the needs

Ideally, the level of security provided through the purchases of ancillary services should be determined through a cost/benefit analysis. This analysis would set this level at the optimal point where the marginal cost of providing more security is equal to the marginal value of this security. While the marginal cost is relatively easy to calculate, the marginal value, which represents mostly the expected cost to consumers of load disconnections that are avoided through the provision of system security, is much harder to compute. Since performing a cost/benefit analysis in every case is not practical, security standards that approximate the optimal solution have been developed. These standards usually specify the contingencies that the system must be able to withstand. Sophisticated models and computational tools have been developed to help system operators manage the power system in accordance with these standards and to quantify the ancillary services that they need to achieve this goal. A discussion of these techniques is beyond the scope of this book. The interested reader is encouraged to consult (Billinton and Allan, 1996) for an exhaustive discussion of the techniques used for calculating reserve requirements. A method for determining and allocating the needs for reactive support is described by Pudjianto *et al.* (2002).

If the cost of running the system is simply passed on to the users, system operators may be tempted to purchase more ancillary services than are strictly needed. Having a bigger pool of resources to call upon in case of difficulties makes operating the system easier and less stressful. It is therefore desirable to develop an incentive scheme that encourages the system operator not only to minimize the cost of purchasing ancillary services but also to limit the amount of services purchased to what is truly necessary to maintain security.

5.4.2 Co-optimization of energy and reserve in a centralized electricity market

Setting the price for ancillary services at the right level is not easy because the procurement of a particular ancillary service often cannot be decoupled from the procurement of electrical energy or other related services. In the early years of competitive electricity markets, this issue was not fully understood. Energy and each type of reserve were traded in separate markets. These markets were cleared successively in a sequence determined by the speed of response of the service. For example, the market for primary reserve would be cleared first, followed by the market for secondary reserve and finally by the energy market. The idea was that resources that had not been successful in one market could then be offered in other markets where the performance requirements are not as demanding. Bids that were successful in one market would not be considered in the subsequent ones. Experience showed that this approach led to problems. It has since been abandoned. See (Oren, 2002) for more details on these problems.

There is now a wide consensus that energy and reserve should be offered in joint markets and that these markets should be cleared simultaneously to minimize the overall cost of providing electrical energy and reserve. This co-optimization is necessary because of the strong interaction between the supply of energy and the provision of reserve. To get a more intuitive understanding of this interaction, consider that to provide spinning reserve, generators must operate part-loaded. This mode of operation has several consequences:

- Part-loaded generators cannot sell as much energy as they might otherwise do;
- To meet the demand, other generators, which are generally more expensive, have to produce more energy;
- The efficiency of the generators that provide spinning reserve may be less than it would be if they were running at full load. These generators therefore may need to be paid more for the energy that they provide.

Meeting the reserve requirements will therefore increase the price of electrical energy. In the rest of this section, we use simple examples to discuss how co-optimization in a centralized electricity market minimizes this additional cost while ensuring that no generator is at a disadvantage when being asked to provide reserve rather than produce electrical energy. For a more detailed discussion of this topic, see Read *et al.* (1995).

5.4.2.1 Example 5.6

Let us consider a small electricity market where the demand varies between 300 and 720 MW. For the sake of simplicity, we will assume that only one type of reserve is needed and that 250 MW of this reserve is required to maintain security for all loading conditions. Four generators are connected to this system. Table 5.2 shows their relevant characteristics.

We observe that these generators are assumed to have a constant marginal cost and that they are ranked in decreasing order of merit. While they have similar capacities, their respective abilities to provide reserve are quite different. Units 1 and 4 cannot provide any reserve that meets the requirements set by the system operator. On the other hand, the amount of reserve that units 2 and 3 can provide is limited not only by their capacity but also by their ability to respond. Figure 5.9 shows how much reserve they can provide as a function of the amount of electrical energy they produce. We ignore all limitations and complications caused by the minimum stable generation of the units.

We will assume that this market operates on a centralized model, that the generators' bids to produce electrical energy are equal to their marginal costs and that the market

Table 5.2 Marginal cost, maximum output and reserve capability of the generating units of Example 5.6

Generating Units	Marginal cost of energy ($/MWh)	P^{max} (MW)	R^{max} (MW)
1	2	250	0
2	17	230	160
3	20	240	190
4	28	250	0

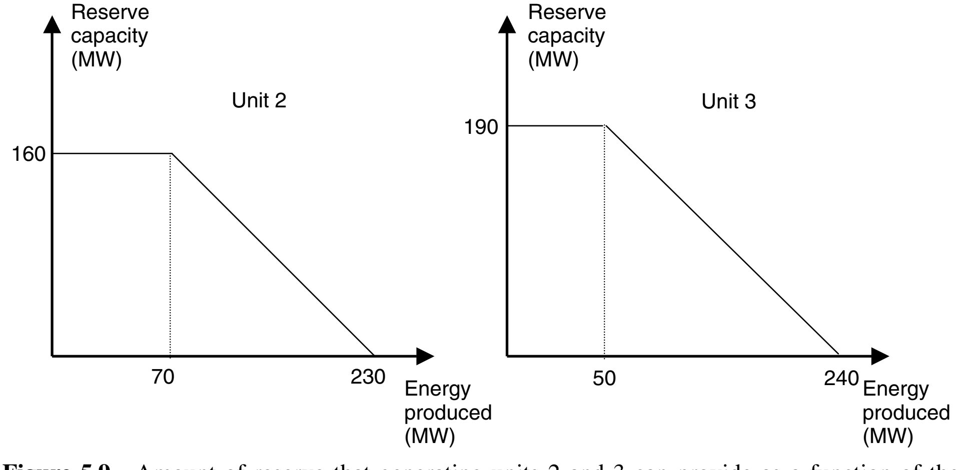


Figure 5.9 Amount of reserve that generating units 2 and 3 can provide as a function of the amount of electrical energy that they produce

rules do not include separate bids for the provision of reserve. This last assumption is reasonable if the generators do not incur a direct cost when providing reserve. We will relax this assumption in the next example. To clear the market, the operator must determine the dispatch that minimizes the cost of production (as measured by the bids) while respecting the operational constraints. Formally, this problem can be expressed as follows:

Find the power produced by each of the four generating units (P_1, P_2, P_3 and P_4) and the amount of reserve provided by these same units (R_1, R_2, R_3 and R_4) that minimize

$$2 \cdot P_1 + 17 \cdot P_2 + 20 \cdot P_3 + 28 \cdot P_4 \tag{5.2}$$

subject to the following constraints:

Balance between production and demand:

$$P_1 + P_2 + P_3 + P_4 = D \tag{5.3}$$

Minimum reserve requirement:

$$R_1 + R_2 + R_3 + R_4 \geq 250 \tag{5.4}$$

Limits on the production of the generating units:

$$\begin{aligned} 0 \leq P_1 \leq 250 \\ 0 \leq P_2 \leq 230 \\ 0 \leq P_3 \leq 240 \\ 0 \leq P_4 \leq 250 \end{aligned} \tag{5.5}$$

Limits on the reserve capabilities of the generating units:

$$\begin{aligned} R_1 &= 0 \\ 0 \leq R_2 &\leq 160 \\ 0 \leq R_3 &\leq 190 \\ R_4 &= 0 \end{aligned} \tag{5.6}$$

Limits on the capacity of the generating units:

$$\begin{aligned} P_1 + R_1 \leq 250 \\ P_2 + R_2 \leq 230 \\ P_3 + R_3 \leq 240 \\ P_4 + R_4 \leq 250 \end{aligned} \tag{5.7}$$

Any linear programming package can easily solve this problem. Table 5.3 shows the results for values of the demand D ranging from 300 to 720 MW. In addition to

Table 5.3 Solution of the optimization problem of Example 5.6 for a range of values of the demand. Each line in this table corresponds to a subrange where the output of only one of the units changes

Demand (MW)	P_1 (MW)	R_1 (MW)	P_2 (MW)	R_2 (MW)	P_3 (MW)	R_3 (MW)	P_4 (MW)	R_4 (MW)
300–420	250	0	50–170	60	0	190	0	0
420–470	250	0	170	60	0–50	190	0	0
470–720	250	0	170	60	50	190	0–250	0

finding the optimal dispatch of energy and reserve, such a package also calculates the dual variables or Lagrange multipliers associated with each constraint. The Lagrange multiplier associated with the constraint on the production–demand balance gives the marginal cost of producing electrical energy. Similarly, the multiplier associated with the minimum reserve requirement constraint gives the marginal cost of providing reserve. In a centralized market, these marginal costs are deemed to be the market clearing prices for electrical energy and reserve, respectively.

For this simple example, we can easily check these solutions by hand and get a better understanding of the physical meaning of the price of reserve and its evolution as the demand changes.

Given that the minimum load we consider is 300 MW and that unit 1 has the lowest marginal operating cost and cannot provide reserve, we conclude immediately that this unit must produce its maximum output of 250 MW for all values of the demand. Since units 2 and 3 are the only ones that can provide reserve and since unit 2 can provide at most 160 MW, unit 3 has to provide at least 90 MW. Given that this unit has a capacity of 240 MW, its energy production must be less than 150 MW.

$$0 \leq P_3 \leq 150 \tag{5.8}$$

Similarly, since unit 3 can provide at most 190 MW of reserve, unit 2 has to provide at least 60 MW. It energy output is thus limited to 170 MW:

$$0 \leq P_2 \leq 170 \tag{5.9}$$

For a demand in the range between 300 and 420 MW, unit 2 is the marginal generator. It produces between 50 and 170 MW, that is, the power that is not supplied by unit 1, which runs at its full output of 250 MW. The marginal cost of unit 2 sets the price for energy at 17 $/MWh. In this range of demand, the inequality constraint for the minimum reserve requirement is not binding because units 2 and 3 provide more than enough reserve. The price of reserve is thus zero.

As the demand increases from 420 to 470 MW, the production of unit 2 is capped at 170 MW because it must provide at least 60 MW of reserve. Unit 3 becomes the marginal generator and progressively increases its output from 0 to 50 MW. The price of energy is thus set at the marginal cost of unit 3, which is 20 $/MWh. To determine the price of reserve, we must figure out where an additional megawatt of reserve would come from and how much it would cost. Figure 5.9 shows that over this range of output, unit 3 provides 190 MW of reserve, which is the maximum it is able to deliver under any circumstance. To get an additional megawatt of reserve beyond the basic 250 MW requirement, we would have to reduce the output of unit 2 by one megawatt. Instead of producing 170 MW, it would thus produce only 169 MW. To compensate for this reduction, we would have to increase the output of unit 3 by one megawatt. This extra megawatt from unit 3 would cost \$20, while the reduction in the output of unit 2 would save \$17. The net cost of getting an additional megawatt of reserve and hence the price of reserve, is thus $20 - 17 = 3$ \$/MWh.

As the demand increases from 470 to 720 MW, the marginal producer is unit 4 and its production increases from 0 to 250 MW. The reserve constraint keeps the energy productions of units 2 and 3 at 170 and 50 MW respectively. In this range, the price of energy is thus 28 \$/MWh.

The price of reserve increases to 11 \$/MWh. This is because in order to make one more megawatt of reserve available, we need to reduce the energy output from unit 2 by 1 MW and increase the production of unit 4 by the same amount. The cost of this marginal redispatch sets the price of the reserve at $28 - 17 = 11$ \$/MWh.

Figure 5.10 summarizes the prices of energy and reserve for the various ranges of demand.

Let us now examine the revenues collected by each generating unit, the costs that they incur and the profits that they achieve by producing energy and providing reserve. This analysis is not particularly interesting in the case of unit 1 because it always operates at full output and sells its energy at a price determined by the marginal costs of other generators. Since its own marginal cost is always lower than this price, it always makes a healthy operating profit.

In the range of demand between 350 and 420 MW, the market price for energy is equal to the bid price of generating unit 2. Since we assume that all generators bid at their marginal cost of production, this unit does not make an economic profit on the sale of energy. Given that the reserve price is zero, it does not make a profit

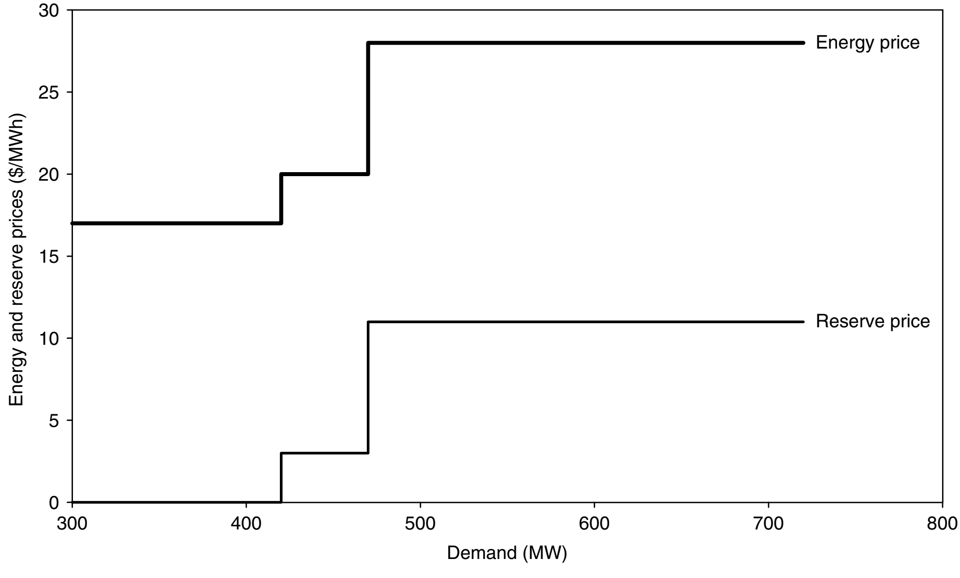


Figure 5.10 Energy and reserve prices for the conditions of Example 5.6

on the provision of reserve either. On the other hand, when the demand is in the range between 420 and 470 MW, even though unit 2 has a lower marginal cost than other units, its output is capped at 170 MW by the reserve requirement constraint. Unit 3 is then the marginal energy producer. The energy price jumps from 17 to 20 \$/MWh, which means that unit 2 makes a 3 \$/MWh profit on each megawatt-hour it produces. At first glance, one might think that the owner of unit 2 is unfairly treated since the reserve constraint prevents it from selling the additional 60 MWh of energy it could sell because it bid at a lower price than unit 3. Observe, however, that the price of reserve in this range of demand is 3 \$/MWh and that unit 2 provides 60 MW of reserve. The revenue it collects for the reserve it provides is thus exactly equal to the opportunity cost of not selling energy. The owner of unit 2 is therefore indifferent to producing more electrical energy or providing reserve. In this same range of demand, unit 3 does not make an economic profit from the sale of energy because it is the marginal producer. On the other hand it makes a profit of 3 \$/MWh from the provision of reserve because the marginal provider of reserve is unit 2.

When the demand increases beyond 470 MW, unit 4 becomes the marginal producer and sets the energy price at 28 \$/MWh. Unit 2 thus makes a profit of 11 \$/MWh on each of the 170 MW that it is allowed to produce. Its owners do not mind the limitation that the reserve constraint puts on their unit's output because they also make an 11 \$/MWh profit on every megawatt of reserve it provides. Unit 2 is still the marginal provider of reserve in the range of demand. On the other hand, unit 3 makes a profit of 8 \$/MWh on its energy production and a profit of 3 \$/MWh on the reserve it provides because it is marginal neither for energy nor for reserve.

Figure 5.11 summarizes the revenues that unit 2 derives from the energy and reserve markets as well as its cost and profit.

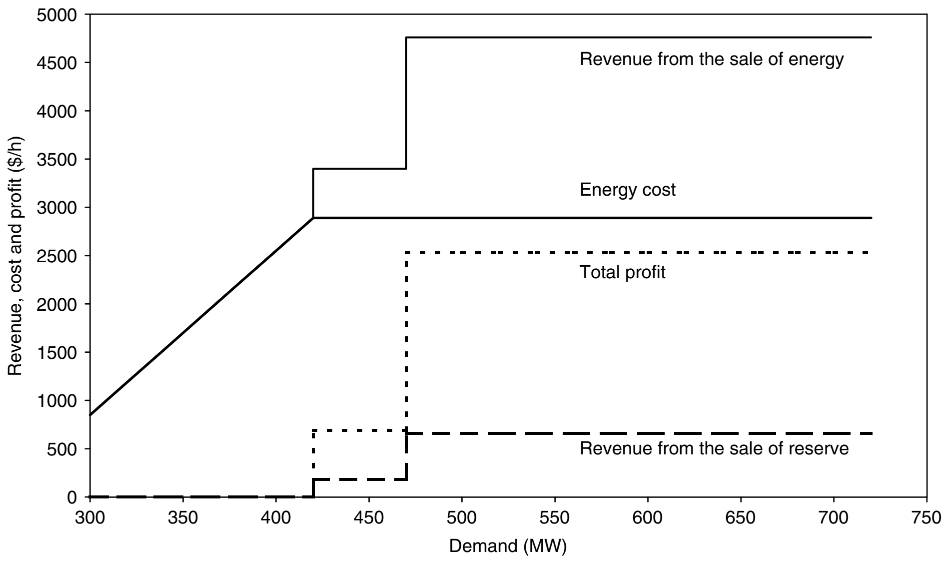


Figure 5.11 Revenues, cost and profit of unit 2 of Example 5.6 for a range of demand

5.4.2.2 Example 5.7

Let us assume that the rules of the market that we considered in our previous example are changed to take into consideration the costs that generators must bear when they provide reserve. These costs may reflect the loss in efficiency of units that operate part-loaded or the additional maintenance costs that the provision of reserve may require. Generators are thus allowed to submit separate bids in the reserve market. In a less than perfectly competitive market, these bids would not reflect the marginal cost of providing reserve, but would reflect the value that generators believe the market places on the reserve they provide. We also assume that unit 4 can now provide a maximum of 150 MW of reserve. Table 5.4 shows the bids that the generators have submitted as well as the relevant unit characteristics.

Given that the generators now bid explicitly to provide reserve, the objective function of the optimization problem that the market operator has to solve becomes

$$\min(2 \cdot P_1 + 17 \cdot P_2 + 20 \cdot P_3 + 28 \cdot P_4 + 0 \cdot R_1 + 0 \cdot R_2 + 5 \cdot R_3 + 7 \cdot R_4) \quad (5.10)$$

The constraints remain the same as in Example 5.6, except for the constraint on the maximum reserve that unit 4 can provide

$$0 \leq R_4 \leq 150 \quad (5.11)$$

Table 5.5 summarizes the dispatch for the conditions of this example, while Figure 5.12 shows the evolution of the energy and reserve prices.

Let us analyze this solution. When the demand is in the range between 300 and 320 MW, unit 1 produces at its maximum capacity of 250 MW while unit 2 produces the rest of the demand and is thus the marginal generator. The price of energy is thus 17 \$/MWh. Unit 2 provides the maximum amount of reserve that it can deliver (160 MW) because it is willing to provide this reserve at no cost. Unit 3 provides the remainder of the reserve requirement and is thus the marginal provider. The price of reserve is thus 5 \$/MWh. Unit 2 makes a profit of \$5.00 per megawatt-hour of reserve while unit 3 just covers its cost of providing reserve.

In the range of demand extending from 320 to 470 MW, the output of unit 2 is kept at 70 MW so that it can provide 160 MW of reserve. Unit 3 is the marginal producer of energy and sets the price at 20 \$/MWh. Unit 3 is also the marginal provider of reserve and the price of reserve thus stays at 5 \$/MWh. Unit 2 makes a 3 \$/MWh profit on the

Table 5.4 Marginal energy and reserve costs, maximum output and reserve capability of the generating units of Example 5.7

Generating units	Marginal cost of energy (\$/MWh)	Marginal cost of reserve (\$/MWh)	P^{max} (MW)	R^{max} (MW)
1	2	0	250	0
2	17	0	230	160
3	20	5	240	190
4	28	7	250	150

energy it sells and a profit of 5 \$/MWh on the reserve it provides. It thus benefits from being limited in the amount it can sell in the energy market.

If the demand increases from 470 to 560 MW, unit 2 continues to produce 70 MW and to provide 160 MW of reserve. Unit 3 is the marginal producer of electrical energy. As its energy production increases, its contribution to reserve must go down. Unit 4 compensates for this reduction. In this case, the price of energy is not equal to the marginal cost of unit 3 because increasing the energy production of unit 3 has an impact on the allocation of reserve. Producing an additional megawatt with unit 3 costs 20 \$/MWh, but reduces its contribution to the reserve by the same amount, thereby saving 5 \$/MWh. This megawatt of reserve is then provided by unit 4 at a cost of 7 \$/MWh. The price of energy is thus $20 - 5 + 7 = 22$ \$/MWh, which is not equal to the marginal cost of production of any individual generator. Unit 3 thus earns 2 \$/MWh above its marginal cost for the energy it produces. The price of reserve is 7 \$/MWh because the marginal provider is unit 4. Observe that this solution is indeed optimum. Keeping the output of unit 3 at 150 MW so that it can provide 90 MW of reserve would require an increase in the energy production of unit 4. This approach

Table 5.5 Solution of the optimization problem of Example 5.7 for a range of values of the demand. Each line in this table corresponds to a subrange where the output of only one of the units changes

Demand (MW)	P_1 (MW)	R_1 (MW)	P_2 (MW)	R_2 (MW)	P_3 (MW)	R_3 (MW)	P_4 (MW)	R_4 (MW)
300–320	250	0	50–70	160	0	90	0	0
320–470	250	0	70	160	0–150	90	0	0
470–560	250	0	70	160	150–240	90–0	0	0–90
560–620	250	0	70–130	160–100	240	0	0	90–150
620–720	250	0	130	100	240	0	0–100	150

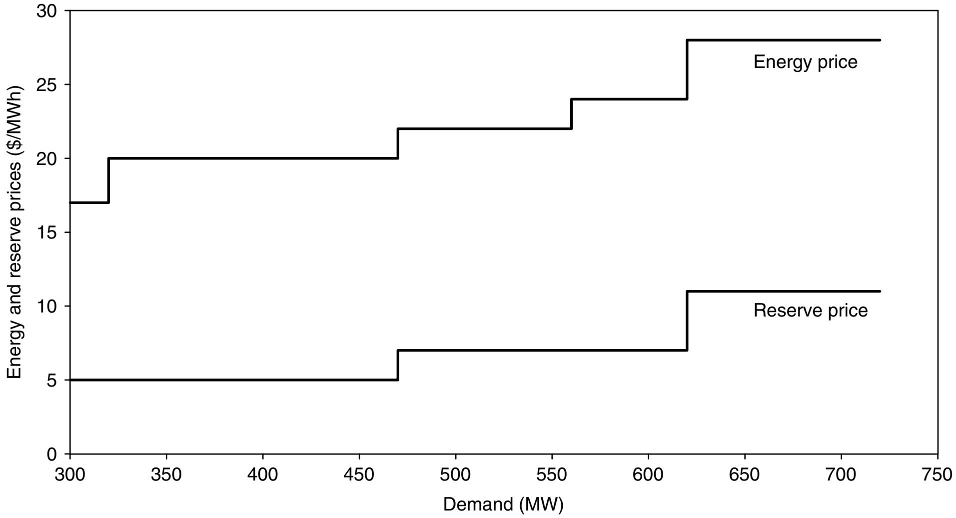


Figure 5.12 Energy and reserve prices for the conditions of Example 5.7

would be more expensive because the extra energy produced would cost 8 $/MWh while the saving in the provision of reserve would only be 2 $/MWh.

For the range of demand between 560 and 620 MW, unit 3 produces its maximum capacity of 240 MW and therefore cannot provide any reserve. Interestingly, unit 2 becomes again the marginal producer of energy. However, the price of energy is not 17 but 24 $/MWh. While producing an additional megawatt with unit 2 costs 17 $/MWh, this reduces its contribution to the reserve by the same amount. However no money is saved because unit 2 is willing to provide reserve for free. The compensating megawatt of reserve is provided by unit 4 at a cost of 7 $/MWh. The price of energy is thus $17 + 7 = 24$ $/MWh. The price of reserve remains 7 $/MWh because the marginal provider is unit 4.

Finally, for a demand greater than 620 MW but smaller than 720 MW, units 2 and 3 produce 130 and 240 MW respectively. Unit 4 is the marginal producer of energy, while unit 2 is the marginal provider of reserve. The price of energy is thus 28 $/MWh. The price of reserve is 11 $/MWh because to get one additional megawatt of reserve we must reduce the production of unit 2 (thereby saving 17 $/MWh) and increase the output of unit 4 (at a cost of 28 $/MWh).

These two examples show that it is possible to clear the energy and reserve markets simultaneously in a way that minimizes the cost to consumers, meets the security requirements but also ensures a fair treatment of all the providers of energy and reserve services.

5.4.3 Allocating the costs

Not all consumers value system security equally. For example, the cost of a service interruption is much larger for a semiconductor factory or a paper mill than it is for residential customers. Some consumers might therefore be willing to pay more for an improved level of security while others would accept a less reliable system in exchange for a reduction in the price they pay for their supply of electricity. Such reliability-based pricing would be economically efficient. Unfortunately, the current state of technology does not enable the system operator to deliver differentiated levels of security. The security standards that it applies must therefore reflect an average level of security that is hoped to be at least acceptable to all. Since all users get the same level of security, it seems logical to share the cost of the ancillary services among all users on the basis of some measure of their use of the system. This measure is typically the energy consumed or produced.

There is, however, another aspect to this issue. The behavior of some users may cause a disproportionate amount of stress in the power system. Penalizing these users might encourage them to change their behavior. Ultimately, this change in behavior should decrease the amount of ancillary services needed and reduce the cost of achieving the desired level of security. Let us explore this concept using two examples.

5.4.3.1 Who should pay for reserve?

Reserve generation capacity is intended to avert a system collapse when a large imbalance develops between load and generation. In most cases, such imbalances originate from the sudden failure of a generating unit or the sudden disconnection of an interconnection with a neighboring system. If such a contingency occurs when the system does not carry enough reserve capacity, the system operator must resort to shedding load to avoid a complete system collapse. Using historical data on the failure rate of generating units and interconnections, it is possible to calculate the amount of reserve required to reduce the probability of load disconnection to an acceptably low level (see for example, Billinton and Allan, 1996). These probabilistic calculations confirm that a system in which generating units fail more frequently requires more reserve than a system in which generators are more reliable. They also show that a system supplied by a few large generating units needs more reserve than one with many smaller generators. The unreliability of a few large generating plants can therefore increase the need for operating reserve. Since our objective is to minimize the cost of reserve services without reducing the level of security, we should give these generators an incentive to reduce their failure rates. If after some time they can demonstrate that they have succeeded in improving their performance, the system operator will be able to reduce the required amount of reserve. Strbac and Kirschen (2000) have argued that the fairest incentive involves charging the cost of the reserve services to the generators in proportion to their contribution to the reserve requirement. Generators would obviously pass on this cost to their consumers in the form of higher prices for electrical energy. Smaller and reliable generating units would then have a competitive advantage over larger and failure-prone units.

5.4.3.2 Who should pay for regulation and load following?

Kirby and Hirst (2000) have analyzed the requirements for load following and regulation services in an actual and typical power system. They have also developed an equitable technique for allocating these requirements between industrial and nonindustrial consumers. For this particular power system, their analysis shows that industrial consumers account for 93% of the regulation and 58% of the load-following requirements even though they represent only 34% of the system load. Since the cost of these services is charged to consumers on the basis of their energy consumption, the residential consumers are clearly subsidizing the industrial ones. It can be shown that there are also wide variations between the contributions of individual consumers within the industrial group. For example, aluminum smelters and paper mills have loads that are nearly time invariant and therefore do not contribute to the requirements for regulation and load-following.

5.5 Selling Ancillary Services

Selling ancillary services represents another business opportunity for generating companies. Technical limitations and cost considerations, however, inextricably link the sale of reserve and voltage control services and the sale of energy. For example, a generator cannot sell spinning reserve or reactive support if the unit is not running and producing at least its minimum power output. Conversely, a unit operating at maximum capacity cannot sell reserve capacity because it does not have any. If it decides to reduce its power output to be able to sell reserve, then it forgoes an opportunity to sell energy. Since the cost of this opportunity can be significant, the generating company must jointly optimize the sale of energy and reserve services.

Rather than attempting to develop a general formulation of this obviously complicated problem, let us explore these interactions using a simple example.

Let us consider the operation of a generating unit that can sell both energy and spinning reserve in competitive markets. The precise characteristics of this spinning reserve service are not important for our analysis and we will not consider the possibility of selling other ancillary services. We will assume that the energy and reserve markets are sufficiently competitive that this unit can be treated as a price taker. This means that its bidding behavior has no effect on either the price of energy or the price of reserve and that it is able to sell any quantity it chooses in either market. We will consider the operation of this unit over a single market period of one hour and we will assume that the unit is running at the beginning of the period. These assumptions allow us to neglect issues related to the start-up cost of the unit, its minimum uptime and its minimum downtime. In an actual application, the optimization would be carried out over a day or longer, and all these issues would have to be taken into account.

We will use the following notations:

π_1: price per MWh on the energy market

π_2: price per MWh of capacity on the spinning reserve market. A MWh corresponds to 1 MW of reserve capacity made available for one hour. Since this reserve capacity may or may not be called, a MWh of reserve is not equivalent to a MWh of energy. For the sake of simplicity, we will assume that the generator does not receive an additional exercise fee when the reserve it provides is actually called upon to provide energy. Considering such an exercise fee would not change the conclusions of this example.

x_1: Quantity bid by the generator in the energy market. Since this generator is a price taker, it is also the quantity of energy sold by the generator.

x_2: Quantity bid by the generator in the reserve market. Since this generator is a price taker, it is also the quantity of reserve sold by the generator.

$P^{\min}$: Minimum power output of the generating unit (minimum stable generation).

$P^{\max}$: Maximum power output of the generating unit.

$R^{\max}$: Upper limit that the ramp rate of the unit and the definition of the reserve service place on the amount of reserve that the unit can deliver. For example, if the unit has a maximum ramp rate of 120 MW per hour and the reserve has to be delivered within 10 min, this unit cannot deliver more than 20 MW of reserve.

$C_1(x_1)$: Cost of producing the amount x_1 of energy. This function must be convex. It includes the costs of fuel and maintenance related to the production of energy but does not include any investment costs.

$C_2(x_2)$: Cost of providing the amount x_2 of reserve. This function must also be convex. It does not include the opportunity cost of selling energy or any investment costs. We assume that the generator can estimate the fraction of the reserve it offers that will be called upon to provide energy. The expected cost of producing this energy is included in this cost.

Let us formulate this example as a constrained optimization problem. Since this generator is trying to maximize the profit it derives from the sale of energy and reserve, the objective function is the difference between the revenues and the costs from both energy and reserve:

$$f(x_1, x_2) = \pi_1 x_1 + \pi_2 x_2 - C_1(x_1) - C_2(x_2) \tag{5.12}$$

Several technical factors place constraints on the energy and reserve that can be provided by this unit. First, the sum of the bids for energy and reserve cannot exceed the maximum power output of the generating unit:

$$x_1 + x_2 \leq P^{\max} \tag{5.13}$$

Second, since the unit cannot operate below its minimum stable generation, the bid for energy should be at least equal to the minimum power output:

$$x_1 \geq P^{\min} \tag{5.14}$$

Third, the unit cannot bid for more reserve than it can deliver within the time allowed by the specification of the reserve service:

$$x_2 \leq R^{\max} \tag{5.15}$$

If $R^{\max} \geq P^{\max} - P^{\min}$, the amount of reserve that the unit can provide is not limited by the ramp rate and condition (5.15) is superfluous. We will therefore assume that $R^{\max} < P^{\max} - P^{\min}$. This restriction implies that constraints (5.13) and (5.14) cannot be binding at the same time. We are not modeling explicitly the fact that reserve cannot be negative. Doing so would complicate our analysis without bringing additional insight. Some generators may obviously decide that, at least part of the time, providing reserve is not worthwhile.

Given this objective function and these constraints, we can form the Lagrangian function for this optimization problem:

$$\begin{aligned} \ell(x_1, x_2, \mu_1, \mu_2, \mu_3) = {} & \pi_1 x_1 + \pi_2 x_2 - C_1(x_1) - C_2(x_2) \\ & + \mu_1(P^{\max} - x_1 - x_2) + \mu_2(x_1 - P^{\min}) + \mu_3(R^{\max} - x_2) \end{aligned} \tag{5.16}$$

Setting the partial derivatives of this Lagrangian with respect to the decision variables equal to zero, we obtain the conditions for optimality:

$$\frac{\partial \ell}{\partial x_1} \equiv \pi_1 - \frac{\mathrm{d}C_1}{\mathrm{d}x_1} - \mu_1 + \mu_2 = 0 \tag{5.17}$$

$$\frac{\partial \ell}{\partial x_2} \equiv \pi_2 - \frac{\mathrm{d}C_2}{\mathrm{d}x_2} - \mu_1 - \mu_3 = 0 \tag{5.18}$$

The solution must also satisfy the inequality constraints

$$\frac{\partial \ell}{\partial \mu_1} \equiv P^{\max} - x_1 - x_2 \geq 0 \tag{5.19}$$

$$\frac{\partial \ell}{\partial \mu_2} \equiv x_1 - P^{\min} \geq 0 \tag{5.20}$$

$$\frac{\partial \ell}{\partial \mu_3} \equiv R^{\max} - x_2 \geq 0 \tag{5.21}$$

and the complementary slackness conditions

$$\mu_1 \cdot (P^{\max} - x_1 - x_2) = 0 \tag{5.22}$$

$$\mu_2 \cdot (x_1 - P^{\min}) = 0 \tag{5.23}$$

$$\mu_3 \cdot (R^{\max} - x_2) = 0 \tag{5.24}$$

$$\mu_1 \geq 0; \mu_2 \geq 0; \mu_3 \geq 0 \tag{5.25}$$

The complementary slackness conditions assert the fact that an inequality constraint is either binding or nonbinding. If it is binding, it behaves like an equality constraint and it can be shown that the corresponding Lagrange multiplier μ_i is equal to the marginal or shadow cost of the constraint. Since a binding constraint always increases the cost of the optimal solution, the Lagrange multipliers of binding inequality constraints must be positive. On the other hand, since a nonbinding inequality constraint has no impact on the cost of the optimal solution, its Lagrange multiplier is equal to zero. Binding inequality constraints are thus associated with strictly positive Lagrange multipliers and vice versa. We will make repeated use of this observation in the discussion that follows.

Equations (5.17) to (5.25) form a set of necessary and sufficient optimality conditions for this problem. They are called the *Karush Kuhn Tucker* (KKT) *conditions*. Unfortunately the KKT conditions do not tell us which inequality constraints are binding. Software packages for optimization try various combinations of binding constraints until they find one that satisfies the KKT conditions. We will examine all the possible combinations because each of them illustrates a different form of interaction between the energy and reserve markets. Since there are three inequality constraints in this problem, we have to consider eight possible combinations.

Case 1: $\mu_1 = 0; \mu_2 = 0; \mu_3 = 0$

Since all the Lagrange multipliers are equal to zero, none of the constraints are binding. Equations (5.17) and (5.18) simplify to

$$\frac{\mathrm{d}C_1}{\mathrm{d}x_1} = \pi_1 \tag{5.26}$$

$$\frac{\mathrm{d}C_2}{\mathrm{d}x_2} = \pi_2 \tag{5.27}$$

These conditions mean that the generating unit will bid to provide energy and reserve up to the point at which their respective marginal costs are equal to their price. Since there are no interactions between energy and reserve, this situation is similar to the sale of energy in a perfectly competitive market, as we described in Chapter 4.

Case 2: $\mu_1 > 0; \mu_2 = 0; \mu_3 = 0$

The generation capacity of the unit is fully utilized by the provision of a combination of energy and reserve:

$$x_1 + x_2 = P^{\max} \tag{5.28}$$

Replacing the values of the Lagrange multipliers in Equations (5.17) and (5.18), we get

$$\pi_1 - \frac{\mathrm{d}C_1}{\mathrm{d}x_1} = \pi_2 - \frac{\mathrm{d}C_2}{\mathrm{d}x_2} = \mu_1 \geq 0 \tag{5.29}$$

Equation (5.29) shows that the provision of energy and reserve are both profitable. Maximum profit is achieved when the unit is dispatched in such a way that the marginal profit on energy is equal to the marginal profit on reserve. The value of the Lagrange multiplier μ_1 indicates the additional marginal profit that would be achieved if the upper limit on the unit's output could be relaxed.

Case 3: $\mu_1 = 0; \mu_2 > 0; \mu_3 = 0$

The unit produces just enough energy to operate at its minimum stable generation:

$$x_1 = P^{\min} \tag{5.30}$$

Equations (5.17) and (5.18) give

$$\frac{\mathrm{d}C_1}{\mathrm{d}x_1} - \pi_1 = \mu_2 \tag{5.31}$$

$$\frac{\mathrm{d}C_2}{\mathrm{d}x_2} = \pi_2 \tag{5.32}$$

In order to be able to provide spinning reserve, the unit must be running and operating at least at its minimum stable generation. Equation (5.32) shows that this unit should provide reserve up to the point at which the marginal cost of providing reserve is equal to the market price for reserve. On the other hand, since the Lagrange multipliers of binding constraints are positive, Equation (5.31) indicates that the production of energy is marginally unprofitable. If it were possible, the generator would prefer to produce less energy.

Note that the KKT conditions determine the operating point that will maximize the profit... or minimize the loss! They do not guarantee that the generator will actually make a profit. In this case, the loss on the sale of energy might exceed the profit on the sale of reserve. To check if an operating point is actually profitable, we would have to replace the values of x_1 and x_2 in the objective function given in Equation (5.12) and check the sign of the result. If an optimal operating point turns out to be unprofitable, the generator might decide to turn off the unit for that hour. However, when the operation of a unit is optimized over a number of periods (e.g. over a day), the overall optimal solution may include some unprofitable periods because of the start-up costs and the minimum time constraints. The sale of reserve may reduce the loss that must be accepted during these unprofitable periods.

Case 4: $\mu_1 > 0; \mu_2 > 0; \mu_3 = 0$ and Case 5: $\mu_1 > 0; \mu_2 > 0; \mu_3 > 0$

Since we assume that the ramp rate limit on reserve is smaller than the operating range of the unit, these cases are not physical and we will not discuss them further.

Case 6: $\mu_1 = 0; \mu_2 = 0; \mu_3 > 0$

The only binding constraint in this case is that the reserve is limited by the ramping rate. We have

$$x_2 = R^{\max} \tag{5.33}$$

$$\frac{dC_1}{dx_1} = \pi_1 \tag{5.34}$$

$$\pi_2 - \frac{dC_2}{dx_2} = \mu_3 \geq 0 \tag{5.35}$$

These equations show that while the profit from the sale of energy is maximized, relaxing the ramp rate constraint would increase the profit from the sale of reserve.

Case 7: $\mu_1 > 0; \mu_2 = 0; \mu_3 > 0$

Both the maximum capacity and the ramp rate constraint are binding:

$$x_1 + x_2 = P^{\max} \tag{5.36}$$

$$x_2 = R^{\max} \tag{5.37}$$

We can rewrite Equation (5.36) as follows:

$$x_1 = P^{\max} - R^{\max} \tag{5.38}$$

The optimality conditions (5.17) and (5.18) give respectively the marginal profitability for energy and reserve:

$$\pi_1 - \frac{dC_1}{dx_1} = \mu_1 \tag{5.39}$$

$$\pi_2 - \frac{dC_2}{dx_2} = \mu_1 + \mu_3 \tag{5.40}$$

Since both μ_1 and μ_3 are positive, Equations (5.39) and (5.40) show that selling more energy and more reserve would be profitable. However, since the marginal profit on the sale of reserve is higher than on the sale of energy, all the capacity of the unit is not devoted to the sale of energy. There is no point in reducing output by more than $R^{\max}$ because of the ramp rate constraint.

Case 8: $\mu_1 = 0; \mu_2 > 0; \mu_3 > 0$

In this case, the values of both x_1 and x_2 are determined by the binding inequality constraints:

$$x_1 = P^{\min} \tag{5.41}$$

$$x_2 = R^{\max} \tag{5.42}$$

Once again, we can use the optimality conditions to determine the marginal profitability of both transactions:

$$\pi_1 - \frac{dC_1}{dx_1} = -\mu_2 \tag{5.43}$$

$$\pi_2 - \frac{dC_2}{dx_2} = \mu_3 \tag{5.44}$$

These equations indicate that the sale of reserve is profitable and would be even more so if it were not for the ramp rate constraint. On the other hand, the sale of energy is unprofitable and would be further reduced if it were not for the minimum stable generation constraint. Once again, the actual profitability of this operating point should be checked using the objective function.

5.6 Further Reading

The concepts of power system security and the application of cost/benefit analysis to the determination of the optimal level of security are discussed by Kirschen (2002). Ejebe *et al.* (2000) describes computationally efficient methods for calculating the available transmission capacity. Billinton and Allan (1996) wrote a standard reference on power system reliability, which explains the techniques used to calculate the amount of reserve needed to achieve a certain level of reliability. A method for calculating the reactive support requirements was proposed by Pudjianto *et al.* (2002). The reader interested in the definition of the various ancillary services should consult the report of the Interconnected Operations Services Working Group (1997) or the paper by Hirst and Kirby (1997). Useful information on the ancillary services used in a particular power system can often be found on the web site of the system operator. For example, the types and amounts of ancillary services required in England and Wales are described on the National Grid Company web site. Oren (2002) compares sequential and simultaneous approaches to clearing energy and reserve markets. Read *et al.* (1995) discuss the co-optimization of energy and reserve and the implementation of this approach in the New Zealand electricity market. Alvey *et al.* (1998) give the detailed formulation of this approach. The allocation of the cost of ancillary services to the parties that are responsible for the need for these services is discussed by Strbac and Kirschen (2000) for the case of reserve and by Kirby and Hirst (2000) for the case of frequency regulation and load-following. The provision of ancillary services by the demand side is again addressed by Kirby and Hirst (1999).

Alvey T, Goodwin D, Ma X, Streiffert D, Sun D, A security-constrained bid-clearing system for the New Zealand wholesale electricity market, *IEEE Transactions on Power Systems*, **13**(2), 1998, 340–346.

Billinton R, Allan R. N, *Reliability Evaluation of Power Systems*, Second Edition, Kluwer Academic/Plenum Publishers, Boston, 1996.

Ejebe G. C, Waight J. G, Santos-Nieto M, Tinney W. F, Fast calculation of linear available transfer capability, *IEEE Transactions on Power Systems*, **15**(3), 2000, 1112–1116.

Hirst E, Kirby B, *Creating Competitive Markets for Ancillary Services*, Report ORNL/CON-448, Oak Ridge National Laboratory, Oak Ridge, TN, October 1997

Interconnected Operations Services Working Group, *Defining Interconnected Operations Services Under Open Access*, EPRI Report TR-108097, Electric Power Research Institute, Palo Alto, CA, May 1997

Kirby B, Hirst E, Load as a resource in providing ancillary services, *American Power Conference*, Chicago, IL, April 1999, On-line: *www.ornl.gov/psr*.

Kirby B, Hirst E, *Customer-specific Metrics for the Regulation and Load-following Ancillary Services*, Report ORNL/CON-474, Oak Ridge National Laboratory, Oak Ridge, Tennessee, January 2000, On-line: www.ornl.gov/psr

Kirschen D. S, Power system security, *IEE Power Engineering Journal*, **16**(5), 2002, 241–248.

National Grid Company plc, Industry Information on Balancing Services, www.nationalgrid.com/uk/indinfo/balancing/mn_commercial.html.

Oren, S. S, Auction design for ancillary reserve products, *IEEE Power Engineering Society Summer Meeting*, Vol. **3**, 2002, pp. 1238–1239.

Pudjianto D, Ahmed S, Strbac G, Allocation of VAr support using LP and NLP based optimal power flows, *IEE Proceedings on Generation, Transmission and Distribution*, **149**(4), 2002, pp. 377–383.

Read E. G, Drayton-Bright G. R, Ring B. J, An integrated energy and reserve market for New Zealand, in *Transmission Pricing and Technology*, R Siddiqi and M Einhorn (eds), Kluwer, Boston, 1995, pp. 183–206.

Strbac G, Kirschen D. S, Who should pay for reserve? *The Electricity Journal*, **13**(8), 2000 pp. 32–37.

5.7 Problems

5.1 A power system is supplied by three generating units that are rated at 150, 200 and 250 MW, respectively. What is the maximum load that can be securely connected to this system if the simultaneous outage of two generating units is not considered to be a credible event?

5.2 Identify the documents describing the security criteria governing the operation of the power system in the region where you live or some other region of your choice. Summarize the main points of these security rules.

5.3 A small power system consists of two buses connected by three transmission lines. Assuming that this power system must be operated according to the N-1 security criterion and that its operation is constrained only by thermal limits on the transmission lines, calculate the maximum power transfer between these two buses for each of the following conditions

a. All three lines are in service and each line has a continuous thermal rating of 300 MW

b. Only two lines rated at 300 MW are in service

c. All three lines are in service. Two of them have a continuous thermal rating of 300 MW and the third is rated at 200 MW.

d. All three lines are in service. All of them have a continuous thermal rating of 300 MW. However, during emergencies, they can sustain a 10% overload for 20 min. The generating units on the downstream bus can increase their output at the rate of 4 MW per minute.

e. Same conditions as in (d), except that the output of the downstream generators can only increase at the rate of 2 MW per minute.

f. Low temperatures and high winds improve the heat transfer between the conductors and the atmosphere. Assume that this dynamic thermal rating increases the continuous and emergency loadings of (d) by 15%.

5.4 A generator is connected to a large power system by a double circuit transmission line. Each line has a negligible resistance and a reactance of 0.2 p.u. The transient reactance of the generator is 0.8 p.u. and its inertia constant is 3 s. The large power system can be modeled as an infinite bus and the bus voltages are kept at their nominal value. Assume that single circuit faults on the transmission line are cleared in 120 ms. Using a transient stability program, calculate the maximum power that this generator can produce without risking instability. Use a 100 MW base.

5.5 Repeat the calculations of Problem 5.4 for the case in which the generator is connected to the power system by two identical double circuit transmission lines.

5.6 Consider a power system with two buses and two lines. One of these lines has a reactance of 0.25 p.u. and the other a reactance of 0.40 p.u. The series resistances and shunt susceptances of the lines are negligible. A generator at one of the buses maintains its terminal voltage at nominal value and produces power that is consumed by a load connected to the other bus. Using a power flow program, calculate the maximum active power that can be transferred without causing a voltage collapse when one of the lines is suddenly taken out of service under the following conditions:

a. The load has unity power factor and there is no reactive power injection at the receiving end.

b. The load has unity power factor and a synchronous condenser injects 25 MVAr at the receiving end.

c. The load has a 0.9 power factor lagging and there is no reactive power injection at the receiving end.

5.7 Consider the small power system shown in Figure 5.13. Each line of this system is modeled by a π equivalent circuit. The parameters of the lines are given in the table below.

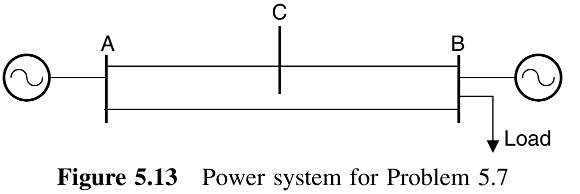


Figure 5.13 Power system for Problem 5.7

Line	R (p.u.)	X (p.u.)	B (p.u.)
A-B	0.08	0.8	0.3
A-C	0.04	0.4	0.15
C-B	0.04	0.4	0.15

Using a power flow program, study the reactive support requirements as a function of the amount of power transferred from bus A to bus B for both normal conditions and abnormal conditions (i.e. avoiding a voltage collapse following the sudden outage of a line). Consider both a unity power factor and a 0.9 power factor–lagging load at bus B. Analyze and discuss the usefulness of a source of reactive power at bus C.

5.8 Identify the documents governing the provision of ancillary services in the region where you live or in another region of your choice. Determine the mechanism used to obtain each service. When services are compulsory, determine their parameters (e.g. minimum lead and lag power factor for generators). When services are procured on a competitive basis, describe the structure of the markets for ancillary services (duration of contracts, bid parameters). Pay particular attention to the definition of reserve services. Identify the mechanism used to pass on the cost of ancillary services to consumers.

5.9 Analyze prices and volumes in the markets for ancillary services in the region where you live or in another region for which you have access to the necessary data.

5.10 The owner of a generating unit would like to maximize its profit by selling both energy and balancing services. Write the objective function and the constraints for this optimization problem. Discuss the various cases that might arise depending on the price paid for energy and balancing. Neglect the constraint introduced by the ramp rate of the generating unit. (Hint: Equation (5.14) of Example 5.6 must be modified because providing balancing services might involve a reduction in the output of the generating unit. On the other hand, you may need to consider explicitly that the amount of balancing service sold must be positive, that is $x_2 \geq 0$.)

6 Transmission Networks and Electricity Markets

6.1 Introduction

In most, if not all, regions of the world, the assumption that electrical energy can be traded as if all generators and loads were connected to the same busbar is not tenable. Transmission constraints and losses in the network connecting these generators and loads can introduce gross distortions in the market for electrical energy. In this chapter, we study the effects that a transmission network has on trading of electrical energy and the special techniques that can be used to hedge against the limitations and price fluctuations that are caused by this network. We consider first and briefly bilateral or decentralized trading. We then turn our attention to centralized or pool-based trading.

6.2 Decentralized Trading Over a Transmission Network

In a decentralized or bilateral trading system, all transactions for electrical energy involve only two parties: a buyer and a seller. These two parties agree on a quantity, a price and any other condition that they may want to attach to the trade. The system operator does not get involved in these transactions and does not set the prices at which transactions take place. Its role is limited to maintaining the balance and the security of the system. This involves the following:

- Buying or selling energy to balance the load and the generation. Under normal circumstances, the amounts involved in these balancing transactions should be small.
- Limiting the amount of power that generators can inject at some nodes of the system if security cannot be maintained through other means.

Let us consider the two-bus power system shown in Figure 6.1 in which trading in electrical energy operates on a bilateral basis. Let us suppose that Generator G_1 has

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

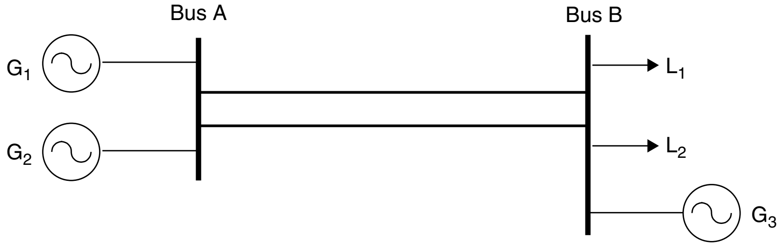


Figure 6.1 Bilateral trading in a two-bus power system

signed a contract for the delivery of 300 MW to load L_1 and that Generator G_2 has agreed to deliver 200 MW to load L_2. Since these transactions are bilateral, the agreed prices are a private matter between the buyer and the seller. On the other hand, the amount of power to be transmitted must be reported to the system operator because this power flows on the transmission system that is open to all parties. The system operator must check that the system will remain secure when all these transactions are implemented. In this case, security is not a problem as long as the capacity of the transmission lines connecting buses A and B is at least 500 MW even under contingency conditions. If the amount of power that can securely be transmitted between buses A and B is less than 500 MW, the system operator has to intervene. Some of the bilateral transactions that were concluded between generators at bus A and loads at bus B must be curtailed.

6.2.1 Physical transmission rights

With modern power system analysis software, determining that a set of transactions would make the operation of the system insecure can be computationally demanding, but is conceptually simple. Deciding which transactions should be curtailed to maintain the required level of security is a much more complex question. Administrative procedures can be established to determine the order in which transactions should be cut back. Such transmission load relief procedures take into account the nature of the transactions (firm or nonfirm), the order in which they were registered with the system operator and possibly some historical factors. They do not, however, factor in the relative economic benefits of the various transactions because a decentralized trading environment does not provide a framework for evaluating these benefits. Administrative curtailments are therefore economically inefficient and should be avoided.

Advocates of decentralized electricity trading believe that the parties considering transactions for electrical energy are best placed to decide whether they wish to use the transmission network. When they sign a contract, producers at bus A and consumers at bus B who do not wish to see their transaction interrupted by congestion should therefore also purchase the right to use the transmission system for this transaction. Since these transmission rights are purchased at a public auction, the parties have the opportunity to decide whether this additional cost is justifiable.

For example, let us suppose that Generator G_1 and load L_1 of Figure 6.1 have agreed on a price of 30.00 $/MWh, while Generator G_2 and load L_2 agreed on 32.00 $/MWh. At the same time, Generator G_3 offers energy at 35.00 $/MWh. Load L_2 should therefore not agree to pay more than 3.00 \$/MWh for transmission rights because this would make the energy it purchases from G_1 more expensive than the energy it could purchase from G_3. The price of transmission rights would have to rise above 5.00 \$/MWh before L_1 reaches the same conclusion. The cost of transmission rights is also an argument that the consumers can use in their negotiations with the generators at bus B to convince them to lower their prices.

Transmission rights of this type are called *physical transmission rights* because they are intended to support the actual transmission of a certain amount of power over a given transmission link.

6.2.2 Problems with physical transmission rights

Our simple example makes physical transmission rights appear simpler than they turn out to be. The first difficulty is practical and arises because the path that power takes through a network is determined by physical laws and not by the wishes of market participants. The second problem is that physical transmission rights have the potential to exacerbate the exercise of market power by some participants. Let us consider these two issues in turn.

6.2.2.1 Parallel paths

Two fundamental laws govern current and power flows in electrical networks: Kirchoff's Current Law (KCL) and Kirchoff's Voltage Law (KVL). KCL specifies that the sum of all the currents entering a node must be equal to the sum of all the currents exiting this node. KCL implies that the active and reactive powers must both be in balance at each node. KVL specifies that the sum of the voltage drops across all the branches of any loop must be equal to zero or, equivalently, that the voltage drops along parallel paths must be equal. Since these voltage drops are proportional to the current flowing through the branch, KVL determines how the currents (and hence the active and reactive power flows) distribute themselves through the network. In the simple example shown in Figure 6.2, a current $\overline{I}$ can flow from node 1 to node 2 along two parallel paths of impedances z_A and z_B. The voltage difference between the two nodes is thus

$$\overline{V_{12}} = z_A \overline{I_A} = z_B \overline{I_B}$$

Since $\overline{I} = \overline{I_A} + \overline{I_B}$, we have

$$\overline{I_A} = \frac{z_B}{z_A + z_B} \overline{I} \tag{6.1}$$

$$\overline{I_B} = \frac{z_A}{z_A + z_B} \overline{I} \tag{6.2}$$

Currents in parallel paths therefore divide themselves in inverse proportion of the impedance of each path. To simplify the following discussion, we will assume that the resistance of any branch is much smaller than its reactance:

$$Z = R + jX \approx jX \tag{6.3}$$

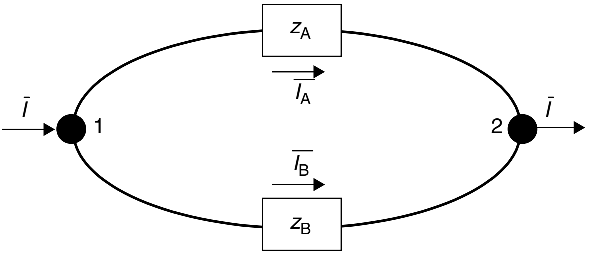


Figure 6.2 Illustration of Kirchoff's Voltage Law

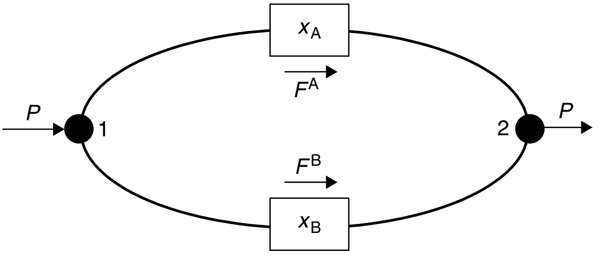


Figure 6.3 Active power flows on parallel paths

We also neglect the flow of reactive power through the network and the losses. Under these assumptions, the system of Figure 6.2 can be depicted in terms of active power flows as shown in Figure 6.3. The active power flows in the parallel paths are related by the following expressions:

$$F^{\mathrm{A}} = \frac{x_{\mathrm{B}}}{x_{\mathrm{A}} + x_{\mathrm{B}}} P \tag{6.4}$$

$$F^{\mathrm{B}} = \frac{x_{\mathrm{A}}}{x_{\mathrm{A}} + x_{\mathrm{B}}} P \tag{6.5}$$

The factors relating the active power injections and the branch flows are called power transfer distribution factors (PTDF).

6.2.2.2 Example

A two-bus system does not illustrate the effect of KVL because there is only one path that the power can follow[1]. We must therefore consider a network with three buses and one loop. Figure 6.4 illustrates such a system and Table 6.1 gives its parameters. To keep matters simple, we assume that network limitations take the form of constant capacity limits on the active power flowing in each line and that the resistance of the lines is negligible.

Let us suppose that Generator B and load Y want to sign a contract for the delivery of 400 MW. If Generator B injects these 400 MW at bus 1 and load Y extracts them

[1]For the sake of simplicity, we treat a line with two identical circuits as a single branch.

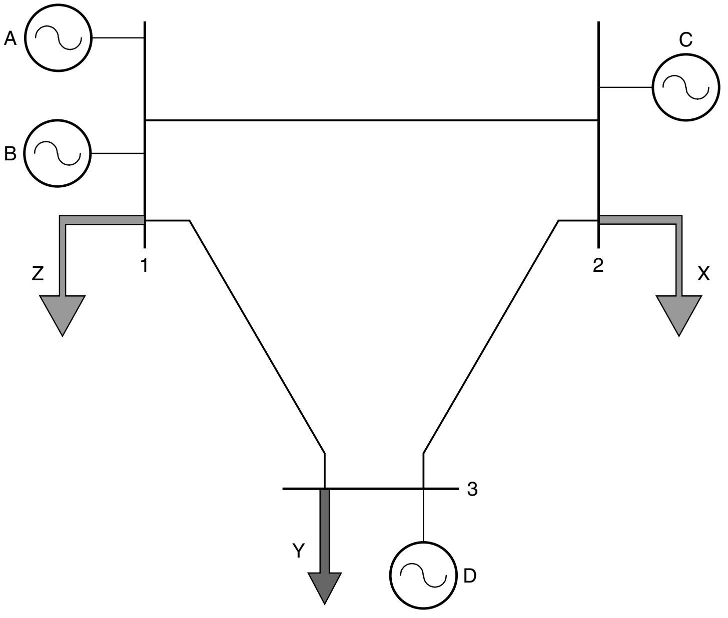


Figure 6.4 A simple three-bus system

Table 6.1 Branch data for the three-bus system of Figure 6.4

Branch	Reactance (p.u.)	Capacity (MW)
1-2	0.2	126
1-3	0.2	250
2-3	0.1	130

at bus 3, this power flows along the two paths shown in Figure 6.5. The amounts of power flowing along paths I and II are given by

$$F^{\mathrm{I}} = \frac{0.2}{0.2 + 0.3} \times 400 = 160\,\mathrm{MW}$$

$$F^{\mathrm{II}} = \frac{0.3}{0.2 + 0.3} \times 400 = 240\,\mathrm{MW}$$

To guarantee that this transaction can actually take place, the parties therefore need to secure 240 MW of transmission rights on Line 1-3 as well as 160 MW of transmission rights on Lines 1-2 and 2-3. This is clearly not possible if this transaction is the only one taking place in this network because the maximum capacity of Lines 1-2 and 2-3 are 126 MW and 130 MW respectively. In the absence of any other transaction, since the most constraining limitation is the capacity of Line 1-2, the maximum amount that

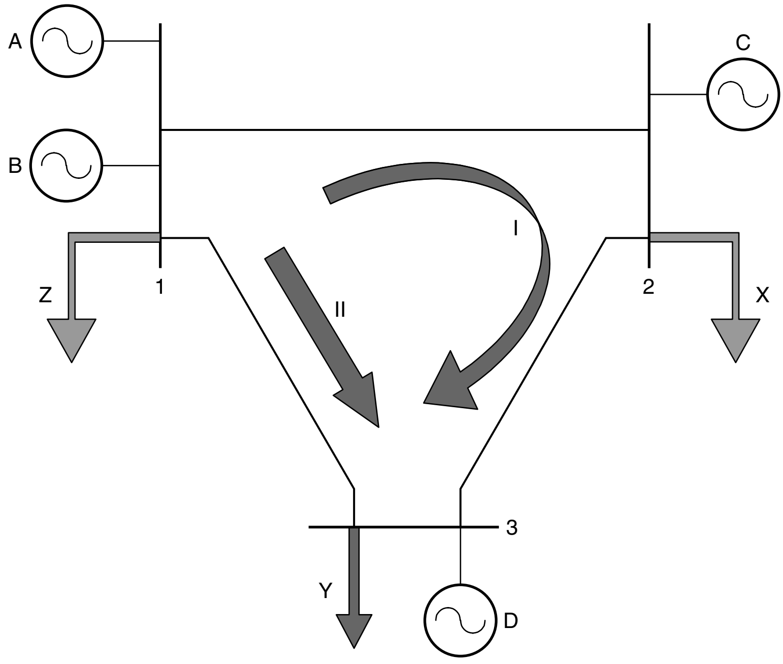


Figure 6.5 Paths for a transaction between Generator B and load Y

A and Y can trade is thus

$$P^{\text{MAX}} = \frac{0.5}{0.2} \times 126 = 315\,\text{MW}$$

However, suppose that load Z would like to purchase 200 MW from Generator D. This power would flow in the following proportions along the paths shown in Figure 6.6:

$$F^{\text{III}} = \frac{0.2}{0.2 + 0.3} \times 200 = 80\,\text{MW}$$

$$F^{\text{IV}} = \frac{0.3}{0.2 + 0.3} \times 200 = 120\,\text{MW}$$

Let us calculate what the flows in this network would be if both of these transactions were to take place at the same time. For this calculation, we can make use of the superposition theorem because our simplifying assumptions have linearized the relations between flows and injections. The flows in the various lines are thus given by

$$F_{12} = F_{23} = F^{\text{I}} - F^{\text{III}} = 160 - 80 = 80\,\text{MW}$$

$$F_{13} = F^{\text{II}} - F^{\text{IV}} = 240 - 120 = 120\,\text{MW}$$

The transaction between Generator D and load Z thus creates a counterflow that increases the power that Generator D and load Y can trade.

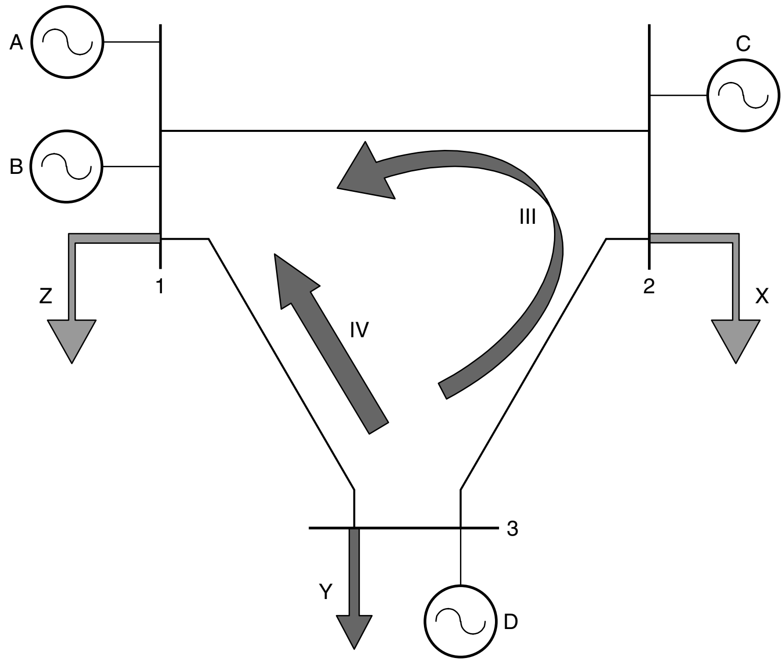


Figure 6.6 Paths for a transaction between Generator D and load Z

If we do not want the transmission network to limit trading opportunities unnecessarily, the amount of physical transmission rights that is made available must take into account possible counterflows. In keeping with the bilateral or decentralized trading philosophy, the system operator should only check that the system would be secure if all the proposed transactions were implemented. If this is not the case, the market participants have to adjust their position through further bilateral contracts until a secure system operating state is achieved. Bilateral energy trading is therefore closely coupled with bilateral trading in physical transmission rights.

In theory, if the market is sufficiently competitive, participants should be able to discover through iterative interactions a combination of bilateral trades in energy and transmission rights that achieves the economic optimum. In practice, in a power system with more than a few capacity constraints, the amount of information that needs to be exchanged is so large that it is unlikely that this optimum could be reached quickly enough through bilateral interactions.

6.2.2.3 Physical transmission rights and market power

We have defined physical transmission rights as giving their owner the right to transmit a certain amount of power for a certain time through a given branch of the transmission network. If physical transmission rights are treated like other types of property rights, their owners can use them or sell them. They can also decide to keep them but not use them. In a perfectly competitive market, buying physical transmission rights but not using them would be an irrational decision. On the other hand, in a less than perfectly competitive market, physical transmission rights can enhance the ability of some participants to exert market power. Consider, for example, the two-bus power system of Figure 6.1. If Generator G_3 is the only generator connected to bus B, it might want to purchase physical transmission rights for power flowing from bus A to bus B. If G_3 does not use or resell these rights, it effectively decreases the amount of power that can be sold at bus B by the other generators. This artificial reduction in transmission capacity enhances the market power that G_3 exerts at bus B and allows it to increase the profit margin on its production. It also has a detrimental effect on the economic efficiency of the overall system. See Joskow and Tirole (2000) for a comprehensive discussion of this issue.

To avoid this problem, it has been suggested that a "use them or lose them" provision be attached to physical transmission rights. Under this provision, transmission capacity that a participant has reserved but does not use is released to others who wish to use it. In theory, this approach should prevent market participants from hoarding transmission capacity for the purpose of enhancing market power. In practice, enforcing this condition is difficult because the unused transmission capacity may be released so late that other market participants are unable to readjust their trading positions.

6.3 Centralized Trading Over a Transmission Network

In a centralized or pool-based trading system, producers and consumers submit their bids and offers to the system operator, who also acts as market operator. The system operator, which must be independent from all the other parties, selects the bids and offers that optimally clear the market while respecting the security constraints imposed by the transmission network. As part of this process, the system operator also determines the market clearing prices. We shall show that, when losses or congestion in the transmission network is taken into account, the price of electrical energy depends on the bus in which the power is injected or extracted. The price that consumers and producers pay or are paid is the same for all participants connected to the same bus. This was not necessarily the case in a decentralized trading system in which prices are determined by bilateral contracts. In a centralized trading system, the system operator thus has a much more active role than it does in the bilateral model. Economic efficiency is indeed achieved only if it optimizes the use of the transmission network.

6.3.1 Centralized trading in a two-bus system

We will begin our analysis of the effects of a transmission network on centralized trading of electrical energy using a simple example involving the two fictitious countries of Borduria and Syldavia. After many years of hostility, these two countries have decided that the path to progress lies in economic cooperation. One of the projects that are under consideration is the reenergization of an existing electrical interconnection between the two national power systems. Before committing themselves to this project, the two governments have asked Bill, a highly regarded independent economist, to

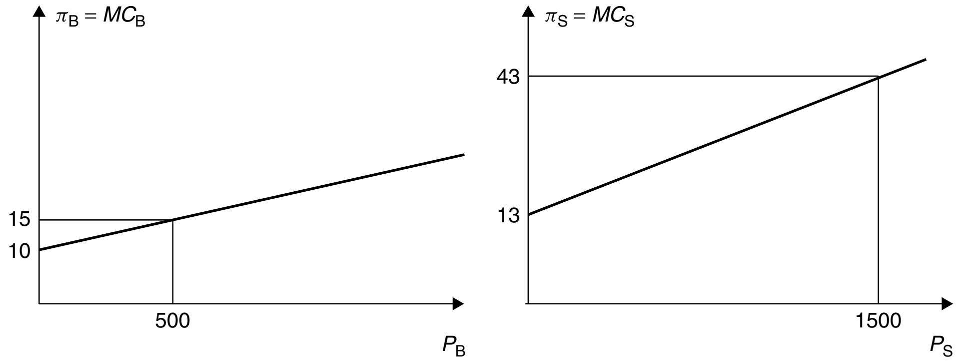


Figure 6.7 Supply functions for the electrical energy markets of Borduria and Syldavia

study the effect that this interconnection would have on their electricity markets and to evaluate the benefit that this interconnection would bring to both countries.

Bill begins his study by analyzing the two national power systems. He observes that both countries have developed centralized electricity markets that are quite competitive. The price of electrical energy in each market thus closely reflects its marginal cost of production. In both countries, the installed generation capacity exceeds the demand by a significant margin. Using regression analysis, Bill estimates the supply function for the electricity market in each country. In Borduria, this function is given by

$$\pi_B = MC_B = 10 + 0.01 P_B \text{ [\$/MWh]} \tag{6.6}$$

While in Syldavia, it is given by

$$\pi_S = MC_S = 13 + 0.02 P_S \text{ [\$/MWh]} \tag{6.7}$$

Like all supply curves, these functions increase monotonically with the demand for electrical energy. Figure 6.7 illustrates these supply functions. For the sake of simplicity, Bill assumes that the demands in Borduria and Syldavia are constant and equal to 500 MW and 1500 MW respectively. He also assumes that these demands have a price elasticity of zero. When the two national electricity markets operate independently, the prices are thus respectively

$$\pi_B = MC_B = 10 + 0.01 \times 500 = 15 \text{ \$/MWh} \tag{6.8}$$

$$\pi_S = MC_S = 13 + 0.02 \times 1500 = 43 \text{ \$/MWh} \tag{6.9}$$

Neither country is interconnected with other countries. Since the transmission infrastructure within each country is quite strong and very rarely affects the operation of the market for electrical energy, Bill decides that the simple model shown in Figure 6.8 is adequate for the study he needs to perform.

6.3.1.1 Unconstrained transmission

Under normal operating conditions, the interconnection can carry 1600 MW. If all the generators in Syldavia were to be shut down, the entire load of that country could still

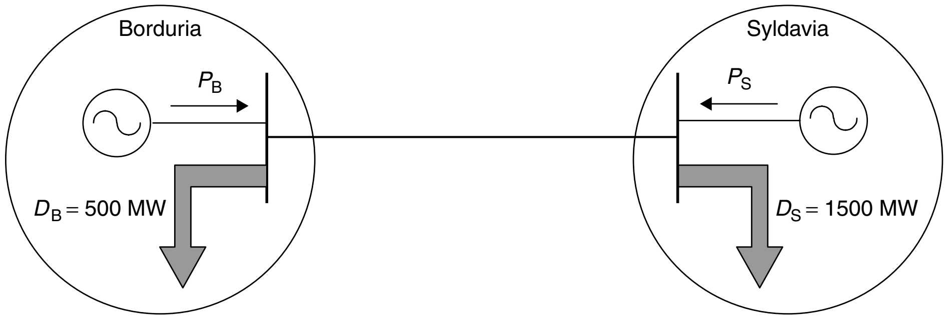


Figure 6.8 Model of the Borduria/Syldavia interconnection

be supplied from Borduria through the interconnection. The capacity of this link is thus larger than the power that could possibly need to be transmitted.

Equations (6.8) and (6.9) show that electricity prices in Borduria are significantly lower than in Syldavia. One might therefore envision that Bordurian generators might supply not only their domestic demand but also the entire demand of Syldavia. We would then have

$$P_B = 2000\,\text{MW} \tag{6.10}$$

$$P_S = 0\,\text{MW} \tag{6.11}$$

Replacing these values in Equations (6.6) and (6.7), we find that the marginal cost of producing electrical energy in the two systems would be

$$MC_B = 30\,\$/\text{MWh} \tag{6.12}$$

$$MC_S = 13\,\$/\text{MWh} \tag{6.13}$$

This situation is clearly not tenable because Bordurian generators would demand 30 \$/MWh while Syldavian generators would be willing to sell energy at 13 \$/MWh. The Bordurian generators are thus not able to capture the entire market because a process of price equalization would take place. In other words, the interconnection forces the markets for electrical energy in both countries to operate as a single market. A single market clearing price then applies to all the energy consumed in both countries:

$$\pi = \pi_B = \pi_S \tag{6.14}$$

Generators from both countries compete to supply the total demand, which is equal to the sum of the two national demands:

$$P_B + P_S = D_B + D_S = 500 + 1500 = 2000\,\text{MW} \tag{6.15}$$

Since the generators in both countries are willing to produce up to the point at which their marginal cost of production is equal to the market clearing price, Equations

(6.6) and (6.7) are still applicable. To determine the market equilibrium, Bill solves the system of Equations (6.6), (6.7), (6.14) and (6.15). He gets the following solution:

$$\pi = \pi_B = \pi_S = 24.30\,\$/\text{MWh} \tag{6.16}$$

$$P_B = 1433\,\text{MW} \tag{6.17}$$

$$P_S = 567\,\text{MW} \tag{6.18}$$

The flow of power in the interconnection is equal to the excess of generation over load in the Bordurian system and the deficit in the Syldavian system:

$$F_{BS} = P_B - D_B = D_S - P_S = 933\,\text{MW} \tag{6.19}$$

A flow of power from Borduria to Syldavia makes sense because the price of electricity in Borduria is lower than in Syldavia when the interconnection is not in service.

Figure 6.9 offers a graphical representation of the operation of this single market. The productions of the Bordurian and Syldavian generators are plotted respectively from left to right and right to left. Since the two vertical axes are separated by the total load in the system, any point on the horizontal axis represents a feasible dispatch of this load between generators in the two countries. This diagram also shows the supply curves of the two national markets. The prices in Borduria and Syldavia are measured along the left and right axes respectively.

When the two systems operate as a single market, the prices in both systems must be identical. Given the way this diagram has been constructed, the intersection of the two supply curves gives this operating point. The diagram then shows the production in each country and the flow on the interconnection.

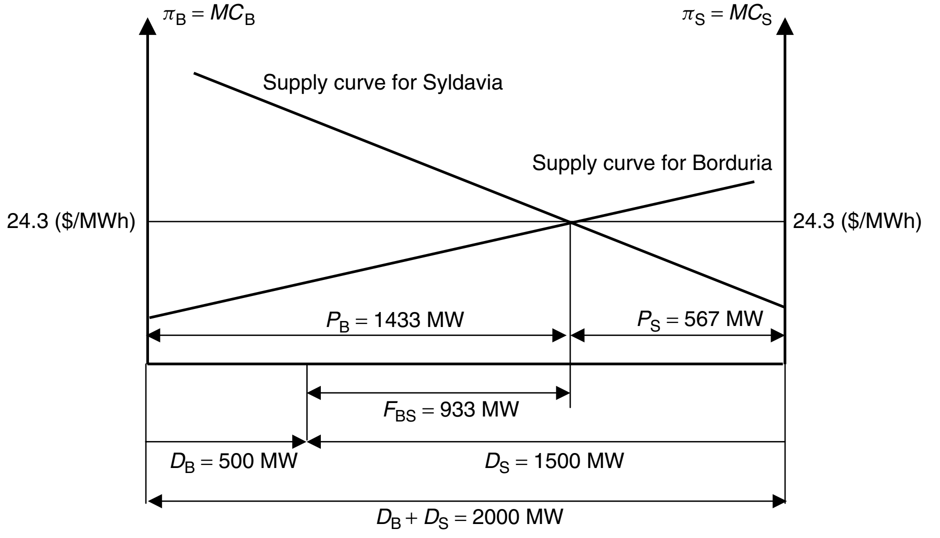


Figure 6.9 Graphical representation of the combination of the Syldavian and Bordurian electricity markets into a single market

6.3.1.2 Constrained transmission

Over the course of a year, various components of the transmission system must be taken out of service for maintenance. These components include not only lines and transformers but also some generating plants that provide essential reactive support services. The Borduria–Syldavia interconnection is therefore not always able to carry its nominal 1600 MW capacity. After consulting transmission engineers, Bill estimates that, during a significant part of each year, the interconnection is able to carry only a maximum of 400 MW. He therefore needs to study how the system behaves under these conditions.

When the capacity of the interconnection is limited to 400 MW, the production in Borduria must be reduced to 900 MW (500 MW of local load and 400 MW sold to consumers in Syldavia). The production in Syldavia is then 1100 MW. Using Equations (6.6) and (6.7), we find that

$$\pi_B = MC_B = 10 + 0.01 \times 900 = 19\,\$/\text{MWh} \tag{6.20}$$

$$\pi_S = MC_S = 13 + 0.02 \times 1100 = 35\,\$/\text{MWh} \tag{6.21}$$

Figure 6.10 illustrates this situation. The constraint on the capacity of the transmission corridor creates a difference of 16 \$/MWh between the prices of electrical energy in Borduria and Syldavia. If electricity were a normal commodity, traders would spot a business opportunity in this price disparity. If they could find a way of shipping more power from Borduria to Syldavia, they could make money by buying energy in one market and selling it in the other. However, this opportunity for *arbitrage* cannot be realized because the interconnection is the only way to transmit power between the two countries and it is already fully loaded. The price difference can thus persist

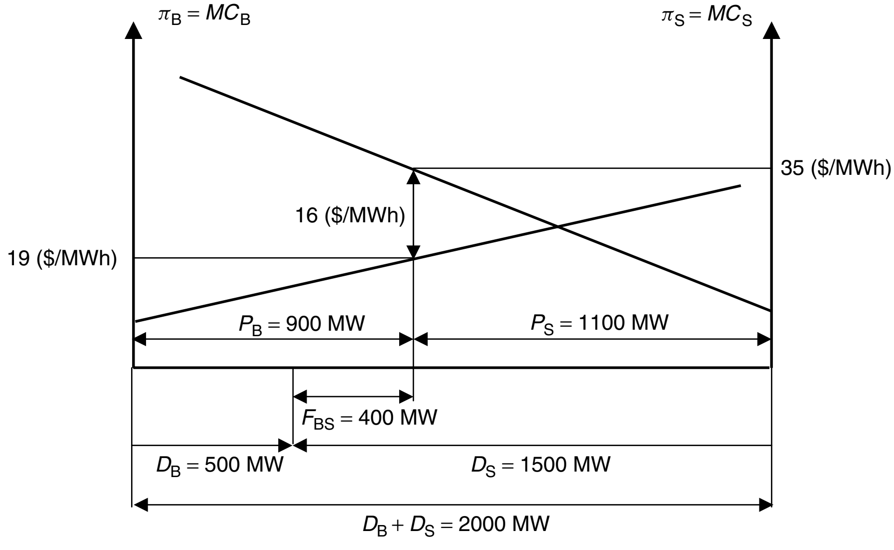


Figure 6.10 Graphical representation of the effect of congestion on the Syldavian and Bordurian electricity markets

as long as the capacity of the interconnection remains below the capacity needed to ensure free interchanges. Constraints imposed by the need to maintain the security of the system can thus create congestion in the transmission network. This congestion divides what should be a single market into separate markets. Because of the congestion, an additional megawatt of load in each country would have to be provided solely by the local generators. The marginal cost of producing electrical energy is therefore different in each country. If these separate markets are still sufficiently competitive, the prices are still equal to the marginal costs. We thus have what is called *locational marginal pricing* because the marginal cost depends on the location where the energy is produced or consumed. If a different price is defined at each bus or node in the system, locational marginal pricing is called *nodal pricing*. Our example shows that locational marginal prices are higher in areas that normally import power and lower in areas that export power.

Bill summarizes his findings so far in Table 6.2. He uses the following notations in this table: R represents the revenue accruing to a group of generators from the sale of electrical energy; E represents the payment made by a group of consumers for the purchase of electrical energy; the subscripts B and S denote respectively Borduria and Syldavia. F_{BS} represents the power flowing on the interconnection. This quantity is positive if power flows from Borduria to Syldavia.

Table 6.2 shows that the biggest beneficiaries of the reenergization of the interconnection are likely to be the Bordurian generators and the Syldavian consumers. Bordurian consumers would see an increase in the price of electrical energy. Syldavian generators would loose a substantial share of their market. Overall, the interconnection would have a positive effect because it would reduce the total amount of money spent by consumers on electrical energy. This saving arises because the energy produced by less efficient generators is replaced by energy produced by more efficient ones. Congestion on the interconnection reduces its overall benefit. Note that this congestion partially shields Syldavian generators from the competition of their Bordurian counterparts.

We have assumed so far that the markets are perfectly competitive. If competition were less than perfect, congestion in the interconnection would allow Syldavian

Table 6.2 Operation of the Borduria/Syldavia interconnection as separate markets, as a single market and as a single market with congestion

	Separate markets	Single market	Single market with congestion
P_B (MW)	500	1433	900
π_B (\$/MWh)	15	24.33	19
R_B (\$/h)	7500	34 865	17 100
E_B (\$/h)	7500	12 165	9500
P_S (MW)	1500	567	1100
π_S (\$/MWh)	43	24.33	35
R_S (\$/h)	64 500	13 795	38 500
E_S (\$/h)	64 500	36 495	52 500
F_{BS} (MW)	0	933	400
$R_{TOTAL} = R_B + R_S$	72 000	48 660	55 600
$E_{TOTAL} = E_B + E_S$	72 000	48 660	62 000

generators to raise their prices above their marginal cost of production. On the other hand, this congestion would intensify competition in the Bordurian market.

6.3.1.3 Congestion surplus

Bill decides that it would be interesting to quantify the effect that congestion on the interconnection would have on producers and consumers in both countries. He calculates the prices in Borduria and Syldavia as a function of the amount of power flowing on the interconnection:

$$\pi_B = MC_B = 10 + 0.01(D_B + F_{BS}) \tag{6.22}$$

$$\pi_S = MC_S = 13 + 0.02(D_S - F_{BS}) \tag{6.23}$$

Bill assumes that consumers pay the going price in their local market independently of where the energy they consume is produced. The total payment made by consumers is thus given by

$$E_{TOTAL} = \pi_B \cdot D_B + \pi_S \cdot D_S \tag{6.24}$$

Combining Equations (6.22), (6.23) and (6.24), Figure 6.11 shows how this payment varies as a function of F_{BS}. As Bill expected, this payment decreases monotonically as the flow between the two countries increases. The curve does not extend beyond $F_{BS} = 933$ MW because we saw earlier that a greater interchange does not make economic sense.

Similarly, Bill assumes that generators are paid the going price in their local market for the electrical energy they produce, independently of where this energy is consumed.

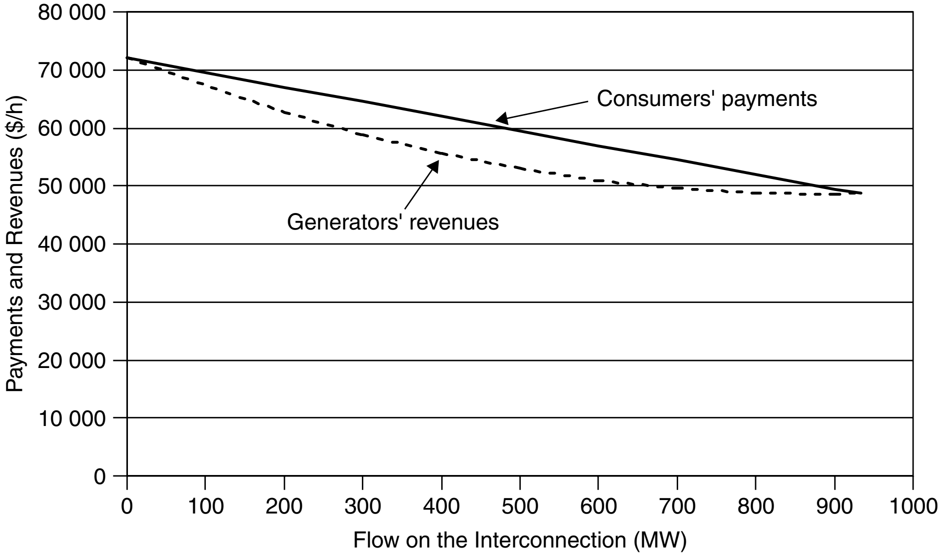


Figure 6.11 Consumers' payments (solid line) and generators' revenue (dashed line) as a function of the flow on the interconnection between Borduria and Syldavia

The total revenue collected by the generators from the sale of electrical energy in both markets is thus given by

$$R_{\text{TOTAL}} = \pi_B \cdot P_B + \pi_S \cdot P_S = \pi_B \cdot (D_B + F_{BS}) + \pi_S \cdot (D_S - F_{BS}) \tag{6.25}$$

This quantity has also been plotted on Figure 6.11 as a function of the power flowing on the interconnection. We observe that this revenue is less than the payment made by the consumers except when the interconnection is not congested ($F_{BS} = 933\,\text{MW}$) or when it is not in service ($F_{BS} = 0\,\text{MW}$). Combining Equations (6.24) and (6.25) while recalling that the flow on the interconnection is equal to the surplus of production over consumption in each country, we can write

$$\begin{aligned} E_{\text{TOTAL}} - R_{\text{TOTAL}} &= \pi_S \cdot D_S + \pi_B \cdot D_B - \pi_S \cdot P_S - \pi_B \cdot P_B \\ &= \pi_S \cdot (D_S - P_S) + \pi_B \cdot (D_B - P_B) \\ &= \pi_S \cdot F_{BS} + \pi_B \cdot (-F_{BS}) \\ &= (\pi_S - \pi_B) \cdot F_{BS} \end{aligned} \tag{6.26}$$

This difference between payments and revenues is called the *merchandizing surplus*. It is thus equal to the product of the differences between the prices in the two markets and the flow on the interconnection between these two markets. In this case, since this surplus is due to the congestion in the network, it is also called the *congestion surplus*.

In particular, for the case where the flow on the interconnection is limited to 400 MW, we have

$$E_{\text{TOTAL}} - R_{\text{TOTAL}} = (\pi_S - \pi_B) \cdot F_{BS} = (35 - 19) \cdot 400 = \$6400 \tag{6.27}$$

Note that its value is identical to the one we obtain if we take the difference between the total payment and the total revenue given in the last column of Table 6.2.

In a pool system in which all market participants buy or sell at the centrally determined nodal price applicable to their location, this congestion surplus is collected by the market operator. It should not, however, be kept by the market operator because this would give a perverse incentive to increase congestion or at least not work very hard to reduce congestion. On the other hand, simply returning the congestion surplus to the market participants would blunt the effect of nodal marginal pricing, which is designed to encourage efficient economic behavior. We will return to this issue when we discuss the management of congestion risks and financial transmission rights (FTRs) later in this chapter.

6.3.2 Centralized trading in a three-bus system

In our discussion of decentralized or bilateral trading, we already mentioned that Kirchoff's current and voltage laws dictate power flows in a transmission network with more than two buses. Now we need to explore the effect that these physical laws have on centralized trading. We will carry out this investigation using the same three-bus system that we used when we discussed bilateral trading. Figure 6.12 shows the diagram

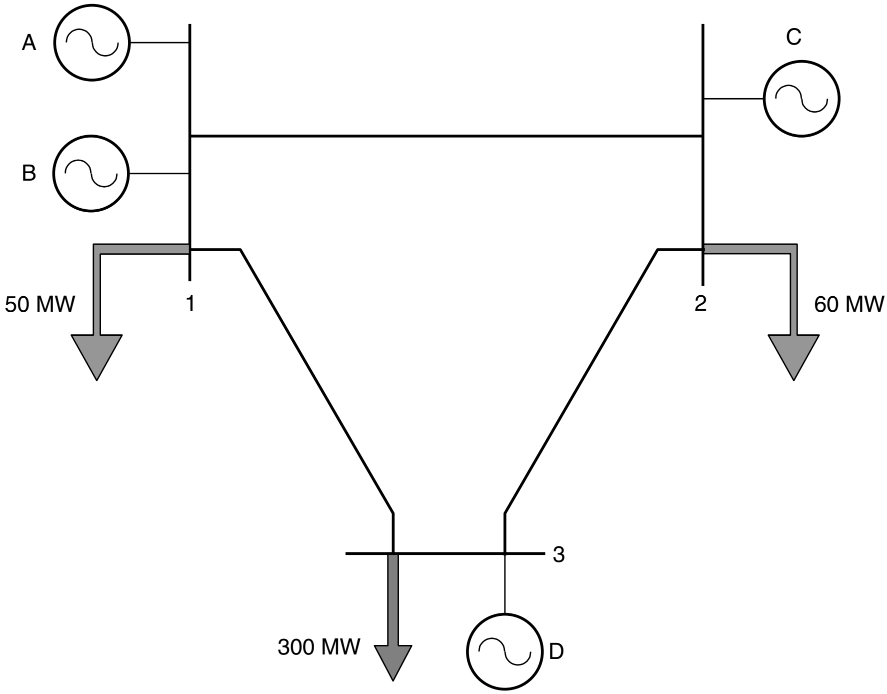


Figure 6.12 Simple three-bus system used to illustrate centralized trading

Table 6.3 Branch data for the three-bus system of Figure 6.12

Branch	Reactance (p.u.)	Capacity (MW)
1-2	0.2	126
1-3	0.2	250
2-3	0.1	130

of the network and Table 6.3 gives its parameters. We will again assume that network limitations take the form of constant capacity limits on the active power flowing in each line and that the resistance of the lines is negligible.

When we analyzed this system in the context of bilateral trading, we did not need to consider price or cost information because this data remains private to the parties involved in each bilateral transaction. On the other hand, in a centralized trading system, producers and consumers submit their bids and offers to the system operator, who uses this information to optimize the operation of the system. Since we are taking the perspective of the system operator, we assume that we have access to the data given in Table 6.4. We also assume that, since the market is perfectly competitive, the generators' bids are equal to their marginal cost. For the sake of simplicity, the marginal cost of each generator is assumed constant and the demand side is represented by the constant loads shown in Figure 6.12.

Table 6.4 Generator data for the three-bus system of Figure 6.12

Generator	Capacity (MW)	Marginal cost ($/MWh)
A	140	7.5
B	285	6
C	90	14
D	85	10

6.3.2.1 Economic dispatch

If we ignore the constraints that the network might impose, the total load of 410 MW should be dispatched solely on the basis of the bids or marginal costs of the generators in a way that minimizes the total cost of supplying the demand. Since we have assumed that these generators have a constant marginal cost over their entire range of operation and that the demand is not price sensitive, this dispatch is easy to compute: the generators are ranked in order of increasing marginal cost and loaded up to their capacity until the demand is satisfied. We get

$$\begin{aligned} P_A &= 125\,\text{MW} \\ P_B &= 285\,\text{MW} \\ P_C &= 0\,\text{MW} \\ P_D &= 0\,\text{MW} \end{aligned} \tag{6.28}$$

The total cost of the economic dispatch is

$$C_{ED} = MC_A \cdot P_A + MC_B \cdot P_B = 2647.50\,\$/\text{h} \tag{6.29}$$

We need to check whether this dispatch would cause one or more flows to exceed the capacity of a line. In a large network, we would calculate the branch flows using a power flow program. For such a simple system, we can do this computation by hand. This exercise will give us a more intuitive understanding of the way power flows through the network. Given the assumed flow directions shown in Figure 6.13, we can write the power balance equation at each bus or node as follows:

$$\text{Node 1: } F_{12} + F_{13} = 360 \tag{6.30}$$

$$\text{Node 2: } F_{12} - F_{23} = 60 \tag{6.31}$$

$$\text{Node 3: } F_{13} + F_{23} = 300 \tag{6.32}$$

In this case, we get three equations in three unknowns. However, these equations are linearly dependent because the power balance also holds for the system as a whole. For example, subtracting Equation (6.31) from Equation (6.30) gives Equation (6.32). Since one of these equations can be eliminated with no loss of information, we are left

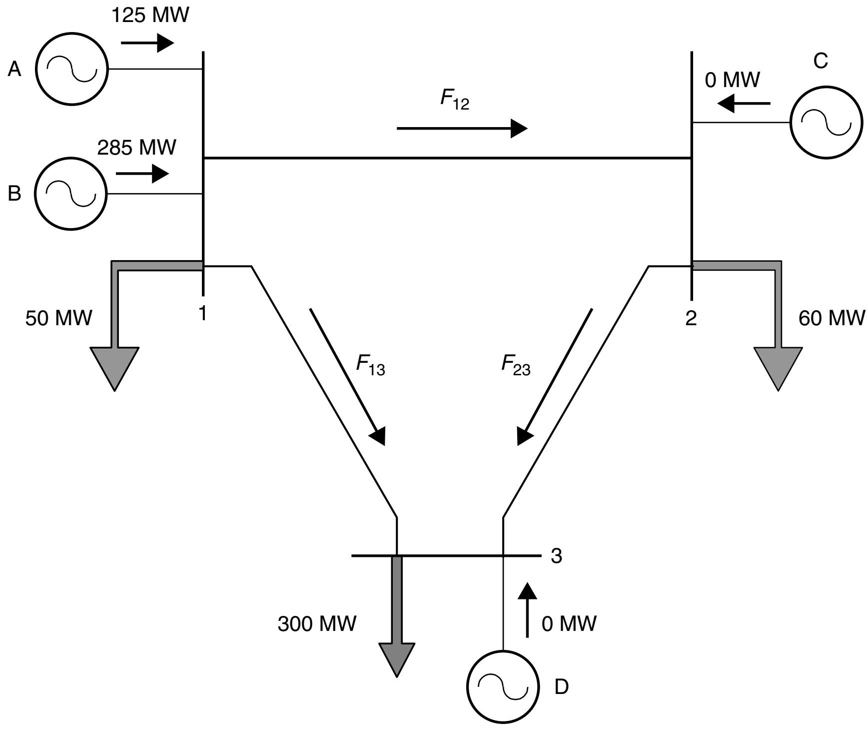


Figure 6.13 Basic dispatch in the three-bus system

with two equations and three unknowns. This is hardly surprising because we have not taken into account the impedances of the branches.

Rather than simply add an equation on the basis of KVL, let us again make use of the superposition theorem. Figure 6.14 shows how our original problem can be decomposed in two simpler problems. If we succeed in determining the flows in these two simpler problems, we can easily find the flows in the original problem because we know from the superposition theorem that

$$F_{12} = F_1^{\mathrm{A}} + F_2^{\mathrm{A}} \tag{6.33}$$

$$F_{13} = F_1^{\mathrm{B}} + F_2^{\mathrm{B}} \tag{6.34}$$

$$F_{23} = F_1^{\mathrm{A}} - F_2^{\mathrm{B}} \tag{6.35}$$

Let us consider the first problem. 300 MW is injected at bus 1 and taken out at bus 3. Since this power can flow along two paths (A and B), we have

$$F_1^{\mathrm{A}} + F_1^{\mathrm{B}} = 300\,\mathrm{MW} \tag{6.36}$$

The reactances of paths A and B are respectively

$$x_1^{\mathrm{A}} = x_{12} + x_{23} = 0.3 \text{ p.u.} \tag{6.37}$$

$$x_1^{\mathrm{B}} = x_{13} = 0.2 \text{ p.u.} \tag{6.38}$$

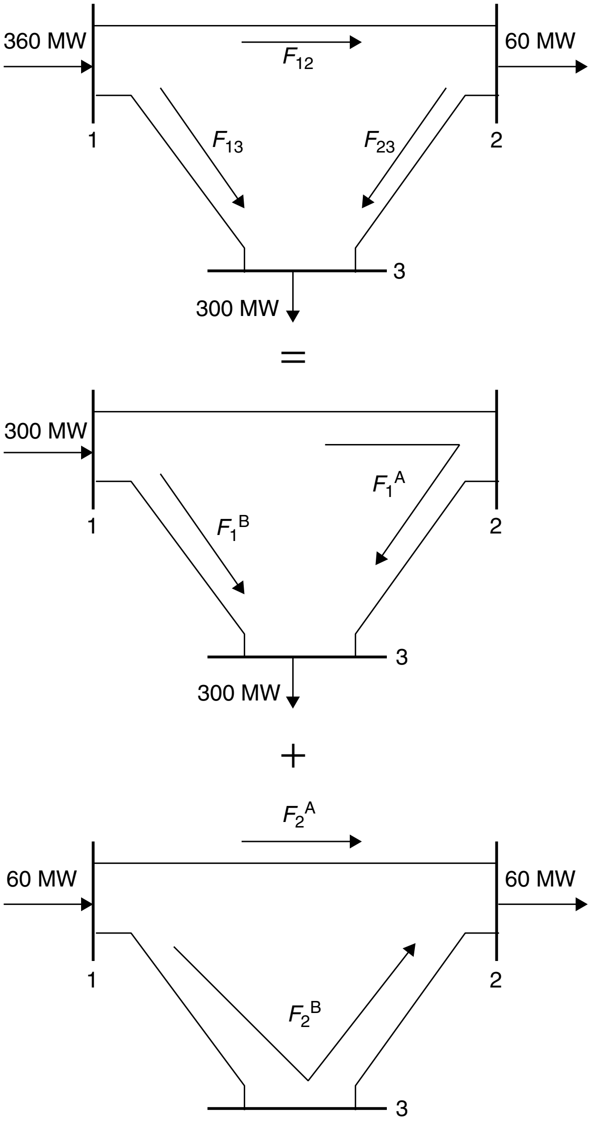


Figure 6.14 Application of the superposition theorem to the calculation of the line flows in the three-bus system

Since these 300 MW divide themselves between the two paths in accordance with Equations (6.4) and (6.5), we have

$$F_1^{\mathrm{A}} = \frac{0.2}{0.3 + 0.2} \cdot 300 = 120\,\mathrm{MW} \tag{6.39}$$

$$F_1^{\mathrm{B}} = \frac{0.3}{0.3 + 0.2} \cdot 300 = 180\,\mathrm{MW} \tag{6.40}$$

Similarly, for the second circuit, 60 MW is injected at bus 1 and taken out at bus 2. In this case, the impedances of the two paths are

$$x_2^A = x_{12} = 0.2\,\text{p.u.} \tag{6.41}$$

$$x_2^B = x_{13} + x_{23} = 0.3\,\text{p.u.} \tag{6.42}$$

Hence,

$$F_2^A = \frac{0.3}{0.3 + 0.2} \cdot 60 = 36\,\text{MW} \tag{6.43}$$

$$F_2^B = \frac{0.2}{0.3 + 0.2} \cdot 60 = 24\,\text{MW} \tag{6.44}$$

Equations (6.33) to (6.35) give the flows in the original system:

$$F_{12} = F_1^A + F_2^A = 120 + 36 = 156\,\text{MW} \tag{6.45}$$

$$F_{13} = F_1^B + F_2^B = 180 + 24 = 204\,\text{MW} \tag{6.46}$$

$$F_{23} = F_1^A - F_2^B = 120 - 24 = 96\,\text{MW} \tag{6.47}$$

Figure 6.15 gives a graphical representation of this solution. From these results, we conclude that the economic dispatch would overload branch 1-2 by 30 MW because it would have to carry 156 MW when its capacity is only 126 MW. This is clearly not acceptable.

6.3.2.2 Correcting the economic dispatch

While the economic dispatch minimizes the total production cost, this solution is not viable because it does not satisfy the security criteria. We must therefore determine the least cost modifications that will remove the line overload. We begin by noting that the economic dispatch concentrates all the generation at bus 1. To reduce the flow on branch 1-2, we can increase the generation either at bus 2 or at bus 3. Let us first

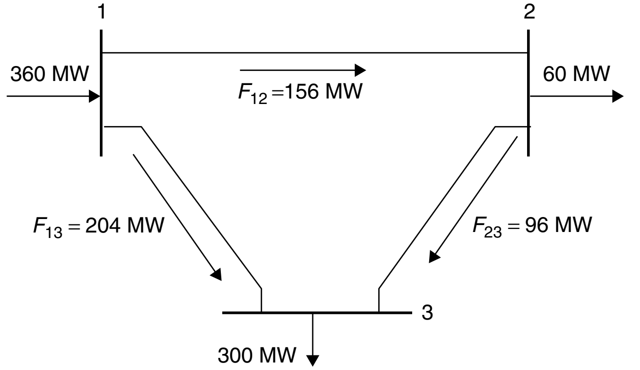


Figure 6.15 Flows for the economic dispatch in the three-bus system

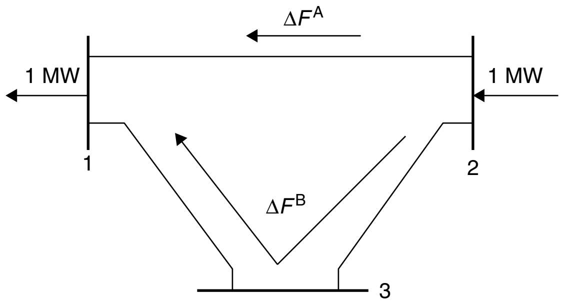


Figure 6.16 Effect of an incremental change in the generation at bus 2

consider what happens when we increase the generation at bus 2 by 1 MW. Since we neglect losses, this implies that we must reduce the generation at bus 1 by 1 MW. Figure 6.16 illustrates this incremental redispatch. Since the incremental flow ΔF^{A} is in the opposite direction as the flow F_{12}, increasing the generation at bus 2 and reducing it at bus 1 will reduce the overload on this branch. To quantify this effect, we can again make use of the superposition theorem. Since the reactances of paths A and B are respectively

$$x^{A} = x_{12} = 0.2\,\text{p.u.} \tag{6.48}$$

$$x^{B} = x_{23} + x_{13} = 0.3\,\text{p.u.} \tag{6.49}$$

and since the sum of the two flows must be equal to 1 MW, we get

$$\Delta F^{A} = 0.6\,\text{MW} \tag{6.50}$$

$$\Delta F^{B} = 0.4\,\text{MW} \tag{6.51}$$

Every megawatt injected at bus 2 and taken out at bus 1 thus reduces the flow on branch 1-2 by 0.6 MW. Given that this line is overloaded by 30 MW, a total of 50 MW of generation must be shifted from bus 1 to bus 2 to satisfy the line capacity constraint. Figure 6.17 illustrates this redispatch and its superposition with the economic dispatch to yield what we will call a constrained dispatch. We observe that the flow on branch 1-3 has also been reduced by this redispatch but that the flow on branch 2-3 has increased. This increase is tolerable, however, because this flow remains smaller than the capacity specified in Table 6.3. To implement this constrained dispatch, the generators connected at bus 1 must produce a total of 360 MW to meet the local load of 50 MW and inject 310 MW net into the network. At the same time, the generator at bus 2 must produce 50 MW. An additional 10 MW is taken from the network to supply the local load of 60 MW. Under these conditions, the least cost generation dispatch is

$$\begin{aligned} P_A &= 75\,\text{MW} \\ P_B &= 285\,\text{MW} \\ P_C &= 50\,\text{MW} \\ P_D &= 0\,\text{MW} \end{aligned} \tag{6.52}$$

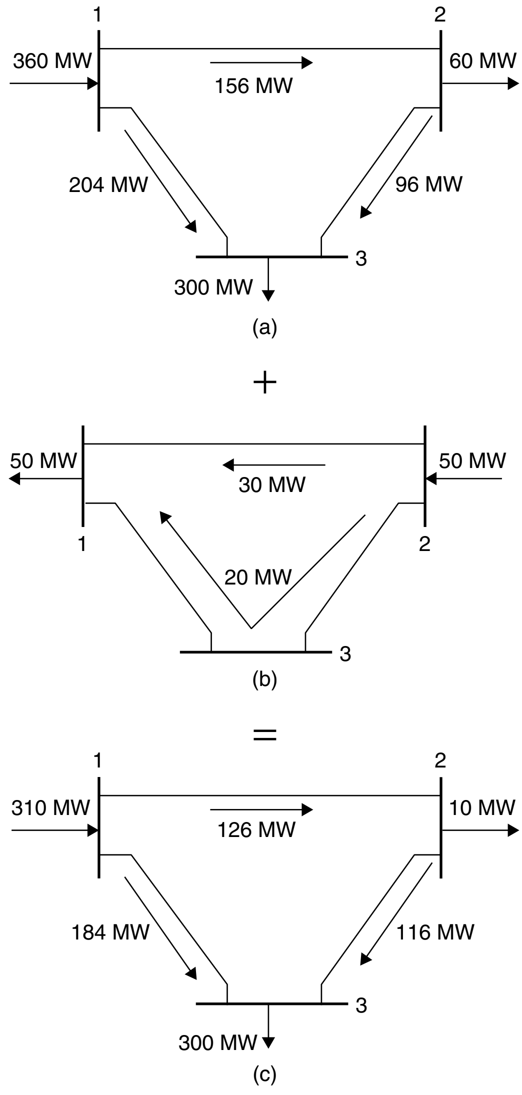


Figure 6.17 Superposition of the redispatch of generation from bus 1 to bus 2 (b) on the economic dispatch (a) to produce a constrained dispatch that meets the constraints on line flows (c)

Comparing with Equation (6.28), we see that the output of Generator A has been reduced rather than the output of Generator B because Generator A has a higher marginal cost. The total cost of this constrained dispatch is

$$C_2 = MC_A \cdot P_A + MC_B \cdot P_B + MC_C \cdot P_C = 2972.50\,\$/\text{h} \qquad (6.53)$$

This cost is necessarily higher than the cost of the economic dispatch that we calculated in Equation (6.29). The difference represents the cost of achieving security using this redispatch.

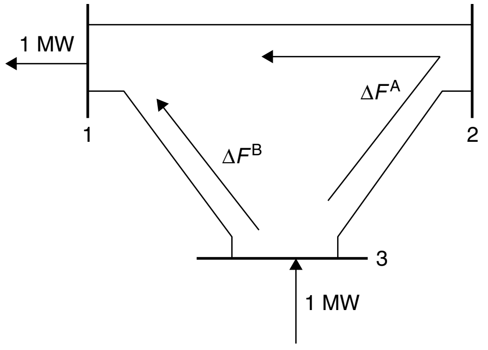


Figure 6.18 Effect of an incremental change in the generation at bus 3

We mentioned above that we could also relieve the overload on branch 1-2 by increasing the output of Generator D connected at bus 3. Let us calculate the extent and the cost of this other redispatch using the same procedure. Figure 6.18 shows the two paths along which an extra MW injected at bus 3 and taken out at bus 1 would divide itself. Given that the reactances of paths A and B are respectively

$$x^{\mathrm{A}} = x_{23} + x_{12} = 0.3\,\text{p.u.} \tag{6.54}$$

$$x^{\mathrm{B}} = x_{13} = 0.2\,\text{p.u.} \tag{6.55}$$

and that the sum of the two flows must be equal to 1 MW, we get

$$\Delta F^{\mathrm{A}} = 0.4\,\text{MW} \tag{6.56}$$

$$\Delta F^{\mathrm{B}} = 0.6\,\text{MW} \tag{6.57}$$

Every MW injected at bus 3 and taken out at bus 1 thus reduces the flow on branch 1-2 by 0.4 MW. This means that we need to shift 75 MW of generation from bus 1 to bus 3 to reduce the flow on branch 1-2 by 30 MW and remove the overload. Figure 6.19 shows how superposing this redispatch on the economic dispatch reduces the flows through all the branches of the network. As expected, the flow on branch 1-2 is equal to the maximum capacity of that branch. Since the total power to be produced at bus 1 is now reduced by 75 MW, the generation dispatch for this case is

$$\begin{aligned} P_{\mathrm{A}} &= 50\,\text{MW} \\ P_{\mathrm{B}} &= 285\,\text{MW} \\ P_{\mathrm{C}} &= 0\,\text{MW} \\ P_{\mathrm{D}} &= 75\,\text{MW} \end{aligned} \tag{6.58}$$

The total cost of this constrained dispatch is

$$C_3 = MC_{\mathrm{A}} \cdot P_{\mathrm{A}} + MC_{\mathrm{B}} \cdot P_{\mathrm{B}} + MC_{\mathrm{D}} \cdot P_{\mathrm{D}} = 2835\,\$/\text{h} \tag{6.59}$$

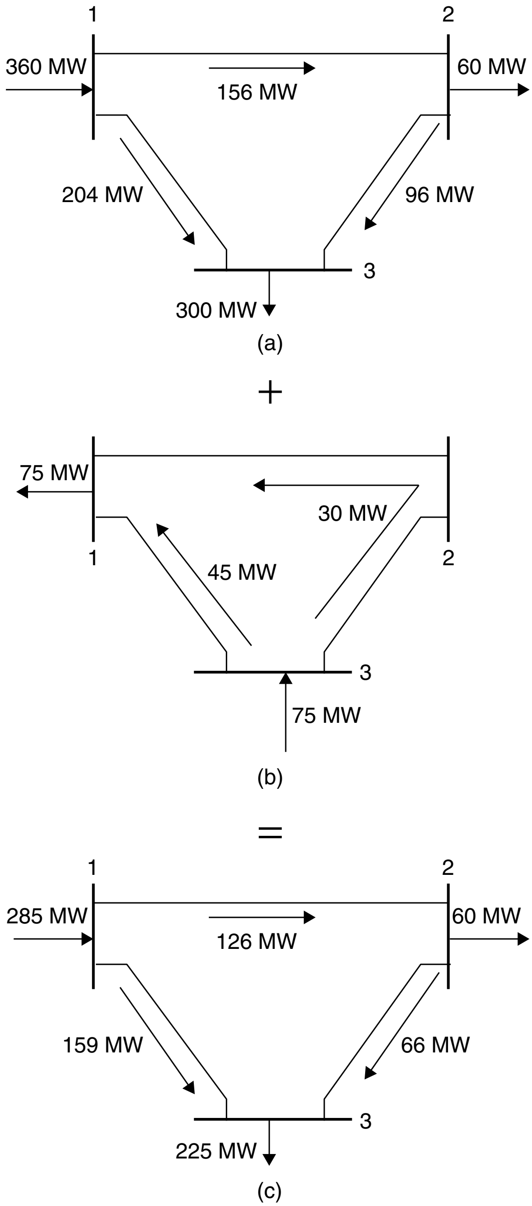


Figure 6.19 Superposition of the redispatch of generation from bus 1 to bus 3 (b) on the economic dispatch (a) to produce a constrained dispatch that meets the constraints on line flows (c)

Let us now compare these two ways of removing the overload on branch 1-2. If we make use of the generation at bus 3, we need to redispatch 75 MW. On the other hand, if we call upon the generation at bus 2, we need to shift only 50 MW. This is because the flow on branch 1-2 is less sensitive to an increase in generation at bus 3 than it is to an increase at bus 2. However, since the marginal cost of Generator D is less than the marginal cost of Generator C, increasing the generation at bus 3 is the preferred solution because it is cheaper. The cost of making the system secure is thus equal to the difference between the cost of this constrained dispatch and the cost of the economic dispatch:

$$C_S = C_3 - C_{ED} = 2835.00 - 2647.50 = 187.50\,\$/\text{h} \tag{6.60}$$

6.3.2.3 Nodal prices

We have already alluded to the concept of nodal marginal price when we discussed the Borduria–Syldavia interconnection. We are now in a position to clarify this concept. The nodal marginal price is equal to the cost of supplying an additional megawatt of load at the node under consideration by the cheapest possible means.

In our three-bus example, this means that we do not start from the economic dispatch but from the constrained dispatch given by Equation (6.58). The output of Generator D has thus been increased to remove the overload on branch 1-2. At node 1 it is clear that an additional megawatt of load should be produced by Generator A. The marginal cost of Generator A is indeed lower than the marginal cost of Generators C and D. While it is higher than the marginal cost of Generator B, this generator is already loaded up to its maximum capacity and is therefore unable to produce an additional megawatt. The network has no impact on the marginal price at this node because the additional megawatt is produced and consumed locally. The nodal marginal price at bus 1 is therefore

$$\pi_1 = MC_A = 7.50\,\$/\text{MWh} \tag{6.61}$$

What is the cheapest way of supplying an additional megawatt at bus 3? Generator A has the lowest marginal cost and is not fully loaded. Unfortunately, increasing the generation at bus 1 would inevitably overload branch 1-2. The next cheapest option is to increase the output of Generator D. Since this generator is located at bus 3, this additional megawatt would not flow through the network. Therefore, we have

$$\pi_3 = MC_D = 10\,\$/\text{MWh} \tag{6.62}$$

Supplying an additional megawatt at bus 2 is a more complex matter. We could obviously generate it locally using Generator C, but this looks rather expensive because at 14 \$/MWh the marginal cost of this generator is much higher than the marginal cost of the other generators. If we choose to adjust the output of the generators at the other buses, we must consider what might happen in the network. Figure 6.20 shows how an additional megawatt of load at bus 2 would flow through the network if it were produced at bus 1 or bus 3. We can see that in both cases we increase the flow on branch 1-2. Since the flow on this branch is already at its maximum value, neither solution is acceptable. Any combination of generation increases at buses 1 and 3 would also be unacceptable. We could, however, increase generation at bus 3 and reduce it at bus 1. For example, as shown in Figure 6.21, we could increase the generation at bus 3 by 2 MW and reduce it at bus 1 by 1 MW. The net increase is then equal to the additional load at bus 2. We can then, once again, use superposition to determine the resulting incremental flows. The first diagram in Figure 6.22 shows that if 1 MW

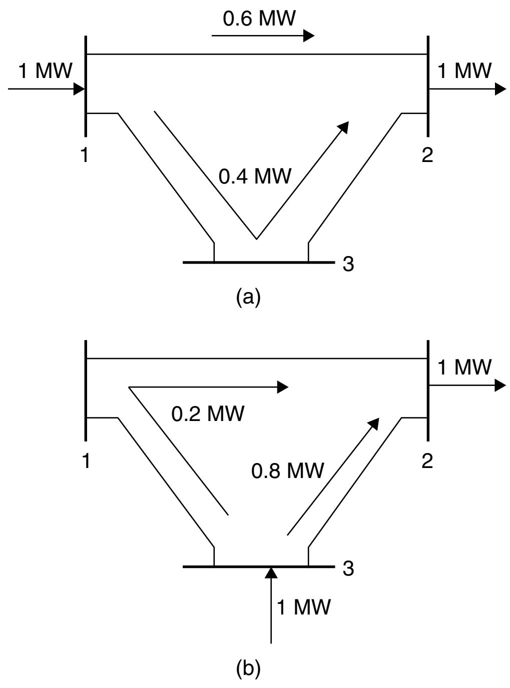


Figure 6.20 Incremental flows in the network due to an additional megawatt of load at bus 2 when this megawatt is produced at bus 1 (a) or at bus 3 (b)

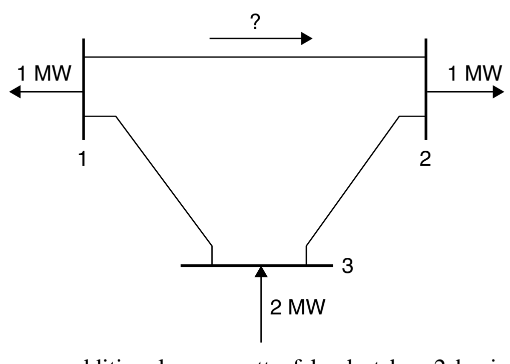


Figure 6.21 Supplying an additional megawatt of load at bus 2 by increasing production at bus 3 and reducing it at bus 3

is injected at bus 3 and taken out at bus 1, the flow on branch 1-2 would decrease by 0.4 MW. The other diagram shows that another 1 MW injected at bus 3 but taken out at bus 2 increases this flow by 0.2 MW. Overall, the flow on branch 1-2 decreases by 0.2 MW. Supplying an additional megawatt at bus 2 by increasing the production at bus 3 by 2 MW and decreasing it at bus 1 by 1 MW is thus acceptable because it keeps the flow on branch 1-2 below the maximum capacity of this branch. But is it optimal? This combination of injections has not simply kept the flow on branch 1-2 at its maximum: it has decreased it to 0.2 MW below this maximum. This means that we

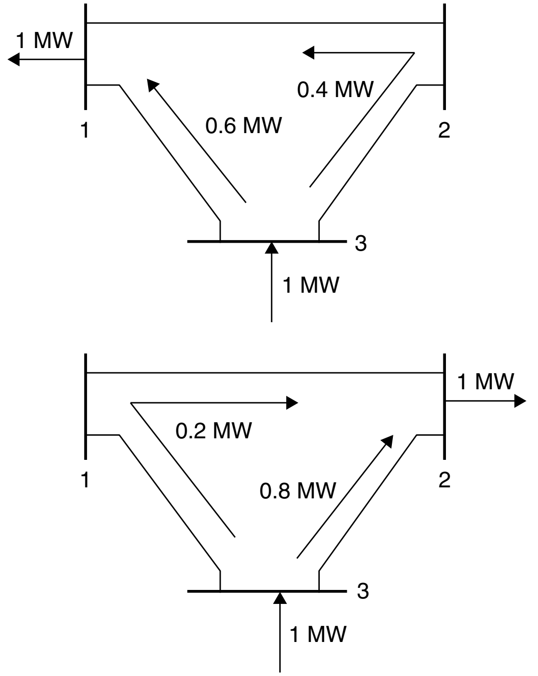


Figure 6.22 Application of the superposition theorem to the analysis of the conditions illustrated in Figure 6.21

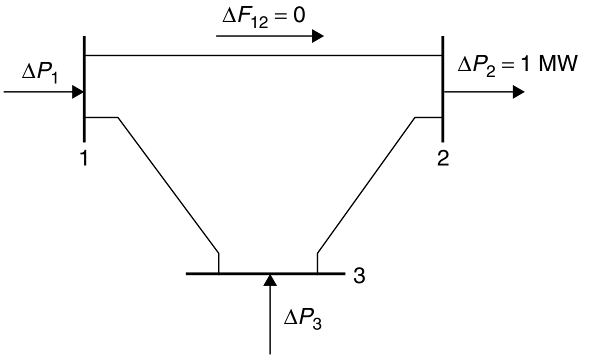


Figure 6.23 Formulation of the problem of supplying an additional megawatt of load at bus 2 without changing the flow on branch 1-2

have reduced too much the generation at bus 1, which is cheaper than the generation at bus 3.

Figure 6.23 illustrates the formulation that we will use to determine how we can supply an additional megawatt at bus 2 by redispatching generation at buses 1 and 3 without overloading branch 1-2. We must have

$$\Delta P_1 + \Delta P_3 = \Delta P_2 = 1\,\text{MW} \tag{6.63}$$

Using the sensitivities shown in Figure 6.20, we can also write

$$0.6\Delta P_1 + 0.2\Delta P_3 = \Delta F_{12} = 0\,\text{MW} \tag{6.64}$$

Solving these equations, we get

$$\Delta P_1 = -0.5\,\text{MW} \tag{6.65}$$

$$\Delta P_3 = 1.5\,\text{MW} \tag{6.66}$$

Supplying at minimum cost an additional megawatt at bus 2 therefore requires that we increase the output of Generator D by 1.5 MW and reduce the output of Generator A by 0.5 MW. The cost of this megawatt, and hence the nodal marginal price at bus 2, is thus

$$\pi_2 = 1.5 \cdot MC_{\text{D}} - 0.5 \cdot MC_{\text{A}} = 11.25\,\$/\text{MWh} \tag{6.67}$$

In summary, we observe that

- Generator A sets a price of 7.50 \$/MWh at bus 1. Generator B has a lower marginal cost (6.00 \$/MWh) but has no influence on the prices because it operates at its upper limit.
- Generator D sets a nodal marginal price of 10.00 \$/MWh at bus 3.
- At bus 2, the price is set at 11.25 \$/MWh by a combination of the prices of the other generators.

These observations can be generalized to more complex networks. In a system without transmission constraints, if we model all generators as having constant marginal costs, all generators, except one, either produce nothing or produce their maximum output. The exception is the marginal generator, whose output is such that the total generation is equal to the total load. Such a generator is said to be part-loaded. The marginal cost of this generator sets the price for the entire system because it provides the hypothetical additional megawatt that determines the marginal price. When a transmission limit constrains the economic dispatch, another generator becomes marginal in the sense that it is neither at its maximum or at its minimum output. In general, if there are m transmission constraints in the system, there will be $m + 1$ marginal generators. Each of these part-loaded generators sets the marginal price at the bus where it is connected. Nodal marginal prices at the other buses are determined by a combination of the prices of the marginal generators. This combination depends on the application of KVL to the constrained network. We will see shortly that this can lead to flows and prices that do not behave in an intuitively obvious manner.

6.3.2.4 Merchandising surplus

Before looking at these counter-intuitive situations, let us summarize the economic operation of this three-bus system. Table 6.5 shows the load and generation as well as

Table 6.5 Summary of the economic operation of the three-bus system

	Bus 1	Bus 2	Bus 3	System
Consumption (MW)	50	60	300	410
Production (MW)	335	0	75	410
Nodal marginal price ($/MWh)	7.50	11.25	10.00	–
Consumer payments ($/h)	375.00	675.00	3000.00	4050.00
Producer revenues ($/h)	2512.50	0.00	750.00	3262.50
Merchandising surplus ($/h)				787.50

the nodal price at each bus. It also shows the payments made by consumers and the revenues collected by generators if energy is bought and sold at nodal marginal prices. All of these quantities are calculated for one hour of operation at constant loads.

If we compare the sum of the consumer payments at all the buses and the sum of the generator revenues at all the buses, we notice that these two quantities do not match. More money is collected from the consumers than is paid to the generators. This difference is the merchandising surplus that we already encountered in our two-bus Borduria–Syldavia example. This surplus is again caused by congestion in the network. If the capacity of branch 1-2 was larger than 156 MW, we could implement the unconstrained economic dispatch. The marginal prices would then be identical at all nodes and the total amount collected by generators would be equal to the total amount paid by consumers.

6.3.2.5 Economically counter-intuitive flows

Locational differences in the price of producing goods are quite common in economics. A very intuitive example is the production of fruits and vegetables, which are cheaper to grow outdoors in a warm climate than in greenhouses in a cold climate. If markets are competitive, the price of produce will be lower in the warmer regions and higher in the colder regions. If trade between these regions is free, fruits and vegetables will be shipped from the regions with low prices to the regions with high prices. No rational trader would transport grapes from Alaska to California and hope to make a profit. In electricity networks, however, such economically counter-intuitive transportation does occur, even when operation is optimal. Figure 6.24 shows the flows and the nodal prices for the constrained dispatch of our three-bus system. The flows in branches 1-2 and 1-3 carry energy from a node with a lower marginal price to nodes with a higher marginal price. On the other hand, power flows from a higher price node to a lower price node on branch 2-3. This phenomenon occurs not because somebody behaves irrationally but because the laws of physics (KVL in this case) take precedence over the "laws" of the market.

Table 6.6 shows the surplus that each branch generates by carrying power in our three-bus network. In the case of Line 2-3, this amount is negative because the power flows from a node with a higher price to a node with a lower price. The sum over all the lines, however, is equal to the merchandising surplus that we calculated on a node-by-node basis in Table 6.5.

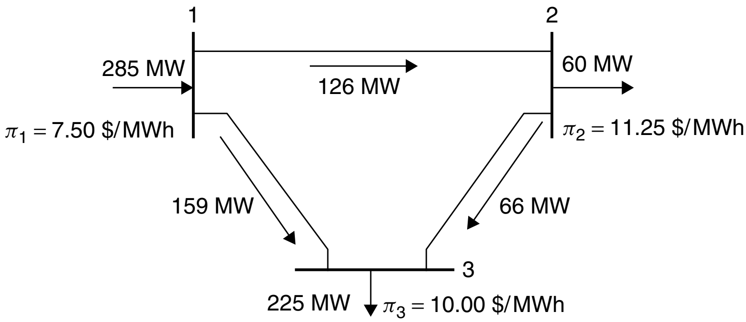


Figure 6.24 Nodal marginal prices and flows in the three-bus system. Power in branch 2-3 flows from a higher marginal price to a lower marginal price

Table 6.6 Contribution of each branch to the merchandising surplus of the three-bus system

Branch	Flow (MW)	"From" price ($/MWh)	"To" price ($/MWh)	Surplus ($/h)
1-2	126	7.50	11.25	472.50
1-3	159	7.50	10.00	397.50
2-3	66	11.25	10.00	−82.50
Total				787.50

6.3.2.6 Economically counter-intuitive prices

In our three-bus example, we have assumed so far that the flow on branch 1-2 could not exceed 126 MW. As we saw in Chapter 5, in a real system, the maximum flow allowable on a line is not necessarily fixed. If this flow is constrained by the thermal rating of the line, the limit depends on the weather conditions because wind and cold ambient temperature reduce the temperature rise inside the conductors. On the other hand, if the limit on the line flow is imposed by stability considerations, it will depend on the configuration of the rest of the system. Studying how the maximum flow on branch 1-2 affects the nodal marginal prices is thus not just a mathematical curiosity.

Table 6.7 summarizes the effect the maximum flow on this branch has on the operation and the economics of the three-bus system. Each row of this table corresponds to a different value of this flow. For each value, we have calculated the constrained dispatch and the nodal prices using the same procedure as above. We have also calculated the amounts collected by the generators and paid by the consumers, as well as the cost of producing the energy, the generator profits and the merchandising surplus. The last row of the table shows that a limit of 160 MW or more does not constrain the dispatch. Generator A is then the only marginal generator, and the nodal prices are uniform across the network. Under these conditions, the network does not produce a surplus. On the other hand, for a limit of less than 70 MW, there is no

Table 6.7 Effect of the maximum flow on branch 1-2 on the operation of the three-bus system

F_{12}^{MAX}	P_A	P_B	P_C	P_D	π_1	π_2	π_3	Generator costs	Generator revenues	Generator profits	Consumer payments	Congestion surplus
70	0.00	238.33	86.67	85.00	6.00	14.00	11.33	3493.33	3606.67	113.33	4540.00	933.33
80	0.00	255.00	70.00	85.00	6.00	14.00	11.33	3360.00	3473.33	113.33	4540.00	1066.67
90	0.00	271.67	53.33	85.00	6.00	14.00	11.33	3226.67	3340.00	113.33	4540.00	1200.00
100	3.33	285.00	36.67	85.00	7.50	14.00	11.83	3098.33	3681.67	583.33	4765.00	1083.33
110	20.00	285.00	20.00	85.00	7.50	14.00	11.83	2990.00	3573.33	583.33	4765.00	1191.67
120	36.67	285.00	3.33	85.00	7.50	14.00	11.83	2881.67	3465.00	583.33	4765.00	1300.00
130	60.00	285.00	0.00	65.00	7.50	11.25	10.00	2810.00	3237.50	427.50	4050.00	812.50
140	85.00	285.00	0.00	40.00	7.50	11.25	10.00	2747.50	3175.00	427.50	4050.00	875.00
150	110.00	285.00	0.00	15.00	7.50	11.25	10.00	2685.00	3112.50	427.50	4050.00	937.50
160	125.00	285.00	0.00	0.00	7.50	7.50	7.50	2647.50	3075.00	427.50	3075.00	0.00

generation dispatch that will supply the load without violating the flow constraint on branch 1-2.

For limits between 70 and 90 MW, Generator A does not produce energy, Generators B and C are part-loaded and Generator D is fully loaded. The nodal prices at buses 1 and 2 are thus 6.00 $/MWh and 14.00 $/MWh respectively while the nodal price at bus 3 is 11.33 $/MWh. This last value falls between the prices at the other two nodes. However, it must be above 10.00 $/MWh because Generator D is fully loaded.

One would expect that further increasing the line capacity should result in lower prices because the system would be less constrained. Table 6.7 shows that this is not necessarily the case. If we raise the limit up to 120 MW, the prices at nodes 1 and 3 increase while the price at node 2 remains constant. The system is not being operated inefficiently, however, because the cheaper generators (A and B) produce more power while the more expensive generator (C) produces less. Overall, the cost to generators of producing electrical energy decreases while the consumer payments, the generators' profits and the congestion surplus increase. In this case, measures to increase the transmission capacity benefit producers at the expense of consumers. Why did this happen? Increasing the flow on branch 1-2 allowed us to increase the output of the generators connected to bus 1. At some point, Generator B reached its maximum capacity and Generator A became the marginal generator, raising the nodal price at bus 1 to 7.50 $/MWh. Using superposition, we can then check that the price at bus 3 is given by

$$\pi_3 = \tfrac{1}{3}\pi_1 + \tfrac{2}{3}\pi_2 = 11.83\ \$/\text{MWh} \tag{6.68}$$

If we increase the limit beyond 120 MW, Generator C produces nothing. The price at node 2 becomes a combination of the prices at nodes 1 and 3 because Generator D is a marginal generator. Nodal prices, generator revenues, generator profits, consumer payments and congestion surplus all decrease until we reach the noncongested state for a limit of 156 MW.

6.3.2.7 More economically counter-intuitive prices

Let us now consider what happens if the capacity of branch 2-3 is reduced to 65 MW. Under these conditions, the minimum cost constrained dispatch is

$$\begin{aligned} P_A &= 47.5\,\text{MW} \\ P_B &= 285\,\text{MW} \\ P_C &= 0\,\text{MW} \\ P_D &= 77.5\,\text{MW} \end{aligned} \tag{6.69}$$

This dispatch produces the following flows

$$\begin{aligned} F_{12} &= 125\,\text{MW} \\ F_{13} &= 157.5\,\text{MW} \\ F_{23} &= 65\,\text{MW} \end{aligned} \tag{6.70}$$

The flow on Line 2-3 is thus the only one that is constrained. The marginal generators are A and D because Generator B is producing its maximum output and Generator C is not producing at all. Generator A thus sets π_1 at 7.50 \$/MWh while Generator D sets π_3 at 10.00 \$/MWh. To calculate the marginal price at node 2, we need to calculate the cost of an additional megawatt of load at that node. Since the marginal generators would supply this megawatt, we have

$$\Delta P_1 + \Delta P_3 = 1 \tag{6.71}$$

These increments in generation must be such that they keep the flow on branch 2-3 at its limit. Considering the relative reactances of the paths, we have

$$-0.4\Delta P_1 - 0.8\Delta P_3 = 0 \tag{6.72}$$

The negative signs arise because increasing the generation at either bus 1 or bus 3 while increasing the load at bus 2 decreases the flow on branch 2-3. Solving Equations (6.71) and (6.72), we get

$$\begin{aligned} \Delta P_1 &= 2\,\text{MW} \\ \Delta P_3 &= -1\,\text{MW} \end{aligned} \tag{6.73}$$

An additional megawatt at bus 2 would therefore cost us the price of 2 MW at bus 1 but we would save the price of 1 MW at bus 3. We therefore have

$$\pi_2 = 2 \times 7.50 - 1 \times 10 = 5.00\,\$/\text{MWh} \tag{6.74}$$

The marginal price at node 2 is thus lower than the price at either of the other buses, that is, lower than the marginal cost of any marginal generator!

6.3.2.8 Nodal pricing and market power

In our discussion, we have assumed so far that the nodal markets are perfectly competitive. The nodal price is thus equal to the marginal cost when the energy is produced using local generators. While this assumption greatly simplifies the analysis, it is highly questionable in practice, especially when the transmission network is congested. We will now show that KVL can make strategic bidding easy and profitable. Let us go back to our three-bus example with, as in the previous section, a constraint on branch 2-3 rather than branch 1-2. Suppose that Generator C at bus 2 desperately wants to produce some power. Such a situation could happen if the start-up cost of C is high and its owner decides that it is cheaper to produce at a loss for a while rather than having to restart the unit later. It could also happen if C is a cogeneration plant and the plant must run to produce the steam required for an industrial process. The owner of Generator C realizes that, if the plant is to run, she must bid below the current nodal marginal price of 5.00 \$/MWh. She therefore decides to bid at 3.00 \$/MWh. If the other generators bid at their marginal cost, the economic dispatch is then

$$
\begin{aligned}
P_A &= 35\,\text{MW} \\
P_B &= 285\,\text{MW} \\
P_C &= 90\,\text{MW} \\
P_D &= 0\,\text{MW}
\end{aligned}
\tag{6.75}
$$

However, this dispatch must be modified as follows to satisfy the constraint on branch 2-3:

$$
\begin{aligned}
P_A &= 32.5\,\text{MW} \\
P_B &= 285\,\text{MW} \\
P_C &= 7.5\,\text{MW} \\
P_D &= 85\,\text{MW}
\end{aligned}
\tag{6.76}
$$

Since Generators A and C are marginal, they set the nodal prices at buses 1 and 2 at 7.50 \$/MWh and 3.00 \$/MWh respectively. Generator D is operating at its upper limit and therefore does not set the price at bus 3. Using the same technique as above, we find that to supply an additional megawatt at bus 3, we would have to increase the output of Generator A by 2 MW and decrease the output of Generator C by 1 MW. The marginal price at node 3 is thus given by

$$\pi_3 = 2\pi_1 - \pi_2 = 12.00\,\$/\text{MWh} \tag{6.77}$$

The submission of a low bid at bus 2 has thus increased the price at bus 3 from 10.00 \$/MWh to 12 \$/MWh and the output of the generator at that bus from 77.5 MW to 85 MW. The low bid of Generator C thus has the counter-intuitive consequence of being very profitable for Generator D!

This fact is unlikely to go unnoticed by the owner of Generator D who may decide to see what happens if he raises his own bid to 20.00 \$/MWh. Under these conditions, the constrained dispatch becomes

$$\begin{aligned} P_{\mathrm{A}} &= 47.5\,\mathrm{MW} \\ P_{\mathrm{B}} &= 285\,\mathrm{MW} \\ P_{\mathrm{C}} &= 0.0\,\mathrm{MW} \\ P_{\mathrm{D}} &= 77.5\,\mathrm{MW} \end{aligned} \tag{6.78}$$

The marginal generators set the nodal prices at buses 1 and 3:

$$\begin{aligned} \pi_1 &= 7.50\,\$/\mathrm{MWh} \\ \pi_3 &= 20.0\,\$/\mathrm{MWh} \end{aligned} \tag{6.79}$$

On the other hand, supplying an additional megawatt at bus 2 would require increasing the injection at bus 1 by 2 MW and decreasing the injection at bus 3 by 1 MW. We therefore have

$$\pi_2 = 2\pi_1 - \pi_3 = 2 \times 7.50 - 1 \times 20.00 = -5.00\,\$/\mathrm{MWh} \tag{6.80}$$

Since the price at bus 2 is negative, consumers connected to that bus would be paid to consume, and generators would have to pay for the privilege of producing energy! Besides making life miserable for Generator C, Generator D increases its profit by raising its bid, even though its output decreases:

$$\Delta\Omega_D = 77.50 \times 20 - 85 \times 10 = \$700$$

Generator D is able to exert market power because it is in a very favorable position with respect to the constraint on branch 2-3. In fact, given the loads in the system, the output of Generator D cannot be reduced below 77.5 MW without violating the constraint. No matter what Generator D bids, its output will not drop below that level. Generator D thus enjoys a locational monopoly.

In general, network constraints increase opportunities for strategic bidding because not all generators are connected in locations where they can relieve a given constraint. In many cases, the number of generators that can effectively affect a constraint is small. Congestion in the transmission network can therefore transform a reasonably competitive global market into a collection of smaller local energy markets. Since these smaller markets inevitably have fewer active participants than the global market, some of them are likely to be able to exert market power. Such scenarios are not easy to detect or analyze. See Day *et al.* (2002) for a discussion of techniques that can be used to model strategic bidding when network constraints are taken into consideration.

6.3.2.9 A few comments on nodal marginal prices

Our collection of small examples has demonstrated that nodal prices at buses without a marginal generator can be higher, lower or in between the prices at buses with marginal generators. A nodal price can even be negative! We have also shown that unlike normal commodities, electrical energy can flow from a high price to a low price. All these effects are a consequence of the interactions between economics and KVL. They demonstrate the wisdom of the statement[2]: "Never trust a technique proven on the basis of a two-bus system".

These results may run against economic common sense, but they are mathematically correct. Trading electrical energy in a pool system that spans a possibly constrained transmission network requires the use of nodal marginal prices. These nodal marginal prices have been computed using an optimization procedure that maximizes the global welfare[3]. The system is thus operated in an economically efficient manner. Unfortunately, as we saw, these prices are dictated not only by economics but also by KVL. Even in our simple three-bus examples, understanding these prices takes time and effort. In a real system, the analysis is even more difficult. This puts electricity traders in the position of "having to take the computer's word for it", which is not entirely satisfactory when compared to trading in normal commodities.

Using a very small example allowed us to explain in details the factors driving the nodal prices. Skeptical readers may be pardoned for suspecting that the phenomena that we have described are an artifact of this small network and would not occur in a real system. This is unfortunately not the case. Counter-intuitive prices have been observed in several systems.

6.3.3 Losses in transmission networks

Transmitting electrical power through a network inevitably results in losses of energy. Since one or more generators must produce this lost energy and since these generators expect to be paid for all the energy they produce, a mechanism must be devised to take losses and their cost into account in electricity markets.

6.3.3.1 Types of losses

Before going further, we should make a distinction between the three different types of losses that are encountered in power systems. The first type is the *variable losses.* These losses are caused by the current flowing through the lines, cables and transformers of the network. Variable losses are also called *load losses, series losses, copper losses* or *transport-related losses*. As Equation (6.81) shows, these losses are proportional to the resistance R of the branch and to the square of the current in this branch. They can also be expressed as a function of the apparent power S or the real and reactive powers P and Q flowing through the branch. Since the voltage in a power system does not normally deviate much from its nominal value and since the active power flow is usually much larger than the reactive power flow, these variable losses can, as a first

[2]Attributed to L.A. Dale.

[3]The optimization procedure that we have used minimizes the production cost. Since we have assumed that the price elasticity of demand is zero, minimizing cost is equivalent to maximizing welfare. See Hogan (1992) for the generalization.

approximation, be treated as a quadratic function of the active power flow.

$$L^{\text{variable}} = I^2 R \approx \left(\frac{S}{V}\right)^2 R = \frac{P^2 + Q^2}{V^2} \cdot R \approx \frac{R}{V^2} \cdot P^2 = K \cdot P^2 \tag{6.81}$$

Note that Equation (6.81) is ambiguous because the power at the receiving end of the line is not the same as the power at the beginning of the line because of the variable losses!

The second type of losses is the *fixed losses*. Most of these losses are caused by hysteresis and eddy current losses in the iron core of the transformers. The rest is due to the corona effect in transmission lines. Fixed losses are proportional to the square of the voltage and independent of the power flows. However, since the voltage varies relatively little from its nominal value, as a first approximation, these losses can be treated as constant. Fixed losses are also called *no-load losses, shunt losses*, or *iron losses*.

The third type of losses is called *nontechnical losses*. This euphemism covers energy that is stolen from the power system.

Because of their quadratic dependence on the power flows, variable losses will be much more significant during periods of peak load. Averaged over a whole year, in western European countries, 1 to 3% of the energy produced is lost in the transmission system and 4 to 9% in the distribution system. Since variable losses are typically much larger than fixed losses, in the remainder of this discussion, we will consider only the variable losses.

6.3.3.2 Marginal cost of losses

Figure 6.25 shows a two-bus system in which a generator connected at bus 1 supplies a load connected at bus 2 through a line of resistance R. For the sake of simplicity, we will assume that the load is purely active and we will neglect the effect on the losses of the reactive power flow on the line. We will also assume that the voltage is equal to its nominal value at both buses. These assumptions allow us to express the losses as follows:

$$L = K \cdot D^2 \tag{6.82}$$

where D is the load at bus 2 and $K = \frac{R}{V^2}$. The generation at bus 1 is thus given by

$$G(D) = D + L = D + K \cdot D^2 \tag{6.83}$$

Figure 6.25 Two-bus system illustrating the calculation of the marginal cost of losses

If the load increases from D to $D + \Delta D$, the generation must increase by

$$\Delta G = G(D + \Delta D) - G(D) = \Delta D + 2\Delta D \cdot D \cdot K = (1 + 2D \cdot K)\Delta D \quad (6.84)$$

where we have neglected the second order term in ΔD. If the marginal cost of generation at bus 1 is c, the increase in the cost of generation due to an increase in load ΔD at bus 2 is

$$\Delta C = c(1 + 2D \cdot K)\Delta D$$

and the marginal cost at bus 2 is

$$\frac{\Delta C}{\Delta D} = c(1 + 2D \cdot K)$$

If we assume that competition is perfect in this system, the prices of energy at buses 1 and 2 are given by

$$\pi_1 = c \quad (6.85)$$

$$\pi_2 = \pi_1(1 + 2D \cdot K) \quad (6.86)$$

The difference in price between the two buses thus increases linearly with the line flow because the losses are a quadratic function of the load.

Because of the losses, the total amount paid by consumers at bus 2 exceeds the amount received by generators at bus 1. A merchandizing surplus MS thus arises in the network. This surplus is equal to the value of the energy sold at bus 2 minus the cost of purchasing the energy produced at bus 1:

$$MS = \pi_2 D - \pi_1(D + K \cdot D^2) \quad (6.87)$$

Using the expressions for the prices given in Equations (6.85) and (6.86), we get

$$\begin{aligned} MS &= c(1 + 2 \cdot K \cdot D)D - c(D + K \cdot D^2) \\ &= c \cdot K \cdot D^2 \end{aligned} \quad (6.88)$$

While less energy is consumed at bus 2 than is produced at bus 1, the difference in price between these two buses is sufficient to ensure that this surplus is always positive. In this case, the merchandising surplus is equal to the cost of supplying the losses because there is only one generator with a defined marginal cost. In a more complex network, one cannot obtain a closed-form expression similar to Equation (6.88). It is therefore impossible to establish a rigorous method for quantifying the cost of losses. The point of Equation (6.88) is thus only to show that the merchandising surplus is a rough indication of the cost of losses.

6.3.3.3 Effect of losses on generation dispatch

Let us go back to the Borduria–Syldavia interconnection that we introduced at the beginning of this chapter to study the effect of losses on the dispatch of the generating units. To keep matters simple, we will first assume that the interconnection is not congested and that its coefficient $K = \frac{R}{V^2} = 0.00005\,\text{MW}^{-1}$

Using Equations (6.6) and (6.7), the variable costs of producing energy in Borduria and Syldavia are given by

$$C_B(P_B) = \int_0^{P_B} MC_B(P)dP = 10P_B + \frac{1}{2} \cdot 0.01P_B^2 \tag{6.89}$$

$$C_S(P_S) = \int_0^{P_S} MC_S(P)dP = 13P_S + \frac{1}{2} \cdot 0.02P_S^2 \tag{6.90}$$

If the joint Borduria–Syldavia electricity market operates efficiently and competitively, at equilibrium it minimizes the total variable cost of producing electrical energy:

$$\min(C_B + C_S) = \min\left(10P_B + \frac{1}{2} \cdot 0.01P_B^2 + 13P_S + \frac{1}{2} \cdot 0.02P_S^2\right) \tag{6.91}$$

This minimization is subject to the power balance constraint. In other words, the power generated in Borduria and Syldavia must be equal to the sum of the load and the losses:

$$P_B + P_S = D_B + D_S + K \cdot F_{BS}^2 \tag{6.92}$$

Where R represents the resistance of the interconnection between the two countries and F_{BS} the active power flow at the Syldavia end of the interconnection. We again make the assumption that the voltage at all buses is kept at nominal value. To solve this optimization problem, we adopt an empirical approach in which we vary the flow F_{BS} and calculate the productions in Syldavia and Borduria using

$$P_S = D_S - F_{BS} \tag{6.93}$$

$$P_B = D_B + F_{BS} + K \cdot F_{BS}^2 \tag{6.94}$$

The total variable production cost can then be computed using Equations (6.89) and (6.90). Figure 6.26 shows how this total cost varies as a function of the flow on the interconnection when we do and do not consider the cost of the losses in the interconnector. This Figure shows that the losses reduce the optimal power transfer from 933 MW to 853 MW. Table 6.8 gives the details of these two optimal solutions. The losses make the Bordurian generators somewhat less competitive because a fraction of the energy that they produce is lost during its transfer to Syldavian customers. Production will therefore decrease in Borduria and increase in Syldavia. It is worth noting that the size of this redispatch is significantly larger than the amount of losses. Because of this redispatch, the marginal costs of production (and hence the local energy prices) are no longer equal in Borduria and Syldavia. A price differential of about 2.00 $/MWh arises. Syldavian consumers are indifferent between buying from local generators at 25.90 $/MWh or from Bordurian producers at 23.89 $/MWh and paying a 2.00 $/MWh transmission charge. Similarly, Bordurian consumers are indifferent

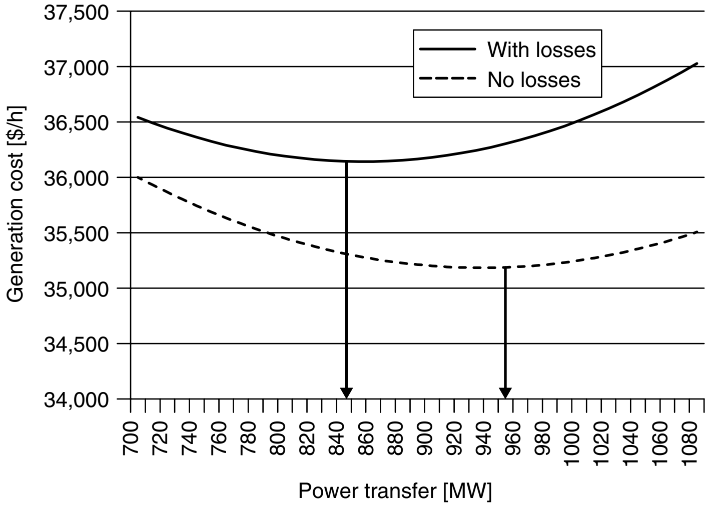


Figure 6.26 Total generation cost in the Borduria–Syldavia interconnection as a function of the flow on the interconnector when the losses in this interconnector are and are not taken into consideration. The coefficient $K = R/V^2$ of the interconnector is 0.00005 MW^{-1}. The resistance of the internal lines of the Bordurian and Syldavian networks has been neglected. The demands in Borduria and Syldavia are 500 MW and 1500 MW respectively

Table 6.8 Effect of losses on the operation of the Borduria–Syldavia interconnection

	Without losses	With losses
P_B (MW)	1433	1389
P_S(MW)	567	647
Losses (MW)	0	36
Power transfer (MW)	933	853
MC_B ($/MWh)	24.33	23.89
MC_S ($/MWh)	24.33	25.94
Total generation cost ($/h)	35 183	36 134

between buying from local producers or from more expensive Syldavian generators because they get rewarded for entering into a transaction that reduces the losses.

6.3.3.4 Merchandising surplus

Table 6.9 summarizes the operation of the Borduria–Syldavia interconnection when losses in the tie-line are taken into consideration. Consumers and producers buy and sell energy at their local price, which is assumed equal to the local marginal cost of production.

The presence of losses thus creates a merchandising surplus of 888.61 $/h. We get the same result if we treat this surplus as the "profit" made by the operator of the interconnection if it were to buy energy in Borduria and sell it in Syldavia. The quantity bought in Borduria would be 889 MW (i.e. 1389 – 500 MW) and the price would be

Table 6.9 Operation of the Borduria–Syldavia system when losses in the interconnection are taken into consideration

	Borduria	Syldavia	System
Consumption (MW)	500	1500	2000
Production (MW)	1389	647	2036
Nodal marginal price ($/MWh)	23.89	25.94	–
Consumer payments ($/h)	11 945.00	38 910.00	50 855.00
Producer revenues ($/h)	33 183.21	16 783.18	49 966.39
Merchandising surplus ($/h)			888.61

23.89 $/MWh. The quantity sold in Syldavia would be 853 MW (i.e. 1500 – 647 MW) and the price 25.94 $/MWh. The profit or surplus would then be

$$853 \times 25.94 - 889 \times 23.89 = 888.61\ \$/\mathrm{h}$$

Note that we do not get the same result if we multiply the price differential by the quantity transported because the losses make the flows at both ends of the line unequal.

6.3.3.5 Combining losses and congestion

Losses occur whether or not the system is congested. Let us consider the case in which the flow on the interconnection is constrained at 600 MW. The generators in Syldavia therefore produce 900 MW to meet the local load of 1500 MW. The nodal price (which we assume is equal to the marginal cost) in Syldavia is then

$$\pi_S = MC_S = 13 + 0.02 P_S = 31.00\ \$/\mathrm{MWh} \tag{6.95}$$

Using Equation (6.94), we can find the production of the generators in Borduria:

$$P_B = D_B + F_{BS} + K \cdot F_{BS}^2 = 500 + 600 + 18 = 1118\ \mathrm{MW} \tag{6.96}$$

The marginal cost and the nodal price in Borduria are then

$$\pi_B = MC_B = 10 + 0.01 P_B = 21.18\ \$/\mathrm{MWh} \tag{6.97}$$

The price differential is 9.82 $/MWh and is primarily due to the constraint. Table 6.10 summarizes the operation of the interconnection under these conditions.

Since the constraint on the interconnection reduces the flow, it also decreases the losses.

6.3.3.6 Handling of losses under bilateral trading

Because losses are not a linear function of the flows in the transmission system, the losses caused by a transaction do not simply depend on the amount of power traded and the location of the two parties involved in the transaction. These losses also depend on

Table 6.10 Operation of the Borduria–Syldavia system when both losses and congestion in the interconnection are taken into consideration

	Borduria	Syldavia	System
Consumption (MW)	500	1500	2000
Production (MW)	1118	900	2018
Nodal marginal price ($/MWh)	21.18	31.00	–
Consumer payments ($/h)	10 590	46 500	57 090
Producer revenues ($/h)	23 679	27 900	51 579
Merchandising surplus ($/h)			5511

all the other transactions taking place in the network. Allocating the losses or their cost between all the market participants is thus a problem that does not have a rigorous solution. Nevertheless, this cost must be paid and shared fairly. A fair mechanism is one in which participants that contribute more to losses (e.g. remote generators and consumers) pay a larger share than the others. See Conejo *et al.* (2002) for a discussion of various methods that have been proposed to allocate the cost of losses on an approximately fair basis.

6.3.4 Mathematical formulation of nodal pricing

In an actual power system, the size and complexity of the network are such that the prices of electrical energy obviously cannot be computed in the ad hoc manner that we have used in the examples of the previous section. A centralized market operator needs a mathematical formulation that can be used to calculate these prices in a systematic fashion. This market operator receives bids and offers from producers and consumers. It must then select the accepted bids and offers and set prices to clear the market. These decisions must maximize the economic welfare generated by the system while satisfying the security considerations that we discussed in Chapter 5. Our formulation is thus a constrained optimization problem. We will consider four, progressively more complex, variants of this optimization problem. Once again, to keep matters simple, we assume that competition is perfect throughout the network. The bids submitted by the generators are thus equal to their marginal costs.

6.3.4.1 Network with a single busbar

Let us first take a step backwards and see how we can formalize trading in electrical energy when the demand and the production are connected to the same busbar. This trivial network does not cause losses and does not limit the transfer of power between generation and load.

The economic welfare is equal to the difference between the benefit that consumers derive from the consumption of electrical energy and the cost of producing this energy. We will assume that the consumers' benefit is given by a function $B(D)$ of the total demand D and that the hourly cost of electrical energy is given by the function $C(P)$ of the total power P produced by the generators. This cost function $C(P)$ represents either the actual cost of production or the bids that the generators have submitted. As we have mentioned before, in a perfectly competitive market, these two functions are identical. Obviously, to maintain the stability of the system, the generation must be equal to the load. We can thus formulate the operation of this system as the following constrained optimization problem:

$$\text{Maximize } B(D) - C(P) \text{ subject to: } P - D = 0$$

The Lagrangian function of this problem is

$$\ell(D, P, \pi) = B(D) - C(P) + \pi(P - D) \tag{6.98}$$

where we have chosen, for reasons that will soon be obvious, to represent the Lagrange multiplier by π. The optimality conditions are obtained by setting the partial derivatives of the Lagrangian to zero:

$$\frac{\partial \ell}{\partial D} \equiv \frac{\mathrm{d}B}{\mathrm{d}D} - \pi = 0 \tag{6.99}$$

$$\frac{\partial \ell}{\partial P} \equiv -\frac{\mathrm{d}C}{\mathrm{d}P} + \pi = 0 \tag{6.100}$$

$$\frac{\partial \ell}{\partial \pi} \equiv P - D = 0 \tag{6.101}$$

From Equations (6.99) and (6.100), we get

$$\frac{\mathrm{d}B}{\mathrm{d}D} = \frac{\mathrm{d}C}{\mathrm{d}P} = \pi \tag{6.102}$$

Equations (6.102) and (6.101) formalize a point that we discussed in Chapter 2, namely, that consumers demand energy up to the point at which the marginal benefit they derive from this consumption equals the price they pay. Similarly, generators produce up to the point at which their marginal cost is equal to the price they receive. At equilibrium, in a perfectly competitive market, the price is equal to the value of the Lagrange multiplier of the optimization problem.

6.3.4.2 Network of infinite capacity with losses

Let us now consider the case in which demand and generation are connected to various nodes of a network. Since we assume that this network has an infinite capacity, transmission constraints are nonexistent and have therefore no effect on the prices of electricity. On the other hand, we will take into consideration the effect that the distribution of the generations and loads has on the losses in the network.

Instead of treating generation and load separately, it is convenient to consider the *net power injection* at each node. If both generators and consumers are connected to a particular node, this net injection is positive when the local production exceeds the demand and negative when the opposite holds. If we denote by I_k the net injection at node k, we have

$$I_k = P_k - D_k \tag{6.103}$$

In the absence of a network, net injections have to be equal to zero and, as we discussed above, the economic optimization has to be carried out independently at each node. A network therefore creates economic welfare by allowing trades between nodes with positive net injections and nodes with negative net injections.

At each node, we define a function $W_k(I_k)$ which is equal to the benefit to consumers at node k of the net injection I_k if it is negative and to minus the cost of producing this net injection if it is positive. Summing over all the nodes, we get the overall welfare created by the network:

$$W = \sum_{k=1}^{n} W_k(I_k) \tag{6.104}$$

As the objective of this optimization problem, we could choose to maximize this total welfare:

$$\max_{I_k}(W) = \max_{I_k}\left[\sum_{k=1}^{n} W_k(I_k)\right] \tag{6.105}$$

Since maximizing a function is equivalent to minimizing its opposite, we could also define the objective function as follows:

$$\min_{I_k}(-W) = \min_{I_k}\left\{\sum_{k=1}^{n}[-W_k(I_k)]\right\} \tag{6.106}$$

The second formulation is preferable because it is consistent with the traditional definition of the *optimal power flow (OPF) problem*[4]. In this problem, the demands are assumed to be completely insensitive to prices and fixed loads are specified at each node. The benefit accruing to consumers is thus constant and does not need to be taken into consideration in the optimization. Under these conditions, Equation (6.106) represents the minimization of the total cost of producing energy:

$$\min_{I_k}(-W) = \min_{I_k}\left\{\sum_{k=1}^{n} C_k(I_k)\right\} \tag{6.107}$$

Since we assume that the network has an infinite capacity, the only constraint on this optimization is the need to maintain a power balance. The sum of the net injections at all nodes must therefore be equal to the power losses in the branches of the network:

$$\sum_{k=1}^{n} I_k = L(I_1, I_2, \ldots I_{n-1}) \tag{6.108}$$

The power losses depend on the flows in the branches and thus on the net injections as shown by the function L in Equation (6.108). This function cannot depend on the injections at all the nodes. If it did, there would be no way to satisfy the power balance

[4]The constrained economic dispatch that we encountered in the three-bus example above is a simplified version of the optimal power flow problem.

because any adjustment in the injections would cause a change in the losses. To get around this difficulty, one bus in the system is designated as the *slack bus* and the injection at this bus is omitted from the variables of the function L. Given all the other net injections, the injection at the slack bus can then be adjusted to satisfy Equation (6.108). Since the concept of slack bus is purely mathematical and has no physical implications, the choice of a slack bus is entirely arbitrary. In Equation (6.108) and the rest of this chapter, we have chosen bus n as the slack bus.

We can now combine Equations (6.107) and (6.108) to build the Lagrangian function of the optimization problem:

$$\ell = \sum_{k=1}^{n} C_k(I_k) + \pi \left[L(I_1, I_2, \ldots, I_{n-1}) - \sum_{k=1}^{n} I_k \right] \tag{6.109}$$

The conditions for optimality are then

$$\frac{\partial \ell}{\partial I_k} \equiv \frac{\mathrm{d}C_k}{\mathrm{d}I_k} + \pi \left(\frac{\partial L}{\partial I_k} - 1 \right) = 0 \quad k = 1, \ldots n-1 \tag{6.110}$$

$$\frac{\partial \ell}{\partial I_n} \equiv \frac{\mathrm{d}C_n}{\mathrm{d}I_n} - \pi = 0 \tag{6.111}$$

$$\frac{\partial \ell}{\partial \pi} \equiv L(I_1, I_2, \ldots, I_{n-1}) - \sum_{k=1}^{n} I_k = 0 \tag{6.112}$$

Combining Equations (6.110) and (6.111), we get

$$\frac{\mathrm{d}C_k}{\mathrm{d}I_k} = \frac{\mathrm{d}C_n}{\mathrm{d}I_n} \left(1 - \frac{\partial L}{\partial I_k} \right) = \pi \left(1 - \frac{\partial L}{\partial I_k} \right) \quad k = 1, \ldots n-1 \tag{6.113}$$

The Lagrange multiplier π thus represents the marginal cost or marginal benefit of an injection of power at the slack bus. In a competitive context, this is the nodal price at the slack bus. The nodal prices at the other buses are related to the price at the slack bus by Equation (6.113). If an increase in the net injection at node k adds to the losses, we have

$$\frac{\partial L}{\partial I_k} > 0 \tag{6.114}$$

Hence, we get

$$\frac{\mathrm{d}C_k}{\mathrm{d}I_k} < \frac{\mathrm{d}C_n}{\mathrm{d}I_n} \tag{6.115}$$

The nodal price paid to generators at node k is thus smaller than the nodal price at the slack bus to penalize them for the additional losses that they would cause by injecting an increment of power in the network at that node. On the other hand, consumers at node k pay a lower price because an increase in load at that bus would reduce the losses. The opposite holds true if an increase in the net injection at node k reduces the losses. Finally, if the losses are neglected, the nodal prices at all buses are equal.

6.3.4.3 Network of finite capacity with losses

Chapter 5 discussed the constraints that security considerations place on the operation of a power system. We saw that the thermal capacity of lines and cables places a direct limit on the amount of power that they can carry. Maintaining the stability of the power system in the face of faults and outages also imposes limits on the flow of power on certain lines or groups of lines. We will model all these constraints as follows:

$$F_l(I_1, I_2, \dots I_{n-1}) \leq F_l^{\max} \quad l = 1, \dots m \tag{6.116}$$

Where F_l is the flow on branch l and $F_l^{\max}$ is the maximum value that this flow is allowed to take. m is the number of branches in the network. Note that the net injection at the slack bus is not included in the expressions for the branch flows to avoid creating an overdetermined problem.

We take these inequality constraints into account by adding them to the Lagrangian function of the previous optimization problem (Equation (6.109)):

$$\begin{aligned} \ell = {} & \sum_{k=1}^{n} C_k(I_k) + \pi \left[L(I_1, I_2, \dots, I_{n-1}) - \sum_{k=1}^{n} I_k \right] \\ & + \sum_{l=1}^{m} \mu_l [F_l^{\max} - F_l(I_1, I_2, \dots, I_{n-1})] \end{aligned} \tag{6.117}$$

The optimality conditions become

$$\frac{\partial \ell}{\partial I_k} \equiv \frac{\mathrm{d}C_k}{\mathrm{d}I_k} + \pi \left(\frac{\partial L}{\partial I_k} - 1 \right) - \sum_{l=1}^{m} \mu_l \frac{\partial F_l}{\partial I_k} = 0 \quad k = 1, \dots n-1 \tag{6.118}$$

$$\frac{\partial \ell}{\partial I_n} \equiv \frac{\mathrm{d}C_n}{\mathrm{d}I_n} - \pi = 0 \tag{6.119}$$

$$\frac{\partial \ell}{\partial \pi} \equiv L(I_1, I_2, \dots, I_{n-1}) - \sum_{k=1}^{n} I_k = 0 \tag{6.120}$$

$$\frac{\partial \ell}{\partial \mu_l} \equiv F_l^{\max} - F_l(I_1, I_2, \dots I_{n-1}) = 0 \quad l = 1, \dots m \tag{6.121}$$

$$\mu_l \cdot [F_l^{\max} - F_l(I_1, I_2, \dots I_{n-1})] = 0; \mu_l \geq 0 \quad l = 1, \dots m \tag{6.122}$$

Let us try to gain a better understanding of the implications of these equations by considering the special case where the flow on one line (say line i) is constrained. Since all the Lagrange multipliers μ_l are then equal to zero except μ_i, we get

$$\frac{\mathrm{d}C_k}{\mathrm{d}I_k} = \pi \left(1 - \frac{\partial L}{\partial I_k} \right) + \mu_i \frac{\partial F_i}{\partial I_k} \quad k = 1, \dots n-1 \tag{6.123}$$

$$\frac{\mathrm{d}C_n}{\mathrm{d}I_n} = \pi \tag{6.124}$$

$$\sum_{k=1}^{n} I_k = L(I_1, I_2, \ldots, I_{n-1}) \tag{6.125}$$

$$F_i(I_1, I_2, \ldots I_{n-1}) = F_l^{\max}; \mu_i > 0 \tag{6.126}$$

Equation (6.123) shows that the nodal price at every node (except the slack bus) is affected by a flow constraint on a single line. This influence depends on the shadow cost of the constraint (the Lagrange multiplier μ_i) and the sensitivity $\partial F_i/\partial I_k$ of the flow on branch i to the net injection at node k.

6.3.4.4 Network of finite capacity, dc power flow approximation

Solving Equations (6.123) to (6.126) is computationally difficult not only because they implicitly involve the solution of the power flow equations but also because they are nonlinear. Instead of using a full and accurate ac model, we could instead carry out this optimization on the basis of a linearized model called a *dc power flow*. The equations for the dc power flow are derived from the equation of the ac power flow by making the following simplifying assumptions:

- The resistance of each branch is negligible compared to the reactance
- The magnitude of the voltage at every bus is equal to its nominal value
- The differences in voltage angles across each branch are sufficiently small to allow the following approximations:

$$\cos(\theta_i - \theta_j) \approx 1$$
$$\sin(\theta_i - \theta_j) \approx \theta_i - \theta_j$$

Under these assumptions, the flow of reactive power in the system is negligible and the net active power injections are related to the bus voltage angles through the following set of equations:

$$I_i = \sum_{j=1}^{n} y_{ij}(\theta_i - \theta_j) \quad i = 1, \ldots n \tag{6.127}$$

Where y_{ij} represents the inverse of the reactance of the branch between nodes i and j and θ_i represents the voltage angle at node i. The flow of active power between nodes i and j is given by

$$F_{ij} = y_{ij}(\theta_i - \theta_j) \quad i, j = 1, \ldots n \tag{6.128}$$

Since the dc power flow neglects the resistances of the branches and thus ignores the losses, we no longer have to take into consideration an equality constraint similar to Equation (6.108). We have, however, introduced a new set of variables θ_i, which is counterbalanced by the new set of Equations (6.127). The constraints on the branch flows are given by

$$y_{ij}(\theta_i - \theta_j) \leq F_{ij}^{\max} \quad i, j = 1, \ldots n \tag{6.129}$$

Note that this formulation distinguishes two constraints for each branch: one on the flow from node i to node j and one on the flow from node j to node i. Obviously, only one of these two constraints can be binding at any time.

The Lagrangian function of this optimization problem is

$$\begin{aligned} \ell = \sum_{i=1}^{n} C_i(I_i) + \sum_{i=1}^{n} \pi_i \left[I_i - \sum_{j=1}^{n} y_{ij}(\theta_i - \theta_j) \right] \\ + \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_{ij}[F_{ij}^{\max} - y_{ij}(\theta_i - \theta_j)] \end{aligned} \tag{6.130}$$

Taking the partial derivatives of this function with respect to the variables, we get the following optimality conditions:

$$\frac{\partial \ell}{\partial I_i} \equiv \frac{\mathrm{d}C_i}{\mathrm{d}I_i} - \pi_i = 0 \quad i = 1, \ldots n \tag{6.131}$$

$$\frac{\partial \ell}{\partial \theta_i} \equiv -\sum_{j=1}^{n} y_{ij}(\pi_i - \pi_j + \mu_{ij} - \mu_{ji}) = 0 \quad i = 1, \ldots n-1 \tag{6.132}$$

$$\frac{\partial \ell}{\partial \pi_i} \equiv \sum_{j=1}^{n} y_{ij}(\theta_i - \theta_j) - I_i = 0 \quad i = 1, \ldots n \tag{6.133}$$

$$\frac{\partial \ell}{\partial \mu_{ij}} \equiv F_{ij}^{\max} - y_{ij}(\theta_i - \theta_j) \geq 0 \quad i, j = 1, \ldots n \tag{6.134}$$

$$\mu_{ij} \cdot [F_{ij}^{\max} - y_{ij}(\theta_i - \theta_j)] = 0; \mu_{ij} \geq 0 \quad i, j = 1, \ldots n \tag{6.135}$$

Note that there are only $n - 1$ equations like (6.132) because the voltage angle at one of the nodes (typically the slack bus) is taken as a reference and is thus not a variable. Equations (6.134) and (6.135) exist only for the pairs ij that correspond to network branches.

Equation (6.131) shows that with this formulation the Lagrange multipliers π_i are equal to the nodal prices. Let us define $C^{\min}$ as the value of the cost at the optimum. This cost depends on the limit on the flow on branch ij. Using Equation (6.130), we get

$$\frac{\partial C^{\min}}{\partial F_{ij}^{\max}} = \mu_{ij} \tag{6.136}$$

The Lagrange multiplier μ_{ij} thus represents the marginal cost of this constraint. It is expressed in \$/MWh because it represents the saving that would accrue each hour if the flow in branch ij could be increased by 1 MW.

In practice, operating a power system involves much more than placing fixed limits on the active power flows on some lines as we do in this linearized formulation. Operators must deal with the multitude of issues that we barely touched upon in Chapter 5. The dc power flow approximation is convenient and computationally efficient, but it would be foolish to think that it provides a sound basis for actually running a power system. However, it can be used to determine nodal marginal prices after the fact (i.e. "a posteriori"). In this approach, we assume that operators run the system in an optimal way using their best judgment and mathematical decision support tools that take into account the full nonlinearity of the system including the interactions between active and reactive power flows. Retrospectively, the actions of the operators can be represented by a set of limits on lines flows and a dispatch of the active power output of the generators that satisfies these constraints. This means that part of the optimization problem defined by Equations (6.131) to (6.135) has been solved. Specifically, the value of the following variables is known:

- The active power injection at each bus.
- The voltage angle at each bus.
- The binding line flow constraints.

If the generator costs are represented by linear or piecewise linear functions, the cost minimization dispatches the generators at their minimum or maximum output, or at an elbow point, except for those that have to be redispatched to satisfy the flow constraints. As we argued in our three-bus example, if there are m constraints, there will be $m + 1$ such marginal generators[5]. We can use Equation (6.131) to determine the price of electrical energy at the buses where these generators are connected. This equation cannot be used for generators that are operating at one of the breakpoints of their cost function because the derivative of this function is not defined at those points. This equation is also of no use at buses where a cost function is not available.

If there are m active constraints, we thus have

- $m + 1$ known prices π_i
- $n - m - 1$ unknown prices π_i
- m unknown Lagrange multipliers μ_{ij}

To find the value of these $n - 1$ unknown variables, we have the $n - 1$ Equations (6.132). If we denote by K and U, the sets of buses where the prices are respectively known and unknown, we can rearrange these equations as follows to have all the unknown variables are on the left-hand side:

$$Y_{ii}\pi_i - \sum_{j \in U} y_{ij}\pi_j + \sum_{j=1}^{n} y_{ij}(\mu_{ij} - \mu_{ji}) = \sum_{j \in K} y_{ij}\pi_j \quad i \in U; \quad i \neq \text{slack bus} \quad (6.137)$$

$$-\sum_{j \in U} y_{ij}\pi_j + \sum_{j=1}^{n} y_{ij}(\mu_{ij} - \mu_{ji}) = -Y_{ii}\pi_i + \sum_{j \in K} y_{ij}\pi_j \quad i \in K; \quad i \neq \text{slack bus} \quad (6.138)$$

[5]If the demand side takes an active part in the operation of the system by submitting bids to increase or decrease load, we could also have marginal loads.

where Y_{ii} represents the ith diagonal element of the admittance matrix of the network. Remember that μ_{ij} is nonzero only if the flow on the branch between nodes i and j is equal to its limits. The Lagrange multipliers μ_{ij} and μ_{ij} cannot be nonzero simultaneously because they correspond to flows on the same branch but in opposite directions. Even though we have written summations covering all the buses, the only nonzero terms are those for which bus j is at the opposite end of a branch connected to node i.

Example
Let us recalculate the nodal marginal prices for the three-bus example using this formulation. We will assume that Equation (6.58) shows the optimal dispatch that an operator has devised to operate the system at minimum cost while respecting the constraint on Line 1-2.

Generators A and D, located respectively at buses 1 and 3, are not operating at one of their limits. The price of electrical energy at these buses is thus known and equal to the marginal cost of these generators:

$$\pi_1 = \frac{dC_A}{dP_A} = 7.5\,\$/\text{MWh} \tag{6.139}$$

$$\pi_3 = \frac{dC_D}{dP_D} = 10.0\,\$/\text{MWh} \tag{6.140}$$

On the other hand, the price at bus 2 is unknown. We thus have

$$K = \{1, 3\}$$
$$U = \{2\}$$

The shadow cost μ_{12} of the constraint on the flow from bus 1 to bus 2 is also unknown. The other Lagrange multipliers μ_{ij} are equal to zero because the corresponding constraints are not binding. If we choose arbitrarily bus 3 as our slack bus, we can use the template provided by Equation (6.141) and (6.142) to write the following equations:

$$i = 1: \quad -y_{12}\pi_2 + y_{12}\mu_{12} = -Y_{11}\pi_1 + y_{13}\pi_3 \tag{6.141}$$

$$i = 2: \quad Y_{22}\pi_2 - y_{12}\mu_{12} = y_{21}\pi_1 + y_{23}\pi_3 \tag{6.142}$$

Since the admittance matrix of this network is

$$Y = \begin{pmatrix} -10 & 5 & 5 \\ 5 & -15 & 10 \\ 5 & 10 & -15 \end{pmatrix} \tag{6.143}$$

Equations (6.141) and (6.142) become

$$\begin{cases} 5\pi_2 - 5\mu_{12} = 25 \\ -15\pi_2 + 5\mu_{12} = -137.5 \end{cases} \tag{6.144}$$

Solving these equations gives

$$\begin{cases} \pi_2 = 11.25\,\$/\text{MWh} \\ \mu_{12} = 6.25\,\$/\text{MWh} \end{cases} \tag{6.145}$$

The nodal price at bus 2 is identical to the value we obtained in Equation (6.67). This is not surprising because that result was obtained under the same set of assumptions. The method that we used then to obtain the nodal prices was an ad hoc version of the general method described here. Note also that the shadow cost of the constraint on branch 1-2 is not equal to the difference between the marginal prices at nodes 1 and 2 because there is more than one path between these two nodes.

Handling of reactive power
In the development of this mathematical formulation, the only control variables that we have considered are the productions of the generators and the demands of the consumers. As we discussed in Chapter 5, to maintain the security of the system operators also adjust controls that affect mostly the reactive power flows. These control variables include the transformer taps, the voltage settings and the injections of reactive compensation devices. If we adopt a full ac OPF model, these variables are taken into account in the formulation of the optimization problem. Optimal prices can then be derived for the reactive power injection at each node in the system. While this result is theoretically interesting, its practical importance is limited. Reactive power does not need to be traded like active power. Indeed, because of the large reactive losses incurred in transmission lines, it cannot be traded over a wide area. Furthermore, the capacity to supply reactive power following a contingency has a value that is much higher than the short-run marginal cost of actually producing VARs.

6.3.5 Managing transmission risks in a centralized trading system

We have already mentioned in previous chapters that it is unusual for producers and consumers of commodities to sell or buy entirely through the spot market. In Chapter 4, we saw how participants in centralized electricity markets use contracts for difference to manage their exposure to the risks associated with fluctuations in the spot price. In that chapter, however, we assumed that the transmission network did not affect trading in electrical energy. We have now seen how security considerations limit the amount of power that can be transmitted across the network and create locational price differences. We must therefore consider the effect that congestion has on the feasibility of these contracts and what new contractual tools are needed to manage the risks associated with this congestion. While losses also create differences in nodal marginal prices, these differences are smaller and more predictable than the differences caused by congestion. We will therefore focus our discussion on the consequences of congestion. Our results can be generalized to cover the effect of losses.

6.3.5.1 The need for new contractual tools

In a centralized trading system, all the energy produced and consumed is traded physically through the pool. Producers and consumers inject or extract power into the network according to the instructions of the system operator. In return, they receive or pay the centrally determined price in effect at the location where they are connected. However, market participants are usually allowed to enter into bilateral financial contracts to protect themselves against the vagaries of the nodal prices. Let us examine what might happen when Borduria Power enters into a simple contract for difference with Syldavia Steel. This contract provides for the continuous delivery of 400 MW at a price of 30 $/MWh. As before, we will assume that there is no congestion within each of these two countries. There is thus a single nodal marginal price for Borduria (at which Borduria Power sells all its production) and a single nodal marginal price for Syldavia (at which Syldavia Steel buys all its consumption).

As long as there is no congestion on the interconnection, these two nodal marginal prices are equal. Generators in Borduria thus see the same price as consumers in Syldavia. In particular, if the spot price is 24.30 $/MWh, the contract between Borduria Power and Syldavia Steel is settled as follows:

- Borduria Power sells 400 MW at 24.30 $/MWh and receives $400 \times 24.30 = \$9720$ in payment.
- Syldavia Steel buys 400 MW at 24.30 $/MWh and pays $400 \times 24.30 = \$9720$.
- Syldavia Steel pays $400 \times (30 - 24.30) = \$2280$ to Borduria Power to settle the contract for difference.
- Borduria Power and Syldavia Steel have thus effectively traded 400 MW at 30 $/MWh.
- If the nodal prices had been higher than 30 $/MWh, Borduria Power would have made a payment to Syldavia Steel to settle the contract for difference.

Let us now consider what happens when the capacity of the interconnection is limited to 400 MW. We saw earlier that in this case, the nodal price for Borduria is 19.00 $/MWh while it rises to 35.00 $/MWh for Syldavia. Under these conditions

- Borduria Power sells 400 MW at 19.00 $/MWh and receives $400 \times 19.00 = \$7600$ in payment. According to the contract, it was supposed to receive $400 \times 30 = \$12\,000$. It is thus \$4400 short (\$12 000 − \$7600) and expects Syldavia Steel to pay this amount to settle the contract.
- Syldavia Steel buys 400 MW at 35.00 $/MWh and pays $400 \times 35.00 = \$14\,000$. According to the contract it was supposed to pay only $400 \times 30 = \$12\,000$. Therefore it expects Borduria Power to pay \$2000 to settle the contract.

These expectations are clearly incompatible. Contracts for difference that cover only the delivery of energy thus do not work when the transmission system is congested. Parties wishing to protect themselves against price variations must therefore contract not only for the energy that they produce or consume but also for the ability of the transmission system to deliver this energy.

6.3.5.2 Financial transmission rights

Bill, the economist who has been asked to study the electrical reconnection between Borduria and Syldavia, knows that without the certainty provided by contracts, the full benefits of the interconnection are unlikely to be realized.

While pondering the example discussed in the previous section, he calculates the total shortfall in the contract for difference, that is, the total amount of money that both parties should have received to settle the contract:

$$\$4400 + \$2000 = \$6400$$

He notices that this amount is exactly equal to the congestion surplus generated in the market, that is, the difference between the total amount paid by the consumers and the total amount received by the generators (see Table 6.2, p. 153):

$$\$62\,000 - \$55\,600 = 6400\,\$/\text{h}$$

Bill realizes that if Borduria Power and Syldavia Steel were given access to this congestion surplus, they would be able to settle equitably their contract for difference. To convince himself that this is not just a coincidence, Bill develops an analytical representation of the settlement of a contract for difference in the presence of congestion. He adopts the sign convention that a positive amount represents a revenue or a surplus and a negative amount an expense or a deficit. Given a contract for difference with a strike price π_C and an amount F, the total amount that a consumer such as Syldavia Steel expects to pay is

$$E_C = -F \cdot \pi_C \tag{6.146}$$

Conversely, the total amount that a producer such as Borduria Power expects to receive is

$$R_C = F \cdot \pi_C \tag{6.147}$$

The amounts that the consumer and producer respectively pay and collect on the spot market are respectively

$$E_M = -F \cdot \pi_S \tag{6.148}$$

and

$$R_M = F \cdot \pi_B \tag{6.149}$$

where Bill has taken into account the fact that the sale and the purchase are concluded at different nodal prices.

The amounts that the consumer and producer expect to pay or receive to settle the contract for difference are thus

$$E_T = E_M - E_C = -F \cdot \pi_S - (-F \cdot \pi_C) = F(\pi_C - \pi_S) \tag{6.150}$$

and

$$R_T = R_M - R_C = F \cdot \pi_B - F \cdot \pi_C = -F(\pi_C - \pi_B) \tag{6.151}$$

If the producer and consumer trade on the same spot market or there is no congestion in the system, we have $\pi_S = \pi_B$ and the contract can be settled because

$$E_{\mathrm{T}} = -R_{\mathrm{T}} \tag{6.152}$$

On the other hand, if $\pi_{\mathrm{S}} \neq \pi_{\mathrm{B}}$ both parties expect a payment and we have a total shortfall given by

$$E_{\mathrm{T}} + R_{\mathrm{T}} = F(\pi_{\mathrm{B}} - \pi_{\mathrm{S}}) \tag{6.153}$$

Bill then compares Equation (6.153) with the expression for the congestion surplus given in Equation (6.26). He observes that both of these two expressions involve the product of a power transfer with a difference in price between two markets. The congestion surplus involves the maximum power that can be transferred between the two locations while the shortfall given in Equation (6.153) pertains to a specific transaction. The congestion surplus should therefore be able to cover the shortfalls for contracts up to the maximum power transfer between the two markets.

Bill concludes that problems with contracts for differences can be solved if the parties acquired what is called Financial Transmission Rights (FTRs). FTRs are defined between any two nodes in the network and entitle their holders to a revenue equal to the product of the amount of transmission rights bought and the price differential between the two nodes. Formally, the holder of FTRs for F MWh between locations B and S is entitled to the following amount taken from the congestion surplus:

$$R_{\mathrm{FTR}} = F(\pi_{\mathrm{S}} - \pi_{\mathrm{B}}) \tag{6.154}$$

This amount is exactly what is needed to ensure that a contract for difference concluded between a producer at location B and a consumer at location S can be settled. Note that if there is no congestion in the transmission system, there is no price difference between locations B and S and the holder of FTRs receives no revenue. In this case, however, contracts for differences balance without problem.

Finally, Bill observes that holders of FTRs are indifferent about the origin or destination of the energy they consumer or produce. For example, a consumer in Syldavia who owns F MWh worth of FTRs between Borduria and Syldavia can either

- buy F MWh of energy on the Bordurian market for a price π_{B} and use its transmission rights to have it delivered "for free" in Syldavia – in this case, it effectively pays $F \cdot \pi_{\mathrm{B}}$;
- buy the F MWh of energy on the Syldavian market for a price π_S and use its share of the congestion surplus to offset the higher price it paid for the energy – in this case it pays $F \cdot \pi_{\mathrm{S}}$ but receives $F \cdot (\pi_{\mathrm{S}} - \pi_{\mathrm{B}})$.

In conclusion, FTRs completely isolate their holders from the risk associated with congestion in the transmission network. They provide a perfect hedge.

Bill must address one more question: how will producers and consumers obtain FTRs? Bill suggests that these rights should be auctioned. For each market period, the system operator would determine the amount of power that can be transmitted over the interconnection. FTRs for this amount of power would then be auctioned to the highest bidders. This auction would be open to all generators, consumers and even to speculators hoping to make a profit from locational differences in the price of electrical energy. The holder of these rights would be able to use them or resell them to another party. How much should bidders pay for an FTR? This depends on their expectations of the price differentials that might arise between the locations where these rights are defined. In the case of our example, if Bill's estimates of the energy prices in Borduria and Syldavia and of the transmission capacity of the interconnection during periods of congestion are correct, the auction should result in a maximum price of

$$35.00\,\$/\text{MWh} - 19.00\,\$/\text{MWh} = 16.00\,\$/\text{MWh}.$$

6.3.5.3 Point-to-point financial transmission rights

An important aspect of the definition of FTRs does not come out clearly from our two-bus Borduria/Syldavia example. FTRs are defined from any point in the network to any other point. These points do not have to be connected directly by a branch. The advantage of this approach from the perspective of a producer and a consumer who wish to enter into a transaction is that they do not have to concern themselves with the intricacies of the network. All they need to know is the bus where the power will be injected and the bus where it will be extracted. As far as they are concerned, the path that this power will take through the network is of no importance.

Let us now check how point-to-point FTRs work in our three-bus example. We will consider first the basic conditions that we analyzed in Sections 6.3.2.2 to 6.3.2.4. Figure 6.27 summarizes the secure economic operation of this system. Suppose that one of the consumers at bus 3 has signed a contract for difference with a generator connected to bus 1. This contract is for the supply of 100 MW at 8.00 $/MWh. The reference price for this contract is the nodal price at bus 1. As part of its risk management strategy, this consumer has also purchased 100 MW of FTRs from bus 1 to bus 3. As we saw, the nodal prices at buses 1 and 3 turn out to be 7.50 $/MWh and 10.00 $/MWh respectively. This contract is settled as follows:

- The consumer pays $100 \times 10.00 = \$1000$ to the market operator for extracting 100 MW at bus 3.
- The generator receives $100 \times 7.50 = \$750$ from the market operator for injecting 100 MW at bus 1.
- The consumer pays $100 \times (8.00 - 7.50) = \$50$ to the generator to settle the contract for difference.
- The consumer collects $100 \times (10.00 - 7.50) = \$250$ from the market operator for the FTRs it owns between buses 1 and 3.

The consumer thus pays a total of $800 for 100 MW, which is equivalent to a price of 8.00 $/MWh.

As we mentioned earlier, the money that the market operator needs to pay the owners of FTRs comes from the merchandising surplus that it collects because of network congestion. The market operator should therefore not sell more FTRs than

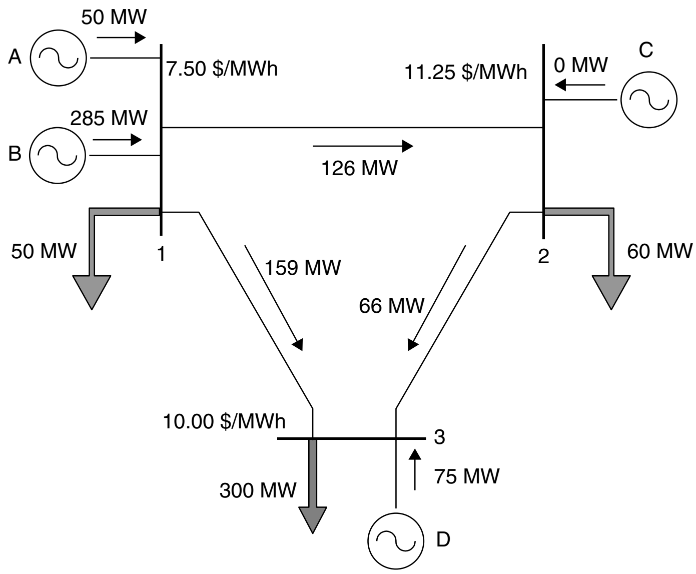


Figure 6.27 Secure economic operation of the three-bus system for the original network conditions

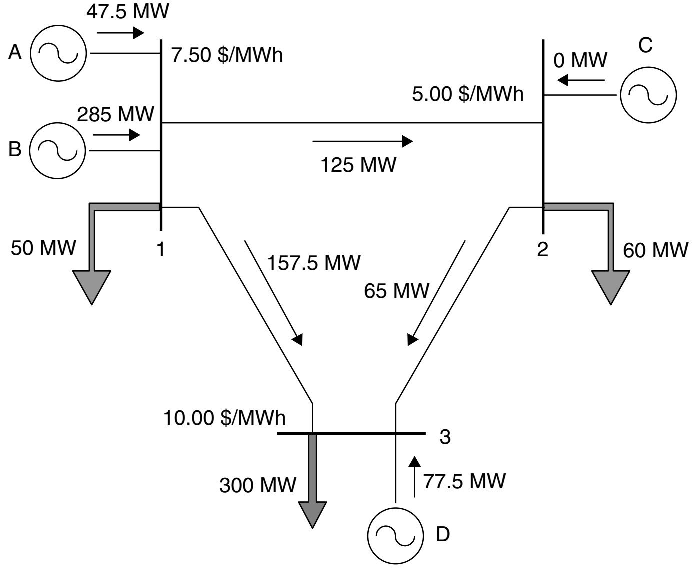


Figure 6.28 Secure economic operation of the three-bus system when the capacity of Line 2-3 is limited to 65 MW

the network can physically handle. Table 6.11 shows three combinations of FTRs that meet this simultaneous feasibility condition for the three-bus example.

Note that in each case, the sum of the revenues that the holders of the rights collect based on the nodal prices is equal to the merchandising surplus collected by the market operator (see Table 6.5, p. 163).

Table 6.11 Some feasible combinations of point-to-point FTRs for the three-bus example

Combination	Transmission rights			Settlement			
	From bus	To bus	Amount (MW)	From bus price ($/MWh)	To bus price ($/MWh)	Revenue ($)	Total ($)
A	1	3	225	7.50	10.00	562.50	787.50
	1	2	60	7.50	11.25	225.00	
B	1	3	285	7.50	10.00	712.50	787.50
	3	2	60	10.00	11.25	75.00	
C	1	3	275	7.50	10.00	687.50	787.50
	1	2	10	7.50	11.25	37.50	
	3	2	50	10.00	11.25	62.50	

Let us see what happens if, as we analyzed in Section 6.3.2.7, the capacity of Line 2-3 is limited to 65 MW. Figure 6.28 summarizes the operation of the system under these conditions. Table 6.12 summarizes the settlement of the three combinations of FTRs that are shown in Table 6.11.

Note that some of these FTRs have a negative value under these conditions. The holders of these rights therefore owe an additional amount to the market operator. This is a bit surprising because they actually paid to obtain these rights. However, it is not as bad as it may sound because contracts for difference can still be settled. Suppose, for example, that the load at bus 2 has signed a contract for difference with a generator at bus 1 for delivery of 60 MW at 8.00 $/MWh. The reference price for this contract is again the nodal price at bus 1. This consumer had also purchased 60 MW of transmission rights between nodes 1 and 2. This contract would be settled as follows:

- The consumer pays $60 \times 5.00 = \$300$ to the market operator for extracting 60 MW at bus 2.
- The generator receives $60 \times 7.50 = \$450$ from the market operator for injecting 60 MW at bus 1.
- The consumer pays $60 \times (8.00 - 7.50) = \$30$ to the generator to settle the contract for difference.
- The consumer pays $60 \times (7.50 - 5.00) = \$150$ to the market operator for the FTRs it owns between buses 1 and 2.

This consumer thus pays a total of $480, which is equivalent to the 8.00 $/MWh strike price in its contract for difference.

A simple calculation, similar to the one we performed in Table 6.5, shows that under these operating conditions, the market operator collects a merchandising surplus of $406.25. Unlike the previous case, this is somewhat short of the $412.50 shown in the last column of Table 6.12, which is the amount that the market operator must disburse to settle the FTRs. This discrepancy arises because the system operator was not able to deliver the point-to-point transmission capacity that it assumed when the FTRs were auctioned. Note that the market operator must collect money from the

Table 6.12 Settlement of combinations of point-to-point FTRs for the three-bus example when the capacity of Line 2-3 is limited to 65 MW

Combination	Transmission rights			Settlement			
	From bus	To bus	Amount (MW)	From bus price ($/MWh)	To bus price ($/MWh)	Revenue ($)	Total ($)
A	1	3	225	7.50	10.00	562.50	412.50
	1	2	60	7.50	5.00	−150.00	
B	1	3	285	7.50	10.00	712.50	412.50
	3	2	60	10.00	5.00	−300.00	
C	1	3	275	7.50	10.00	687.50	412.50
	1	2	10	7.50	5.00	−25.00	
	3	2	50	10.00	5.00	−250.00	

FTRs that have a negative value to be able to roughly balance its account book. FTRs must therefore be treated not as an option (where the contract is exercised only if it is profitable for its holder), but as an obligation that must be met in all cases.

6.3.5.4 Flowgate rights

Instead of being defined from point to point, FTRs can be attached to a branch or flowgate in the network. They are then called *flowgate rights* (FGRs). FGRs operate like FTRs except that the value of these rights is not tied to the difference in nodal prices, but to the value of the Lagrange multiplier or shadow cost associated with the maximum capacity of the flowgate. When a flowgate is not operating at its maximum capacity, the corresponding inequality constraint is not binding, and the corresponding Lagrange multiplier μ has a value of zero. The only FGRs that produce revenues are thus those that are associated with congested branches.

Since there is no difference between FTRs and FGRs in a two-bus system, we will jump directly to our three-bus example. Let us consider again the case of a consumer at bus 3 that wants to purchase 100 MW from a generator at bus 1 under the conditions depicted in Figure 6.27. To protect itself against fluctuations in nodal prices, this consumer must buy 100 MW of FGRs. We saw in Section 6.3.2.1 that only 60% of the power injected at bus 1 and extracted at bus 3 flows directly on branch 1-3. The other 40% flow through branches 1-2 and 2-3 as shown in Figure 6.29. The consumer should therefore purchase the following FGRs:

- 60 MW on branch 1-3
- 40 MW on branch 1-2
- 40 MW on branch 2-3

For the original network conditions depicted in Figure 6.28, the only branch that is operating at maximum capacity is branch 1-2. We saw in the example of Section 6.3.4.4 that the Lagrange multiplier $\mu_{12} = 6.25$ \$/MWh for these conditions. The Lagrange

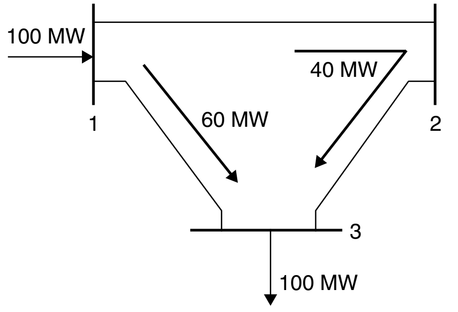


Figure 6.29 Transmission rights that must be acquired for a 100-MW transaction between buses 1 and 3

multipliers for the other inequality constraints are equal to zero. The consumer will therefore collect 40 MW × 6.25 $/MWh = $250 from its FGRs. This is exactly equal to the amount that this consumer would collect for 100 MW of point-to-point FTRs between buses 1 and 3. In this case, FGRs provide the same perfect hedge as FTRs.

Proponents of FGRs argue that in practice, market participants would not have to buy FGRs on all the branches through which the power they trade flows. Since at most a few branches of the transmission network are congested, they would only have to purchase transmission rights on these critical flowgates. This approach, however, leaves participants only partially hedged against the risk of congestion because it is often difficult to predict all the branches that will turn out to be congested.

Note that since the Lagrange multipliers corresponding to an inequality constraint are never negative, holders of FGRs never find themselves in the position of having to pay money back to the market operator. FGRs thus always behave like options.

6.3.5.5 The FTR versus FGR debate

At the time of writing this book, there is considerable debate on the advantages and disadvantages of FTRs and FGRs. Here is a summary of the main points of this debate:

- The market for FGRs should be more liquid than the market for FTRs because there are many more possible combinations of point-to-point rights than there are branches likely to be operated at maximum capacity.
- However, it may be difficult to predict which branches will be congested. Trading on a fixed set of critical flowgates may cause other branches to become congested.
- The value of FTRs is difficult to determine because the point-to-point transmission capacity changes with the configuration of the network. On the other hand, the maximum capacity of a given branch is much more constant, particularly if the flow on this branch is only limited by its thermal capacity.
- FGRs are simpler because there are typically only a few congested branches in a network. On the other hand, as soon as one branch is congested, all the nodal prices are different.

- Participants must take into consideration and understand the operation of the network when purchasing flowgate rights. In practice, this means that they must know the matrix of PTDFs. Participants who buy FTRs do not need to concern themselves with the operation of the network. They can base their decisions on their perception of the fluctuations in nodal prices.
- In a perfectly competitive market, FTRs, FGRs and even physical transmission rights are equivalent. If competition is less than perfect FGRs may provide more opportunities for gaming, particularly if trading focuses on a fixed set of flowgates.

It has been suggested that the best way to resolve this debate is to let the market decide which types of rights it prefers. See, for example, O'Neill *et al*. (2003) for a market design that combines both approaches.

6.4 Further Reading

Momoh (2000) provides a very readable discussion of the optimization techniques used in power systems. Hsu (1997) and Wu *et al*. (1996) are useful references on the principles of nodal pricing. Conejo *et al*. (2002) discusses the various methods used to handle losses. The seminal work on financial transmission rights was published by Hogan (1992). Flowgate rights were initially proposed by Chao and Peck (1996). For an introduction to the FTR versus FGR debate, see Chao *et al*. (2000) and Hogan (2000). Joskow and Tirole (2000) discuss in detail market power issues with physical transmission rights. Day *et al*. (2002) present a numerical method to analyze market power issues in large networks.

Chao H, Peck S, A market mechanism for electric power transmission, *Journal of Regulatory Economics*, **10**(1), 1996, 25–29.

Chao H, Peck S, Oren S, Wilson R, Flow-based transmission rights and congestion management, *Electricity Journal*, **13**(8), 2000, 38–58.

Conejo A J, Arroyo J. M, Alguacil N, Guijarro A. L, Transmission loss allocation: a comparison of different practical algorithms, *IEEE Transactions on Power Systems*, **17**(3), 2002, 571–576.

Day C J, Hobbs B. F, Pang J. -S, Oligopolistic competition in power networks: a conjectured supply function approach, *IEEE Transactions on Power Systems*, **17**(3), 2002, 597–607.

Hogan W W, Contract networks for electric power transmission, *Journal of Regulatory Economics*, **4**(3), 1992, 211–242.

Hogan W W, *Flowgate Rights and Wrongs*, 2000, John F. Kennedy School of Government, Harvard University, available online from ksghome.harvard.*edu/~.whogan.cbg.ksg/*.

Hsu M, An introduction to the pricing of electric power transmission, *Utilities Policy*, **6**(3), 1997, 257–270.

Joskow P, Tirole J, Transmission rights and market power on electric power networks, *RAND Journal of Economics*, **31**(3), 2000, 450–487.

Momoh J A, *Electric Power System Applications of Optimization*, Marcel Dekker, New York, 2000.

O'Neill R P, Helman U, Hobbs B. F, Stewart W. R Jr, Rothkopf M. H, A joint energy and transmission rights auction: proposal and properties, *IEEE Transactions on Power Systems*, **17**(4), 2002, 1058–1067.

Wu F, Varaiya P, Spiller P, Oren S, Folk theorems on transmission access: proofs and counterexamples, *Journal of Regulatory Economics*, **10**(1), 1996, 5–23.

6.5 Problems

6.1 Consider the power system shown in Figure P6.1. Assuming that the only limitations imposed by the network are imposed by the thermal capacity of the transmission lines and that the reactive power flows are negligible, check that the following sets of transactions are simultaneously feasible.

	Seller	Buyer	Amount
Set 1	B	X	200
	A	Z	400
	C	Y	300
Set 2	B	Z	600
	A	X	300
	A	Y	200
	A	Z	200
Set 3	C	X	1000
	X	Y	400
	B	C	300
	A	C	200
	A	Z	100

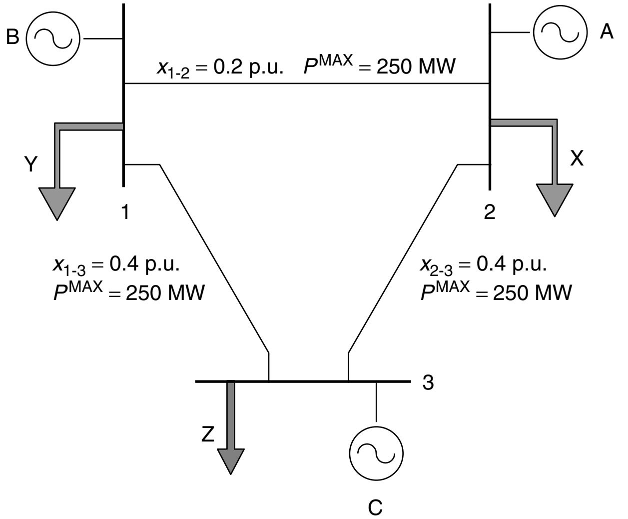


Figure P6.1 Three-bus power system for Problem 6.1

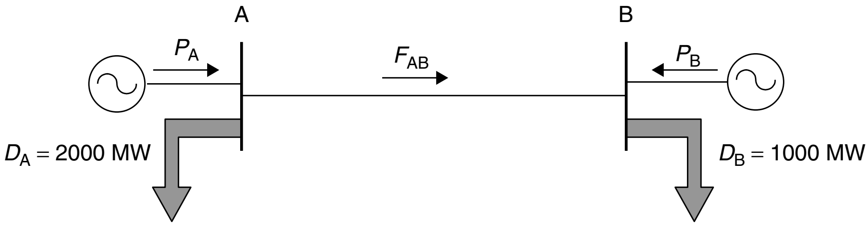


Figure P6.2 Two-bus power system for Problems 6.2, 6.3, 6.4, 6.10 and 6.11

6.2 Consider the two-bus power system shown in Figure P6.2. The marginal cost of production of the generators connected to buses A and B are given respectively by the following expressions:

$$MC_A = 20 + 0.03P_A \text{ \$/MWh}$$
$$MC_B = 15 + 0.02P_B \text{ \$/MWh}$$

Assume that the demand is constant and insensitive to price, that energy is sold at its marginal cost of production and that there are no limits on the output of the generators. Calculate the price of electricity at each bus, the production of each generator and the flow on the line for the following cases:

a. The line between buses A and B is disconnected

b. The line between buses A and B is in service and has an unlimited capacity

c. The line between buses A and B is in service and has an unlimited capacity, but the maximum output of Generator B is 1500 MW

d. The line between buses A and B is in service and has an unlimited capacity, but the maximum output of Generator A is 900 MW. The output of Generator B is unlimited.

e. The line between buses A and B is in service but its capacity is limited to 600 MW. The output of the generators is unlimited.

6.3 Calculate the generator revenues and the consumer payments for all the cases considered in Problem 6.2. Who benefits from the line connecting these two buses?

6.4 Calculate the congestion surplus for case (e) of Problem 6.2. Check your answer using the results of Problem 6.3. For what values of the flow on the line between buses A and B is the congestion surplus equal to zero?

6.5 Consider the three-bus power system shown in Figure P6.5 The table below shows the data about the generators connected to this system. Calculate the unconstrained economic dispatch and the nodal prices for the loading conditions shown in Figure P6.5.

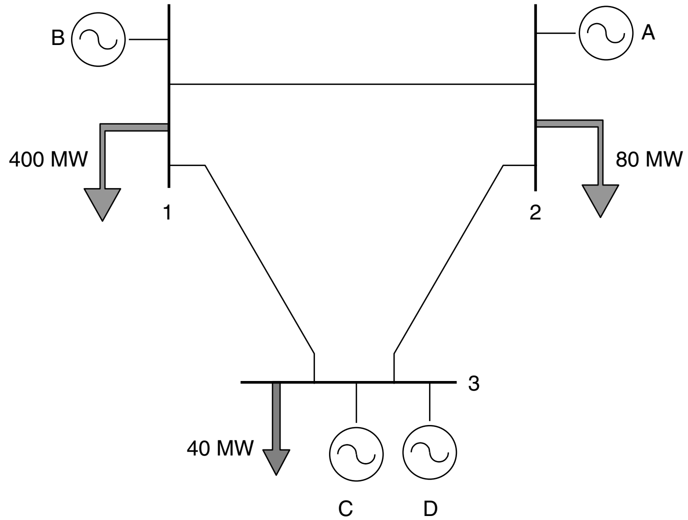


Figure P6.5 Three-bus power system for Problems 6.5 to 6.9 and 6.12 to 6.17

Generator	Capacity (MW)	Marginal cost ($/MWh)
A	150	12
B	200	15
C	150	10
D	400	8

6.6 The table below gives the branch data for the three-bus power system of Problem 6.5. Using the superposition principle, calculate the flow that would result if the generating units were dispatched as calculated in Problem 6.5. Identify all the violations of security constraints.

Branch	Reactance (p.u.)	Capacity (MW)
1-2	0.2	250
1-3	0.3	250
2-3	0.3	250

6.7 Determine two ways of removing the constraint violations that you identified in Problem 6.6 by redispatching generating units. Which redispatch is preferable?

6.8 Calculate the nodal prices for the three-bus power system of Problems 6.5 and 6.6 when the generating units have been optimally redispatched to relieve the constraint violations identified in Problem 6.7. Calculate the merchandising surplus and show that it is equal to the sum of the surpluses of each line.

6.9 Consider the three-bus power system described in Problems 6.5 and 6.6. Suppose that the capacity of branch 1-2 is reduced to 140 MW while the capacity of the other lines remains unchanged. Calculate the optimal dispatch and the nodal prices for these conditions. [Hint: the optimal solution involves a redispatch of generating units at all three buses]

6.10 Consider the two-bus power system of Problem 6.2. Given that $K = R/V^2 = 0.0001\,\text{MW}^{-1}$ for the line connecting buses A and B and that there is no limit on the capacity of this line, calculate the value of the flow that minimizes the total variable cost of production. Assuming that a competitive electricity market operates at both buses, calculate the nodal marginal prices and the merchandising surplus. [Hint: use a spreadsheet].

6.11 Repeat Problem 6.10 for several values of K ranging from 0 to 0.0005. Plot the optimal flow and the losses in the line, as well as the marginal cost of electrical energy at both buses. Discuss your results.

6.12 Using the linearized mathematical formulation (dc power flow approximation), calculate the nodal prices and the marginal cost of the inequality constraint for the optimal redispatch that you obtained in Problem 6.7. Check that your results are identical to those that you obtained in Problem 6.8. Use bus 3 as the slack bus.

6.13 Show that the choice of slack bus does not influence the nodal prices for the dc power flow approximation by repeating Problem 6.12 using bus 1 and then bus 2 as the slack bus.

6.14 Using the linearized mathematical formulation (dc power flow approximation), calculate the marginal costs of the inequality constraints for the conditions of Problem 6.9.

6.15 Consider the three-bus system shown in Figure P6.5. Suppose that Generator D and a consumer located at bus 1 have entered into a contract for difference for the delivery of 100 MW at a strike price of 11.00 \$/MWh with reference to the nodal price at bus 1, Show that purchasing 100 MW of point-to-point financial rights between buses 3 and 1 provides a perfect hedge to Generator D for the conditions of Problem 6.8.

6.16 What FGRs should Generator D purchase to achieve the same perfect hedge as in Problem 6.15?

6.17 Repeat Problems 6.15 and 6.16 for the conditions of Problem 6.9.

6.18 Determine whether trading is centralized or decentralized in your region or country or in another area for which you have access to sufficient information. Determine also the type(s) of transmission rights that are used to hedge against the risks associated with network congestion.

6.19 Determine how the cost of losses is allocated in your region or country or in another area for which you have access to sufficient information.

7
Investing in Generation

7.1 Introduction

In the previous chapters, we studied the economics of operating a power system using a given set of generating plants. In this chapter, we consider the possibility of adding or removing generation capacity. We first treat each generating plant independently. Taking the perspective of a potential investor, we examine the factors that affect the decision to build a new generating plant. We also consider the retirement of existing plants when their profitability becomes insufficient. For the sake of simplicity, we assume that all the revenues produced by these plants are derived from the sale of electrical energy and we neglect the additional revenues that a plant could obtain from the sale of ancillary services. We also assume that generating plants are not remunerated for providing capacity, that is, simply for being available in case their output is needed.

In the second part of the chapter, we discuss the provision of generation capacity from the perspective of the consumers. Electricity is so central to economic activity and personal well-being that consumers want a system that ensures a reliable supply of electrical energy. They expect their supply to remain available and affordable not only when the demand fluctuates but also when some generators are unable to produce because of technical difficulties. We must therefore consider whether profits from the sale of electrical energy result in a total generation capacity that is and remains sufficient to meet the consumers' expectations. Since in many electricity markets it has been decided that the answer to this question is negative, we discuss the additional incentives that can entice generating companies to provide the necessary capacity.

7.2 Generation Capacity from an Investor's Perspective

7.2.1 Building new generation capacity

An investor will finance a production facility if he or she believes that the plant will earn a satisfactory profit over its lifetime. More specifically, the revenues produced

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

by the plant should exceed the cost of building and operating the plant. Furthermore, this profit should be larger than the profit that this investor could realize by any other venture with a similar level of risk.

To make such an investment decision, this capitalist must compute the long-run marginal cost of the plant (including the expected rate of return) and forecast the price at which the output of this plant might be sold. Building a plant is a rational decision as long as the forecasted price exceeds the long-run marginal cost of the plant. In a liberalized electricity market, this reasoning is applicable to investments in generation capacity. Relying on this type of an investment decision leads to what is called "merchant generation expansion".

In practice, deciding to invest in a new generating plant is considerably more complex than this simplified theory would suggest. Both sides of the equation are indeed affected by a considerable amount of uncertainty. Construction delays and fluctuations in the price of fuel can affect the long-run marginal cost. On the other hand, the evolution of wholesale electricity prices over a long period is notoriously difficult to forecast because demand might change, competitors might enter the market or new, more efficient generation technologies might be developed. The development of merchant plants is often possible only when backed by upstream and downstream contracts. Upstream contracts guarantee a supply of fuel at a fixed price. Downstream contracts ensure that the electrical energy produced by the plant is sold at a price that is also fixed. Such arrangements eliminate the price risks over which the plant owner usually has very little control. This owner thus carries only the risk associated with operating the plant, that is, the risk that a failure might prevent the plant from producing electrical energy and honoring its contracts.

A generating plant, like any other machine, is designed to operate satisfactorily for a certain number of years. Investors who decide to build a generating plant base their decision on this estimated lifetime. For generating plants, this lifetime usually ranges from 20 to 40 years. Some hydro power plants, however, have considerably longer lifetimes.

7.2.1.1 Example 7.1

Bruce, a young consulting engineer, has been asked by Borduria Power to help them reach a preliminary decision on whether a new 500-MW coal-fired power plant should be built. Bruce begins by collecting some typical values for the essential parameters of a plant of this type. The table below shows the values he has gathered.

Investment cost	1021 $/kW
Expected plant life	30 years
Heat rate at rated output	9419 Btu/kWh
Expected fuel cost	1.25 $/MBtu

Adapted from DOE data cited by S. Stoft (2002).

Since he is only asked for a rough estimate, Bruce neglects the costs associated with starting up and maintaining the plant. Borduria Power has asked Bruce to use the

Internal Rate of Return (IRR) method to estimate the profitability of the plant. This method, which is also called *the discounted cash flow method*, measures the internal earning rate of an investment. To implement this method, Bruce must determine the net cash flow for each year of this plant's lifetime. Bruce starts by calculating the cost of building the plant:

Investment cost:

$$1021\,\$/\mathrm{kW} \times 500\,\mathrm{MW} = \$510\,500\,000$$

Next, Bruce needs to estimate the annual production of this plant. Ideally, the plant should operate at full capacity at all times. In practice, this is not possible because the plant has to be shut down periodically for maintenance and because there will inevitably be failures resulting in unplanned outages. Bruce therefore postulates a utilization factor of 80%. Under these conditions, we have:

Estimated annual production:

$$0.8 \times 500\,\mathrm{MW} \times 8760\,\mathrm{h/year} = 3\,504\,000\,\mathrm{MWh}$$

Bruce can then calculate the annual cost of producing this energy:

Annual production cost:

$$3\,504\,000\,\mathrm{MWh} \times 9419\,\mathrm{Btu/kWh} \times 1.25\,\$/\mathrm{MBtu} = \$41\,255\,220$$

Finally, to estimate the revenue, Bruce assumes that this generating plant will be able to sell the energy it produces for 32 $/MWh. The annual revenue is thus as follows:

Annual revenue:

$$3\,504\,000\,\mathrm{MWh} \times 32\,\$/\mathrm{MWh} = \$112\,128\,000$$

At this point, Bruce's spreadsheet looks like this:

Year	Investment	Production	Production cost	Revenue	Net cash flow
0	$510 500 000	0	0	0	−$510 500 000
1	0	3 504 000	$41 255 220	$112 128 000	$70 872 780
2	0	3 504 000	$41 255 220	$112 128 000	$70 872 780
3	0	3 504 000	$41 255 220	$112 128 000	$70 872 780
...	0	...	...	...	...
30	0	3 504 000	$41 255 220	$112 128 000	$70 872 780

Bruce has thus assumed that all the investment costs are incurred during the year immediately before the plant starts generating and that the production, revenue, production cost and net cash flow remain constant over the 30-year productive life of the plant. Using one of the functions provided by the spreadsheet software, Bruce then calculates the Internal Rate of Return for this stream of net cash flow. (See Sullivan *et al.* (2003) or a similar book for a detailed explanation of the calculation of the Internal Rate of Return. Most spreadsheets provide a function for calculating this quantity.) He obtains a value of 13.58%. Borduria Power must then compare this value with their "minimum acceptable rate of return" (MARR) before they make their decision.

Before committing itself, Borduria Power will also want to consider the risks associated with this project. Two issues are of particular concern in this case: what happens if the price of electrical energy does not meet expectations or if the plant cannot achieve the targeted utilization factor? Using his spreadsheet, Bruce can easily recalculate the Internal Rate of Return for a range of prices and utilization factors and produce the graph shown in Figure 7.1. Assuming that the plant achieves a utilization factor of 80%, the average price at which it sells electricity cannot drop below 30 \$/MWh if the plant is to achieve a MARR of 12%. On the other hand, the average selling price would have to increase considerably if the utilization factor drops much below 80%.

7.2.1.2 Example 7.2

After considering the results shown above, the board of Borduria Power is concerned by the size of the initial investment and the risk associated with this project. It therefore

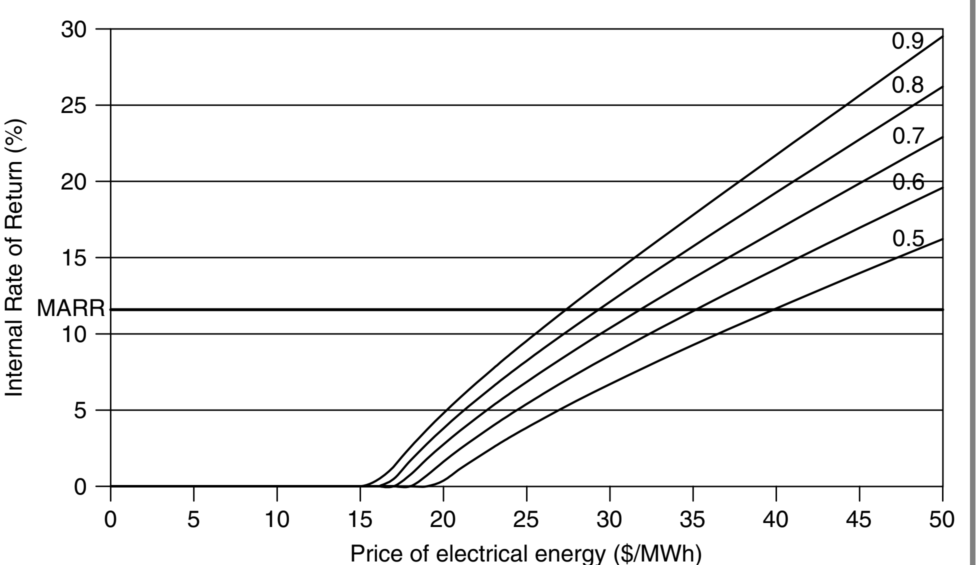


Figure 7.1 Internal Rate of Return for the coal unit of Example 7.1 as a function of the expected price of electrical energy for various values of the utilization factor

asks Bruce to perform a similar analysis for a combined-cycle gas turbine (CCGT) plant. As the table below shows, the economics of this technology are quite different from those of a coal plant. The initial investment is much smaller and the energy-conversion efficiency is much higher (because the heat rate is lower). On the other hand, a CCGT burns gas, a fuel that is much more expensive than coal.

Investment cost	533 $/kW
Expected plant life	30 years
Heat rate at rated output	6927 Btu/kWh
Expected fuel cost	3.00 $/MBtu

Adapted from DOE data cited by S. Stoft (2002).

Assuming the same utilization factor (80%) and the same electricity price (32 $/MWh), the annual production and the annual revenue from a CCGT plant of the same capacity would clearly be the same as for a coal plant:

Annual production:

$$0.8 \times 500\,\text{MW} \times 8760\,\text{h/year} = 3\,504\,000\,\text{MWh}$$

Annual revenue:

$$3\,504\,000\,\text{MWh} \times 32\,\$/\text{MWh} = \$112\,128\,000$$

On the other hand, the investment cost and the annual production cost would be different:

Investment cost:

$$533\,\$/\text{kW} \times 500\,\text{MW} = \$266\,500\,000$$

Annual production cost:

$$3\,504\,000\,\text{MWh} \times 6927\,\text{Btu/kWh} \times 3.00\,\$/\text{MBtu} = \$72\,816\,624$$

Using his spreadsheet, Bruce again analyzes how the Internal Rate of Return varies with the projected price of electrical energy and the utilization factor. The results of this analysis, which are shown in Figure 7.2, suggest that a CCGT plant might yield a higher rate of return than a coal plant. A decision between mutually exclusive investment alternatives, however, should not be based simply on a comparison of their respective rates of return. If the smaller investment (in this case the CCGT plant) produces an acceptable rate of return, the larger investment (the coal plant) should be treated as an incremental investment. Bruce therefore calculates the incremental net cash flow derived from this additional investment. This part of his spreadsheet looks like this:

Year	CCGT plant investment (A)	Coal plant investment (B)	CCGT plant production cost (C)	Coal plant production cost (D)	Incremental net cash flow (A) − (B) + (C) − (D)
0	\$266 500 000	\$510 500 000	0	0	−\$244 000 000
1	0	0	\$72 816 624	\$41 255 220	\$31 561 404
2	0	0	\$72 816 624	\$41 255 220	\$31 561 404
3	0	0	\$72 816 624	\$41 255 220	\$31 561 404
...	...	...	...	...	...
30	0	0	\$72 816 624	\$41 255 220	\$31 561 404

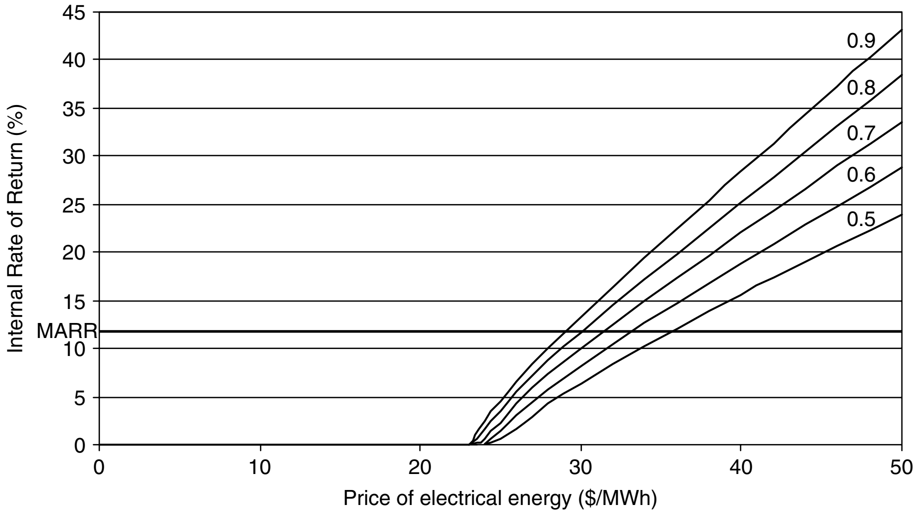


Figure 7.2 Internal Rate of Return for the CCGT unit of Example 7.2 as a function of the expected price of electrical energy for various values of the utilization factor

The columns showing the annual production and the annual revenue are not shown because they are identical for both technologies. Bruce then calculates the Internal Rate of Return corresponding to the cash flow stream shown in the last column and gets a value of 12.56%. If the MARR of Borduria Power is set at 12%, building a coal plant rather than a CCGT plant would be justified, at least for this value of the utilization factor. In his report, Bruce includes the graph shown in Figure 7.3 and points out to the board of Borduria Power that this incremental internal rate of return drops below 12% if the plant does not achieve an 80% utilization factor.

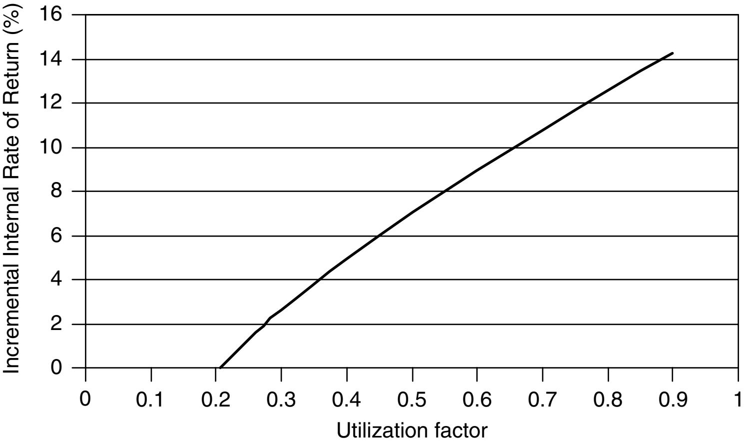


Figure 7.3 Incremental Internal Rate of Return that Borduria Power would achieve by investing in a coal plant rather than a CCGT plant

7.2.1.3 Example 7.3

While Borduria Power is considering building a plant that burns fossil fuel, Nick, the managing director of Syldavian Wind Power Ltd., has identified a promising site for the development of a 100-MW wind farm. The table below shows the plant parameters that Nick considers in his preliminary profitability calculation.

Investment cost	919 $/kW
Expected plant life	30 years
Heat rate at nominal output	0
Expected fuel cost	0

Adapted from DOE data cited by S. Stoft (2002).

The initial investment cost is thus: 919 $/kW × 100 MW = $91 900 000

Since the wind is free and the maintenance and operation costs are neglected in this first approximation, Nick does not need to consider an annual production cost. At 32 $/MWh, his best estimate of the average price of electricity during the lifetime of the plant happens to be identical to the one used by Borduria Power. Even though the site that Nick is considering has an excellent wind regime, the utilization factor of a wind farm is unlikely to exceed 35%.

Annual production:

$$0.35 \times 100\,\text{MW} \times 8760\,\text{h/year} = 306\,600\,\text{MWh}$$

Annual revenue:

$$306\,600\,\text{MWh} \times 32\,\$/\text{MWh} = \$9\,811\,200$$

Nick's spreadsheet thus looks like this:

Year	Investment	Production	Production cost	Revenue	Net cash flow
0	$91 900 000	0	0	0	–$91 900 000
1	0	306 600	0	$9 811 200	$9 811 200
2	0	306 600	0	$9 811 200	$9 811 200
3	0	306 600	0	$9 811 200	$9 811 200
...	0	...	...	...	...
30	0	306 600	0	$9 811 200	$9 811 200

Over the 30-year expected lifespan of the wind farm, the stream of net cash flow shown in the last column yields an Internal Rate of Return of 10.08%. This is less than the 12% return that Borduria Power considers acceptable but it meets the less exacting 10% MARR used by Syldavian Wind Power Ltd.

7.2.2 Retiring generation capacity

Once a generating plant goes into operation, its designed lifetime becomes a theoretical reference point around which the actual lifetime can deviate significantly. Market conditions may indeed change so much that the revenues generated by the plant no longer cover its operating costs. Unless there are reasons to believe that market conditions will improve, the plant must be retired. It is worth emphasizing that, in a competitive environment, such a decision is based solely on the future revenue and cost prospects for the plant and does not take into account the technical fitness of the plant or the sunk costs. On the other hand, recoverable costs (such as the value of the land on which the plant is built) are taken into consideration in such a decision because they represent revenues that become available.

7.2.2.1 Example 7.4

On the basis of Bruce's report, the board of Borduria Power decided to build the coal plant discussed in Example 7.1. Unfortunately, after only 15 years of operation, this plant has run into trouble. Because of increased demand, the price of the low-sulfur coal burnt by the plant has climbed to 2.35 \$/MBtu. Moreover, the government of Borduria has imposed an environmental tax of 1.00 \$/MWh on the output of fossil-fuel plants. Under these conditions, the marginal cost of production of the plant has risen as follows:

Marginal cost of production:

$$2.35\,\$/\text{MBtu} \times 9419\,\text{Btu/kWh} + 1\,\$/\text{MWh} = 23.135\,\$/\text{MWh}$$

At the same time, competitors have put into service more efficient CCGT plants that have depressed the average price of electrical energy to 23.00 \$/MWh. Assuming a utilization factor of 80%, the plant makes an annual loss as follows:

Annual loss:

$$(23.13465 - 23.00)\ \$/\text{MWh} \times 0.8 \times 500\,\text{MW} \times 8760\,\text{h/year} = \$473\,040$$

A market analysis commissioned by Borduria Power suggests that this situation is unlikely to change with more high efficiency plants due to come online over the next few years, leading to a drop in the price of electrical energy to 22.00 \$/MWh. Furthermore, a further increase in the price of low-sulfur coal is predicted. On this basis, the plant should be decommissioned immediately to recover the value of the land, which is estimated at \$10 000 000.

Before making a final decision on the decommissioning of the plant, Borduria Power investigates another possibility. Instead of burning low-sulfur coal, it could switch to high-sulfur coal, which costs only 1.67 \$/MBtu. This change would require an investment of \$50 000 000 for the installation of flue gas desulfurization (FGD) equipment. This installation would take a year and would have a detrimental effect on the heat rate of the plant, which would increase to about 11 500 Btu/kWh. Keeping the other economic assumptions unchanged, the effect of this plant refurbishment over the remaining 15 years of the plant's life is summarized in the following spreadsheet:

Year	Investment	Production	Production cost (incl. tax)	Revenue	Net cash flow
0	\$50 000 000	0	0	0	–\$50 000 000
1	0	3 504 000	\$70 798 320	\$77 088 000	\$6 289 680
2	0	3 504 000	\$70 798 320	\$77 088 000	\$6 289 680
3	0	3 504 000	\$70 798 320	\$77 088 000	\$6 289 680
...	0	...	...	...	...
15	0	3 504 000	\$70 798 320	\$87 088 000	\$16 289 680

The revenue for year 15 includes the estimated value of the land. While this investment would create a positive cash flow, the net present value of this cash flow stream is equal to −£4 763 285. This demonstrates that this investment would not be profitable and that the plant should definitely be retired.

7.2.3 Effect of a cyclical demand

If the demand for electricity increases without a compensating increase in generation capacity, or if the available capacity decreases because generating units are decommissioned, the market price for electrical energy will rise. This price increase will provide an incentive for generating companies to invest in new plants. As we discussed in Chapter 2, production capacity expands up to the point where the market price is equal to the long-run marginal cost of producing electrical energy. On a superficial level, investments in electrical generation capacity are thus governed by the same principles that apply to the production of any commodity. We must, however, take into consideration the cyclical nature of the demand for electrical energy and the significant effect that the weather has on this demand. While electrical energy is by no means the only commodity that exhibits such fluctuations in demand, it is the only one that cannot easily be stored. Its production must therefore match the consumption, not only over a period of days or weeks but on a second-by-second basis. When considering investments in generation capacity, we are not interested in the timing of the peaks and valleys of the daily, weekly or yearly load profile. Instead, we need to know for how many hours each year the load is less than a given value. This information is encapsulated in what is called a *load-duration curve*. Figure 7.4 shows this curve for the Pennsylvania–Jersey–Maryland (PJM) system during the year 1999. From this curve, we can observe that the load in this system was never less than 17 500 MW or greater than 51 700 MW during that year. We can also see that it exceeded 40 000 MW for only about 8% of the 8760 h that made up 1999.

Since the shape of this curve is typical, we can conclude that the installed generation capacity in a power system must be substantially larger than the demand averaged over a whole year. This means that not all generators can expect to have a utilization factor close to unity. In an efficient competitive market, a generator with a lower marginal production cost will usually have the opportunity to produce before a generator with a higher marginal cost. Cheap generators therefore achieve higher utilization factors than less efficient ones. As one would expect, prices are thus lower during periods of low demand than during periods of high demand. Competition during periods of low demand should also be more intense than during periods of high demand. During periods of high demand, most generators are indeed fully loaded and are not actively competing. Competition is then limited to a smaller number of expensive generating

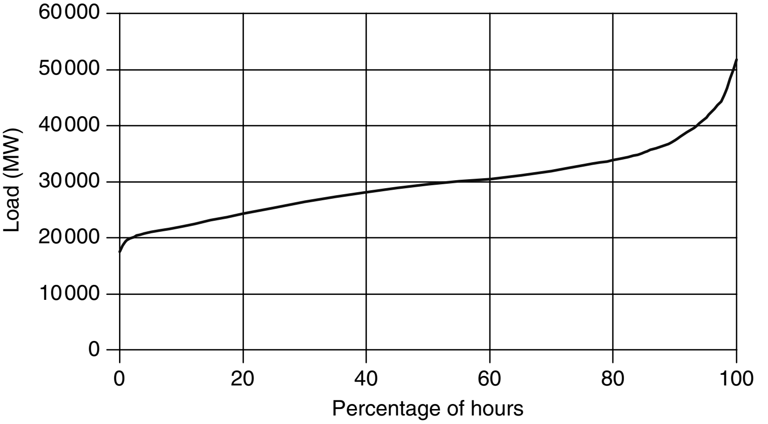


Figure 7.4 Load-duration curve for the PJM system during the year 1999. (Source: www.pjm.com)

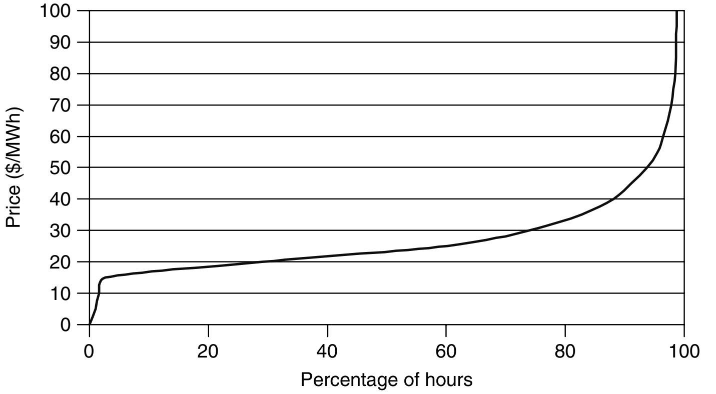


Figure 7.5 Price-duration curve for the PJM system during the year 1999. To enhance readability, the price axis has been limited to 100 $/MWh. In fact, the curve peaks at 1000 $/MWh for 100% of hours. The price shown is the average locational marginal price in the system. (Source: www.pjm.com)

units. On the other hand, during periods of low demand, even efficient generators may have to compete to remain on-line and avoid incurring start-up costs.

These conjectures are confirmed by the *price-duration curve* of the PJM system for 1999 shown in Figure 7.5. This curve shows the fraction of the number of hours of the year during which the price was less than a given value. The shape of this curve is similar to the shape of the load-duration curve, but its extremities are distorted by the variations in the intensity of competition.

As we discussed in previous chapters, the marginal generator sets the market price. In an efficient competitive market, this generator has no incentive to bid higher or lower than its marginal cost of production. The market price is therefore equal to the cost of *producing* the last MWh. While the marginal generator will not lose money on the sale of the electrical energy it produces, it will not collect any economic profit either. On the other hand, the infra-marginal generators collect an economic profit. Despite its name, all of this economic profit cannot be passed on to the shareholders to remunerate their investment. Part of this money must be used to cover the fixed costs of the plant. These fixed costs include the maintenance costs that do not vary with production, the personnel costs, taxes on the value or capacity of the plant and the opportunity cost on the recoverable value of the plant.

But what happens to the least efficient and therefore most marginal of generators? This generator is only called when the load reaches its maximum on a very hot summer day or a very cold and dark winter evening. By definition, it is never infra marginal and thus never collects economic profits while other generators set the market price. If this generator bids purely on the basis of its short-run marginal cost of production, it will never collect money to pay its fixed costs. If this generator is to stay in business, it must factor its fixed costs in its bids.

7.2.3.1 Example 7.5

Faced with the introduction of a competitive electricity market, Harry, the vice president for operations of Syldavia Electric, must decide what to do with the Skunk

River plant, an aging oil-fired 50-MW generating station. This plant has a heat rate of 12 000 Btu/kWh and burns a fuel that costs 3.00 $/MBtu. Since it is among its least efficient, Syldavia Electric has used the Skunk River plant in recent years only during periods of extremely high load. Harry first calculates the fixed costs associated with this plant and gets a figure of about $280 000 per year. He then tries to estimate the minimum bid that the Skunk River plant should submit to recover all of its costs. This means that the revenue produced by the plant should be equal to its total operating cost, which can be expressed as follows:

$$\text{Production [MWh]} \times \text{bid [\$/MWh]} = \text{fixed cost [\$]} + \text{production [MWh]} \times \text{heat rate [Btu/kWh]} \times \text{fuel cost [\$/MBtu]}$$

Since Harry does not really know what the production will be, he decides to calculate the minimum bid for a range of values. To simplify the calculations, he assumes that, if it runs, the unit will produce at maximum capacity. Harry can then calculate the minimum bid price as a function of the number of hours of operation or of the utilization factor. Figure 7.6 summarizes his results and shows that the minimum bid increases beyond 1000 $/MWh if the unit runs only 5 h per year. For comparison, the marginal cost of production of the Skunk River plant is 36 $/MWh. Such prices might seem totally unreasonable, but from Harry's perspective they are entirely justified. Moreover, he is also very likely to obtain whatever bids he submits during these few hours a year because the only competition he is likely to face would be from plants that are in the same situation as his. While consumers might balk at these prices, the alternative is not to consume at all during these periods because the Skunk River Plant could be described as the plant of last resort.

Owners of very marginal plants who use this approach to set their prices need to estimate the number of hours that the plant is likely to run each year. This is not an easy task because this quantity is affected by several random factors. An average or expected value can be predicted on the basis of historical data and predictions for load growth

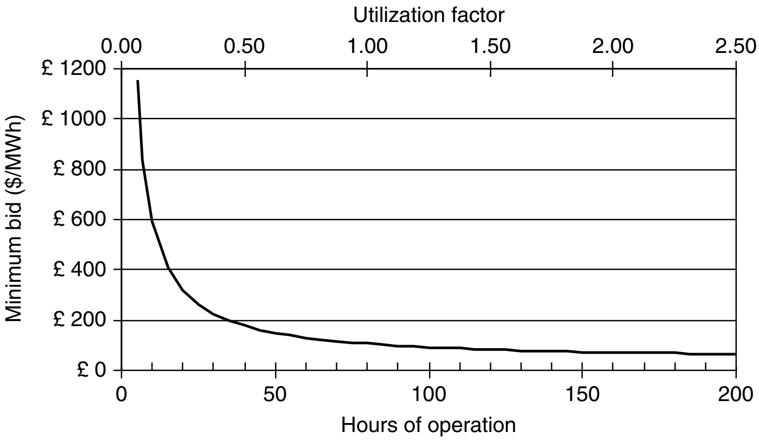


Figure 7.6 Minimum bid that the Skunk River unit of Example 7.5 should submit to recover both its fixed and variable costs, as a function of the number of hours of operation at full load that this unit expects to achieve

and the retirement of other generating plants. The actual value, however, may deviate significantly from this average. For example, during a warm winter or a cool summer the demand may reach critical values less frequently and for fewer hours than expected. Similarly, insufficient precipitations may reduce the availability of hydroenergy and increase the need for thermal generation. Depending on the conditions, very marginal plants therefore might be called frequently or not at all. While their revenues may be acceptable if averaged over a number of years, the possibility of losing money during one or more years might be more than a risk-averse owner would tolerate. Such plants would therefore be prime candidates for retirement.

7.3 Generation Capacity from the Customers' Perspective

In the first part of this chapter, we took the perspective of a potential investor who tries to decide whether to build a new generating plant. We also considered the decision process of the owner of a plant who is trying to decide whether the time has come to shut it down. In this section, we consider the provision of generation capacity from the consumers' perspective. In a completely deregulated environment, there is no obligation on any company to build power plants. The total generation capacity that is available for supplying the demand therefore arises from individual decisions based on perceptions of profit opportunities.

We will first discuss whether the decision to build generation capacity can be driven entirely by the profits obtainable in the markets for electrical energy. If this is not satisfactory, market-based expansion needs to be supplemented by a centralized mechanism designed to ensure or encourage the availability of a certain amount of capacity. We will discuss four forms that this mechanism can take.

At this point, we must note that consumers have some reliability expectations when they purchase electrical energy. This means that the energy should be delivered when consumers demand it and not at some other time. Since generating units are occasionally unavailable because of breakdowns or the need to perform maintenance, the power system must have available more generation capacity than is necessary to meet the peak demand. Increasing this generation capacity margin improves the reliability of the system. The mechanisms described below should therefore be judged not only on their ability to provide enough generation capacity to deliver the electrical energy demanded by consumers but also to meet their reliability expectations.

7.3.1 Expansion driven by the market for electrical energy

Some power-system economists (see for example the comprehensive exposition of this point of view by Stoft, 2002) argue that electrical energy should be treated like any other commodity. They insist that if electrical energy is traded on a free market, there should be no centralized mechanism for controlling or encouraging investments in generating plants. If left alone, markets should determine the optimal level of production capacity called for by the demand. Interfering with the market distorts prices and incentives. Centralized planning and subsidies lead to overinvestment or underinvestment, both of which are economically inefficient.

As we saw in Chapter 2, if the demand for a commodity increases, or its supply decreases, the market price rises and encourages additional investments in production capacity and a new long-run equilibrium is ultimately reached. Because of the cyclical nature of the demand for electricity and its lack of elasticity, price increases on electricity markets are usually not smooth and gradual. Instead, we are likely to observe price spikes (i.e. very large increases in price over short periods of time) when the demand approaches the total installed generation capacity. Figure 7.7 illustrates this phenomenon. A typical supply function is represented by a stylized, three-segment, piecewise linear curve. The first, moderately sloped, segment represents the bulk of the generating units in a reasonably competitive market. The second segment, which has a much steeper slope, represents the peaking units that are called infrequently. The third segment is vertical and represents the supply function when all the existing generation capacity is in use. An almost vertical line represents the low-elasticity demand function. This demand function moves horizontally as the demand fluctuates over time.

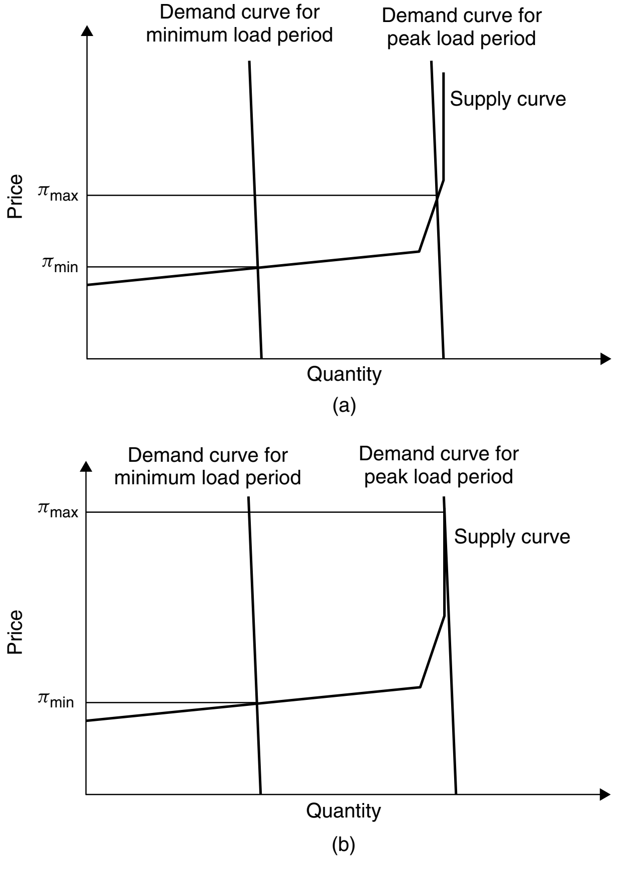


Figure 7.7 Illustration of the mechanism leading to price spikes in markets for electrical energy. (a) Sufficient generation capacity, (b) Insufficient generation capacity

Two curves are shown: one representing the minimum demand period and the other the peak demand period. The intersections of these curves with the supply function determine the minimum and the maximum prices. When the generation capacity is tight but sufficient to meet the load (Figure 7.7(a)), the price rises sharply during periods of peak demand because the market price is determined by the bids of generating units that operate very infrequently. These price spikes are much higher when all the generation capacity is in use under peak load conditions (Figure 7.7(b)). Such a situation could happen because the installed generation capacity has not kept up with the load growth, because some generation capacity has been retired or because some generation capacity is unavailable (for example, because a drought has reduced the amount of available hydroenergy). Under these conditions, the only factor that would limit the price increase is the elasticity of the demand. Note that this reasoning assumes that the demand does respond to prices. If this is not the case, load must be shed to prevent a collapse of the system.

In practice, these price spikes are significantly higher than what Figure 7.7 suggests and they are sufficient to increase substantially the average price of electricity even if they occur only a few times a year. Price spikes therefore provide a vivid signal that there is not enough capacity to meet the demand and the "extra" revenue that they produce is essential to give generating companies the incentive that they need to invest in new generation capacity or keep older units available.

These price spikes are obviously very expensive (one might say painful) for the consumers. They should thus encourage them to become more responsive to price signals. As the price elasticity of the demand increases, the magnitude of the spikes decreases, even if the balance between peak load and generation capacity does not improve. Price spikes also give consumers a strong incentive to enter into contracts that encourage generators to invest in generation capacity.

According to this theory, which is supported by quite sophisticated mathematical models (see the classical work of Schweppe *et al.*, 1988), an equilibrium should eventually be reached. At this equilibrium, the balance between investments in generation capacity and investments in load control equipment is optimal and the global welfare is maximum. However, several practical and socio-behavioral problems and their political consequences may prevent this equilibrium from being reached. First, the technology required to make a sufficient portion of the demand responsive to short-term price signals is not yet available. Until such technology becomes widely available and accepted, it may be necessary to implement quantity rationing rather than price rationing when demand exceeds supply. In other words, the system operator may have to disconnect loads to keep the system in balance during periods of peak demand. Widespread load disconnections are extremely unpopular and often have disastrous social consequences (accidents, vandalism). They are also economically very inefficient. Their impact can be estimated using the value of lost load (VOLL), which is several orders of magnitude larger than the cost of the energy not supplied. Consumers are not used to such disruptions and it is unlikely that their political representatives would tolerate them for any length of time.

When consumers are exposed to spot prices and should be adjusting their demand, price spikes are very unpopular. Since the origin of these spikes is rather hard to explain and justify to nonspecialists, consumers often believe that they are being ripped off. Price spikes also have socially unacceptable consequences such as forcing poor and vulnerable people to cut back on their consumption of electrical energy for essential needs such as heating, cooking and air conditioning. To be politically acceptable, many electricity markets therefore incorporate a price cap designed to prevent large price spikes. Such a price cap obviously removes a good part of the incentive for building or keeping generation capacity.

An electricity market that relies on spikes in the price of electrical energy to encourage the development of generation capacity is not necessarily good for investors either. Price spikes may not materialize and the average price of electricity may be substantially lower if the weather is more temperate or if higher than average precipitations make hydroenergy more abundant. Basing investment decisions on such signals represents a significant risk for investors. This risk may deter them from committing to the construction of new plants.

Finally, simulation models developed by Ford (1999, 2001) suggest that the time it takes to obtain planning permission and build a power plant can create instability in the market. Instead of increasing smoothly in response to load growth, the generation capacity goes through a series of boom-and-bust cycles. A lack of generation capacity produces very high electricity prices and triggers a boom in power plant construction. This boom results in a glut of capacity that depresses prices and discourages construction until the overcapacity has been resorbed. Such boom-and-bust cycles are not in the long-term interest of either producers or consumers.

In conclusion, it appears that relying solely on the market for electrical energy and its price spikes to bring about enough generation capacity is unlikely to give satisfactory results. This approach presumes that consumers are only buying electrical energy and that this transaction can be treated as the purchase of a commodity. In practice, consumers do not purchase only electrical energy but a service that can be defined as the provision of electrical energy with a certain level of reliability.

7.3.2 Capacity payments

The risk associated with leaving generation development to the invisible hand of the electrical energy market is often judged to be too great. Market designers in several countries and regions have decided that, rather than occasionally paying generators large amounts of money because of shortage-induced price spikes, it was preferable to pay to them a smaller amount on a regular basis. Payments would be proportional to the amount of capacity made available by each generator. These capacity payments form a stream of revenue that is separate from the money that generators derive from the market for electrical energy. They should cover at least part of the capital cost of new generating units and encourage generating companies to keep available units that are rarely called upon to produce energy. By increasing the total available capacity, these payments reduce, but do not eliminate, the likelihood of shortages. More production capacity also enhances competition and moderates prices in the market for electrical energy.

Capacity payments thus reduce the risks described in the previous section and spread them among all consumers, irrespective of the timing of their demand for electrical energy. At least in the short term, this socialization of the cost of peaking energy benefits the risk-averse market participants, be they consumers or producers. In the long term, this approach reduces the incentive for economically efficient behavior – too much capital may be invested in generation capacity and too little on devices that consumers can use to control their demand.

There are also practical difficulties. First, there is no clear way to determine either the total amount of money to be spent on capacity payments or the rate to be paid per megawatt of installed capacity. Second, such a system can also lead to endless debates about how much should be paid to each generator. For example, it can be argued that thermal and hydro plants do not make the same contribution to reliability because a drought may limit the output of the hydro units. Finally, since capacity payments are not tied to any performance criteria, it is not obvious that they actually do enhance reliability.

In an attempt to get around these difficulties, the Electricity Pool of England and Wales adopted an alternative approach. The centrally determined price of electrical energy during each period t was increased by a capacity element equal to

$$CE_t = VOLL \times LOLP_t \tag{7.1}$$

where $VOLL$ is the value of lost load (determined through a customer survey and updated annually to take inflation into account) and $LOLP_t$ is the loss of load probability during period t. Since this probability depends on the margin between the load and the available capacity and on the outage rates of the units, this capacity element fluctuated from one period to the next and occasionally caused significant price spikes. In exchange for this capacity element, the energy price was capped at the value of lost load. The money collected during each half-hourly period through this capacity element was divided among all the generating units that submitted bids to supply energy, regardless of whether or not they had been scheduled to produce energy. The capacity element was intended to send a short-term signal to consumers, while the associated capacity payments were designed to provide a long-term incentive to producers. While the capacity payments procured significant revenues to the generators and helped maintain a substantial generation capacity, their dependence on LOLP (loss of load probability), which is a short-term variable, made them easy to manipulate by the large generating companies. These payments were abandoned when the New Electricity Trading Arrangements were introduced.

7.3.3 Capacity market

Rather than fixing the total amount or the rate of capacity payments, some regulatory authorities set a generation adequacy target and determine the amount of generation capacity required to achieve this target. All energy retailers and large consumers (i.e. all entities that buy energy) are then obligated to buy their share of this requirement on an organized capacity market. While the amount of capacity to be purchased on this market is determined administratively, its price depends on the capacity on offer and may be quite volatile.

Implementing a capacity market that achieves its purpose is not a simple matter. Several important issues must be considered carefully. The first, and probably most fundamental, of these issues is the time step of the market, that is, the period over which each retailer's capacity obligations are calculated. Retailers prefer a shorter period (say a month or less) because it reduces the amount of capacity that they have to purchase during periods of light load. A shorter time step also increases the liquidity of the capacity market. On the other hand, a longer time step (e.g. a season or a year) favors generators and encourages the building of new capacity. In an interconnected system, it discourages existing generators from selling their capacity in a neighboring market. A longer time step also matches more closely the period over which the regulatory authorities evaluate the reliability of the system.

The installed generation capacity must exceed the peak demand because generators can fail at any time. Unreliable generators therefore increase the size of the required generation capacity margin and impose a cost on the entire system. Choosing an appropriate method to evaluate and reward the performance of generators is thus the second major issue in the design of a capacity market. This method should track as closely as possible the reliability of the system. It should reward reliable plants and encourage the retirement of unreliable units. For example, in the Pennsylvania–Jersey–Maryland market, the amount of capacity that generators are allowed to offer in the capacity market is derated by their historical forced outage rate. Generators thus have an incentive to maintain or improve the availability of their units. Ideally, these performance criteria should be refined to encourage generators not only to build or retain capacity but also to operate in such a way that they are available during critical periods.

An energy buyer that does not purchase its share of the target generation capacity benefits from the installed capacity margin paid for by the other market participants. It also has a cost advantage in the energy market. A deficiency payment or penalty must therefore be imposed on any player that does not meet its obligations. The level of this payment and the rules for its imposition should be set in a way that encourages proper behavior and discourages free riders.

7.3.4 Reliability contracts

Ideally, every consumer should decide freely and independently how much it is willing to pay for reliability. In a mature electricity market, it would then be able to enter into a long-term contract with a generator that would guarantee the delivery of energy with this level of reliability. Such long-term contracts would give generators the incentive to build the amount of capacity required to achieve the desired level of reliability.

Until electricity markets achieve the level of maturity where this approach becomes possible, a central authority (for example, the regulator or the system operator) could purchase reliability on behalf of the consumers. Instead of setting a target for installed capacity as happens in capacity markets, this central authority could auction reliability contracts as proposed by Vazquez *et al.* (2002). Such contracts consist essentially of long-term call options with a substantial penalty for nondelivery. The central authority uses reliability criteria to determine the total amount Q of contracts to be purchased and sets the strike price s of these contracts, typically at 25% above the variable cost of the most expensive generator that is expected to be called. It also sets the duration of the contracts. Bids for these contracts are ranked in terms of the premium fee asked by the generators. The premium fee P that clears the quantity Q is paid for all contracts.

Let us consider a generator that has sold an option for q MW at a premium P. This generator receives a premium fee of $P \cdot q$ for every period of the duration of the contract. For each period during which the spot price of electrical energy π exceeds the strike price s, this generator must reimburse $(\pi - s) \cdot q$ to the consumers. If this generator is only producing g MW during this period, it must pay an additional penalty of $pen \cdot (q - g)$.

Reliability contracts have a number of desirable features:

- They reduce the risks faced by marginal generators because the highly volatile and uncertain revenues derived from price spikes are replaced by a steady income from the option fees.
- The central authority can set the amount of contracts to be auctioned at a level that is likely to achieve the desired level of reliability.
- Generators have an incentive to maintain or increase the availability of their generating units because periods of high prices caused by shortages of capacity are less profitable. The penalty for nondelivery during periods of high prices discourages generators from bidding for contracts with less-reliable units.
- In exchange for the money they pay above the cost of electrical energy, consumers get a hedge against very high prices. This is in direct contrast with capacity payments and capacity markets where the benefit for consumers is not tangible. Consumers also get the reassurance that the option fees are determined through a competitive auction.
- Finally, because the strike price is set significantly above competitive prices, the options become active only when the system is close to rationing. Interferences with the normal energy market are thus minimized.

7.4 Further Reading

Sullivan *et al.* (2003) have written an easy-to-read introduction to the techniques used to make investment decisions. Stoft (2002) discusses in considerable detail the perceived advantages of electricity markets, where decisions to invest in generation capacity are driven entirely by prices for electrical energy. Schweppe *et al.* (1988) is the classic reference on spot pricing. Ford (1999, 2001) presents some very interesting simulations that explore the factors that might lead to boom-and-bust cycles in the building of generation capacity. de Vries and Hakvoort (2003) discuss the advantages and disadvantages of the various methods used to encourage investments in generation capacity. Vazquez *et al.* (2002) introduce the concept of reliability contracts. Billinton and Allan (1996) explain in great detail the relation between generation capacity and system reliability.

Billinton R, Allan R N, *Reliability Evaluation of Power Systems*, Second Edition, Kluwer Academic/Plenum Press, Boston, 1996.

de Vries L J, Hakvoort R A, The question of generation adequacy in liberalized electricity markets, *Proceedings of the 26th IAEE Annual Conference*, June 2003.

Ford A, Cycles in competitive electricity markets: a simulation study of the western United States, *Energy Policy*, **27**, 1999, 637–658.

Ford A, Waiting for the boom: a simulation study of power plant construction in California, *Energy Policy*, **29**, 2001, 847–869.

Schweppe F C, Caramanis M C, Tabors R D, Bohn R E, *Spot Pricing of Electricity*, Kluwer Academic Publishers, Boston, 1988.
Stoft S, *Power System Economics*, John Wiley & Sons, New York, 2002.
Sullivan W G, Wicks E M, Luxhoj J T, *Engineering Economy*, Twelfth Edition, Prentice Hall, Upper Saddle River, NJ, 2003.
Vazquez C, Rivier M, Perez-Arriaga I J, A Market Approach to Long-Term Security of Supply, *IEEE Transactions on Power Systems*, **17**(2), 2002, 349–357.

7.5 Problems

Most of these problems require the use of a spreadsheet.

7.1 Calculate the Internal Rate of Return (IRR) for an investment in a 400-MW power plant with an expected life of 30 years. This plant costs 1200 $/kW to build and has a heat rate of 9800 Btu/kWh. It burns a fuel that costs 1.10 $/MBtu. On average, it is expected to operate at a maximum capacity for 7446 h per year and sell its output at an average price of 31 $/MWh. What should be the average price of electrical energy if this investment is to achieve a MARR of 13%?

7.2 What would be the Internal Rate of Return of the unit of Problem 7.1 if the utilization rate drops by 15% after 10 years and by another 15% after 20 years?

7.3 What would be the Internal Rate of Return of the unit of Problem 7.1 if the price of electrical energy was 35 $/MWh during the first 10 years of the expected life of the plant before dropping to 31 $/MWh? What would be the value to the Internal Rate of Return if this price was 31 $/MWh during the first 20 years and 35 $/MWh during the last 10 years. Compare these results with the Internal Rate of Return calculated in Problem 7.1 and explain the differences.

7.4 In an effort to meet its obligation under the Kyoto agreement, the government of Syldavia has decided to encourage the construction of renewable generation by guaranteeing to buy their output at a fixed price of 35 $/MWh. Greener Syldavia Power Company is considering taking advantage of this program by building a 200-MW wind farm. This wind farm has an expected life of 30 years and its building cost amounts to 850 $/kW. On the basis of an analysis of the wind regime at the proposed location, the engineers of Greener Syldavia Power Company estimate that the output of the plant will be as shown in the table below:

Output as a fraction of capacity (%)	**Hours per year**
100	1700
75	1200
50	850
25	400
0	4610

Given that the Greener Syldavia Power Company has set itself a Minimum Acceptable Rate of Return of 12%, should it take the government's offer and build this wind farm?

7.5 Syldavia Energy is exploring the possibility of building a new 600-MW power plant. Given the parameters shown in the table below, which technology should it adopt for this plant, assuming that the plant would have a utilization factor of 0.80 and would be able to sell its output at an average price of 30 $/MWh? Syldavia Energy uses a Minimum Acceptable Rate of Return of 12%.

	Technology A	**Technology B**
Investment cost	1100 $/kW	650 $/kW
Expected plant life	30 years	30 years
Heat rate at rated output	7500 Btu/kWh	6500 Btu/kWh
Expected fuel cost	1.15 $/MBtu	2.75 $/MBtu

7.6 Borduria Power has built a plant with the following characteristics:

Investment cost	1000 $/kW
Capacity	400 MW
Expected plant life	30 years
Heat rate at rated output	9800 Btu/kWh
Expected fuel cost	1.10 $/MBtu
Expected utilization factor	0.85
Expected average selling price	31 $/MWh

After five years of operation, market conditions change dramatically. The fuel price increases to 1.50 $/MBtu, the utilization factor drops to 0.45 and the average price at which Borduria Power can sell the energy produced by this plant drops to 25 $/MWh.

What should Borduria Power do with this plant? What should Borduria Power have done if it had known about this change in market conditions? Assume that Borduria Power uses a MARR of 12% and ignore the recoverable cost of the plant.

7.7 Assume that Borduria Power decides to continue operating the plant of Problem 7.6 and that the market conditions do not improve. Five years later, the plant suffers a major breakdown that would cost $120 000 000 to repair. It is expected that this repair would allow the plant to continue operating for the rest of its design life. What should Borduria Power do? What should it do if this breakdown occurs 15 years after the plant was built?

7.8 An old 100-MW power plant has a heat rate of 13 000 Btu/kWh and burns a fuel that costs 2.90 $/MBtu. The owner of the plant estimates the fixed cost of keeping the plant available at $360 000 per year. What is the minimum price that would justify keeping this plant available if it has a 1% utilization rate? Compare this price with the average production cost of the plant.

7.9 The investment analysis illustrated by examples 7.1 and 7.2 is quite simplified. Discuss what factors should be considered in a more detailed analysis.

7.10 Plot the load-duration curves and the price-duration curve for the power system in the region where you live or another system for which you have access to the necessary data. Compare the peak demand to the installed generation capacity in the system.

7.11 Repeat the previous problem for several years. Try to explain any significant differences that you may observe in terms of weather conditions, commissioning of new generating plants, retirement of old plants and other relevant factors.

7.12 Determine if there is a mechanism to encourage the provision of generation capacity in the region where you live (or in another region for which you have access to sufficient information).

8

Investing in Transmission

8.1 Introduction

In Chapter 6, we studied the effect that an existing transmission network has on electricity markets. Expanding this transmission network through the construction of new lines or the upgrade of existing facilities increases the amount of power that can be traded securely and the number of generators and consumers that can take part in this market. Transmission expansion thus enhances the competitiveness of the market. On the other hand, investments in new transmission equipment are costly and should therefore be undertaken only if they can be justified economically. In order to deliver maximum economic welfare to society, the electricity supply industry should follow the path of least-cost long-term development. This requires a coordinated approach to the optimization of the generation and transmission operation and development. Optimizing the transmission network in isolation from the generation resources would almost certainly not meet the above objective. Before the introduction of competition, the vertical integration of electric utilities was considered necessary to ensure a sufficient level of coordination.

Among other reasons, competition was introduced in the electricity supply to respond to growing concerns about the inefficiency of established operation and investment practices. One of the consequences of the deregulation process has been the separation of generation from transmission. This separation is indeed frequently considered indispensable to achieve open and nondiscriminatory access to the energy market. In this environment, pricing of transmission becomes the key to achieving both efficient operation and least-cost system development of the entire system. Coordination of investments in generation and transmission, which are now operated as separate entities, is to be achieved through efficient network pricing mechanisms.

Much of the early work on transmission network pricing has focused on short-term operational efficiency and transmission congestion management. Location specific, short-run marginal cost (SRMC) based pricing is now a well-established method for allocating scarce network resources. Locational marginal prices computed from a bid-based security constrained dispatch and combined with financial transmission rights (FTRs) have been successfully implemented in several major electricity markets.

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

More recently, the discussion has turned to the need to restructure the framework for investments in the transmission network. This restructuring can be taken along two, possibly complementary, directions: merchant transmission investments and transmission investments based on regulatory incentives.

The first approach starts from the view that market forces are the keys to investment and expansion in transmission. The fundamental function of any transportation business is to buy a product at a low price in one location and sell it at a higher price in another. Transportation is a viable business if the price differential between the two markets is larger than the cost of transporting the product. In principle, the same logic could be applied to transmission of electricity. While locational marginal prices, coupled with FTRs, provide a conceptual framework for merchant transmission investments, a number of theoretical and practical difficulties must still be overcome.

The second approach starts from the premise that transmission networks is inherently a monopoly and hence needs to be regulated. The key responsibility of the regulatory agencies that determine the income of transmission developers is then to organize incentives that encourage an efficient transmission expansion. These incentives should financially reward decisions that increase economic efficiency. They should also penalize inefficient expenditures. Setting the targets that measure efficient operation is particularly difficult with this approach. Allocating the costs and benefits of transmission expansion to all the network users is another major challenge.

Since both of these approaches are still under development, this chapter does not discuss the implementation of either approach but focuses on the theoretical foundations common to both.

After a brief review of the essential characteristics of the transmission business, we discuss the traditional approach to transmission investments where investors are remunerated on the basis of the cost of the installed transmission equipment. We then consider the relationship between short-term locational marginal prices and transmission investment. We also develop the concept of economically adapted or reference transmission network.

8.2 The Nature of the Transmission Business

In liberalized electricity markets, transmission is usually separated from the other components of a traditional, vertically integrated utility. It is therefore useful to begin our discussion of transmission investments by considering some of the characteristics of transmission as a standalone business.

Rationale for a transmission business The transmission business exists only because the generators and loads that use the network are in the wrong place. Market opportunities for transmission increase with the distance that separates producers and consumers. On the other hand, if a reliable and environmentally friendly generation technology became cost effective for domestic installations, the transmission business would probably disappear.

Transmission is a natural monopoly It is currently almost inconceivable that a group of investors would decide to build a completely new transmission network designed to operate in competition with an existing one. Because of their visual impact on the environment, it is indeed most unlikely that the construction of competing transmission lines along similar routes would be allowed. Furthermore, the minimum efficient size of a transmission network is such that electricity transmission is considered a good example of a natural monopoly.

Like all monopolies that provide an essential service, electricity transmission must be regulated to ensure that it delivers an economically optimal combination of quality of service and price. This is not an objective that is easy to achieve. Even though the consumers and generators pay for using the transmission network, the regulator in essence "buys" transmission capacity on their behalf. Its best judgment about how much capacity is needed thus replaces the multitude of independent purchasing decisions that make up the demand curve in a competitive market.

In exchange for being granted a regional monopoly, a transmission company must accept that the regulatory authorities will determine its revenues. These revenues are usually set in such a way that investors get a relatively modest return on their capital. However, compared to other stock market investments, transmission companies are relatively safe because they do not face competition. In fact, the biggest risk that these companies face is the regulatory risk, that is, the risk that a change in regulatory principles or practices may decrease their allowed revenues.

Transmission is a capital-intensive business Transmitting electric power securely and efficiently over long distances requires large amounts of expensive equipment. While the most visible items of equipment are obviously the aerial transmission lines, the cost of the transformers, switchgear and reactive compensation devices is very high. Maintaining the security of the system while operating it close to its physical limits requires large amounts of protection and communication equipment as well as sophisticated control centers. The cost of these investments is high compared to the recurring cost of operating the system. Making good investment decisions is thus the most important aspect of running a transmission company.

Transmission assets have a long life Most transmission equipment is designed for an expected life ranging from 20 to 40 years or even longer. A lot of things can change over such a long period. Generation plants that were expected to produce the bulk of the demand for electrical energy may become prematurely obsolete because of changes in the cost of fuel or because of the emergence of a competing technology. At the same time, an uneven economic development may shift the geographical distribution of the demand. A transmission line that was built on the basis of erroneous forecasts may therefore be used at only a fraction of its rating.

Transmission investments are irreversible Once a transmission line has been built, it cannot be redeployed in another location where it could be used more profitably. Other types of transmission equipment may be easier to move, but the cost of doing so is often prohibitive. The resale value of installed assets is very low. Owners of transmission networks therefore have to live with the consequences of their investment decisions for a very long time. A large investment that is not used as much as was initially expected is called a *stranded investment*. Investors must therefore analyze what the performance of an asset might be under a wide variety of scenarios. In a regulated environment, they usually get some form of assurance that they will be able to recover the value of their investment even if it becomes stranded because of unforeseen changes in the demand for transmission.

Transmission investments are lumpy Manufacturers sell transmission equipment in only a small number of standardized voltage and MVA ratings. It is therefore often not possible to build a transmission facility whose rating exactly matches the need. While it is occasionally possible to upgrade facilities as demand increases, this standardization and the low resale value of the installed equipment often make this process impractical and economically difficult to justify. Investments in transmission facilities thus occur infrequently and in large blocks. Early in its life, the capacity of a facility tends to exceed the demand. Later on, it is likely to be utilized much more intensively, at least if the situation evolves as forecast.

Economies of scale Ideally, investments should be proportional to the capacity they provide. For transmission lines, this is clearly not the case. The cost of building the line itself is primarily proportional to its length because of the need to acquire the right of way, adapt the terrain and erect the towers. The rating of the line affects the cost only through the size of the conductors and the height that the towers must have to accommodate higher voltages. In addition, new substations must be built at both ends or existing ones must be expanded. This cost is significant and almost independent of the amount of active power that the line can transport. Because of these fixed costs, the average cost of transmitting electricity decreases with the amount transported. Transmission networks thus involve important economies of scale.

Merchant transmission While the overwhelming majority of transmission investments are still remunerated on a regulated basis, over the last several years a few transmission links have been built on a merchant basis. The regional regulated transmission company did not build these links. Instead, an unregulated company provided the capital needed for their construction. Rather than getting a modest but safe rate of return, these unregulated companies hope to obtain much larger revenues through the operation of these links. On the other hand, they carry the risk that these revenues may be insufficient to recover the cost of their investment.

8.3 Cost-based Transmission Expansion

Under the traditional regulatory compromise, regulated transmission companies collect enough revenues to cover the costs of their investment plus a rate of return sufficient to attract capitalists seeking a relatively safe investment. While this approach is conceptually simple, we need to explore two important questions:

- How much transmission capacity should be built?
- How should the cost of transmission be allocated among the users of the transmission network?

8.3.1 Setting the level of investment in transmission capacity

Under the traditional model, investments in transmission facilities are carried out according to a process that typically works as follows:

- Using demographic and economic projections, the transmission company forecasts the needs for transmission capacity.

- On the basis of this forecast, it prepares an expansion plan that it submits to the regulatory authorities.
- The regulatory authorities review this plan and decide which facilities may be built or upgraded.
- The transmission company builds these new facilities using the capital provided by its shareholders or its bondholders.
- Once the new facilities are commissioned, the transmission company begins recovering the cost of these investments through the charges that users of the network have to pay.

The price that consumers pay for electricity is clearly a function of the capacity of the transmission network. If the regulator allows the transmission company to build too much transmission capacity, the users pay more for capacity that is not used. On the other hand, if too little capacity is available, congestion in the network reduces trading opportunities, increases prices in some areas and depresses them in others.

In theory, the regulator should try to get it exactly right because too little capacity or too much capacity causes a loss of global welfare. Because of the inevitable uncertainty on the evolution of the demand and the generation, this is not easily achieved. In practice, one might argue that, from an economic perspective, it is better to err on the side of too much transmission capacity. Transmission indeed accounts for only about 10% of the total cost of electricity to consumers. While the cost of overinvesting is not small, the potential cost of underinvesting is much higher because even a small deficit of transmission capacity can have a very large effect on the price of electrical energy, which represents about 60% of the total cost to consumers.

On the other hand, remunerating transmission companies on a rate-of-return basis encourages them to overstate the need for transmission capacity because building more facilities increases the revenue that they are allowed to collect from the users of the network. Regulatory authorities rarely have the manpower and technical expertise required to evaluate expansion plans prepared by the transmission company.

In conclusion, remunerating transmission investments on the basis of their cost keeps the transmission company in business, which is usually in the best interest of all parties involved. This approach also ensures some predictability in the cost of transmitting electricity. On the other hand, it does not guarantee that the level of investment in transmission capacity is economically optimal.

8.3.2 Allocating the cost of transmission

Once the regulator has determined the revenue that the transmission company can collect to recover its investment, this embedded cost must be divided between the producers and consumers that use the transmission network. In the following paragraphs, we briefly discuss the principles of the main allocation methods that have been proposed. Readers interested in the details of these embedded cost methods should consult Marangon Lima (1996).

8.3.2.1 Postage stamp method

Under this method, all users must pay a "use of system charge" to gain access to the network of their local transmission company. This charge usually depends on the megawatt rating of the generating units for a producer or the peak demand for a consumer. It can also factor in the annual energy produced or consumed (in MWh). It may also be a function of the voltage level at which a user is connected to reflect the fact that a user connected directly to the transmission network does not make use of the subtransmission and distribution networks. However, like a postage stamp, this charge usually does not depend on where the energy is coming from or going to, as long as it is within the local system.

The charge that each user pays thus reflects the average usage of the entire network rather than the use of specific transmission facilities. Charges are adjusted proportionally to ensure that the transmission company recovers all the revenue that it is entitled to collect.

Because of its simplicity, this method is the most common charging mechanism for the utilization of the local transmission network. Its main disadvantage is that the charges paid by each user do not reflect the actual use that they make of the network or the value they derive from being connected. In many cases, some users cross-subsidize others. This is not economically desirable because it distorts competition. For example, generators connected close to the main load centers could argue that they should not pay the same charges as remote generators because the energy they produce does not need to transit through long and expensive transmission lines to reach the consumers.

Another problem with the postage stamp approach is that it only covers the cost of using the local transmission network. If a producer wants to sell energy in a neighboring system, it may have to buy an additional postage stamp to get access to the neighboring network. If two trading partners are not located in adjacent networks, each intermediate transmission company may require the purchase of a separate postage stamp. Like in a stack of pancakes, the cost of each stamp may not be very high, but the overall expense may be substantial. This phenomenon is dubbed the "pancaking of rates". It is usually undesirable because the overall charge overestimates the actual cost of transmitting the energy and may make economically justified transactions unprofitable.

8.3.2.2 Contract path method

The contract path method finds its origins in the days when the electricity supply industry consisted mostly of vertically integrated utilities and energy transactions were infrequent. When a consumer wanted to buy energy from a producer other than its local utility, it was still making use of the network of this utility and should therefore bear a proportionate share of the embedded costs of this network. A wheeling contract has to be set up to formalize this arrangement. In this method, the contract specifies an electrically continuous path (the contract path) along which the power is assumed to flow from the generator to the point of delivery. The producer and the consumer agree to pay, for the duration of the contract, a wheeling charge proportional to the amount of power transmitted. This wheeling charge provides part of the revenue that the utility needs to recover the cost of the transmission assets included in the contract path.

The producer and consumer thus pay only for the usage of specific network facilities and not a fraction of the average cost of the entire network. While remaining simple, this method thus appears more cost reflective than the postage stamp approach. In reality, the power traded does not follow only the contract path but a multitude of other paths, as dictated by Kirchhoff's laws. Whether it is truly more cost reflective is thus highly questionable.

8.3.2.3 MW-mile method

With the MW-mile method, power flow calculations are used to determine the actual paths that the power follows through the network. The amount of MW-mile of flow that each transaction causes is calculated. This amount is then multiplied by an agreed per-unit cost of transmission capacity to get the wheeling charge. The method can be refined to take into account the fact that some transactions reduce the flow on some lines. If transmission networks were linear systems, this approach would be rigorous. Unfortunately, they are not. The base case from which transactions are evaluated and the order in which they are considered has a significant and undesirable effect on the results.

8.3.2.4 Comments on the embedded cost allocation methods

All the methods discussed above have been criticized because they lack a credible foundation in economic theory. In particular, they produce charges that are proportional to the average rather than the incremental cost of the network. This means that they do not provide correct economic signals. Nevertheless, because of their simplicity and ease of implementation, they have been used extensively, mainly in the United States.

8.4 Value-based Transmission Expansion

In a competitive electricity market, transmission can be viewed as being in competition with generation. The transmission network indeed allows remote generators to compete with local ones. We can thus estimate the value of transmission on the basis of the differences in the marginal cost or price of generation across the network. This value provides a sound basis for setting the price that producers and consumers should pay to use the network.

8.4.1 Quantifying the value of transmission

8.4.1.1 A first example

Let us consider the two-bus, one-line system shown in Figure 8.1. For the sake of simplicity, we neglect losses and ignore security considerations. We also assume that the capacity of both generators is such that each of them can supply the 1000 MW load

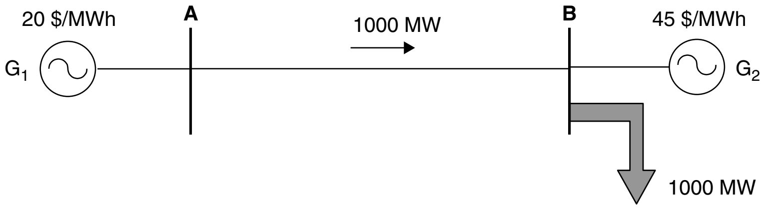


Figure 8.1 Simple example illustrating the value of transmission

on its own. Finally, we assume that the capacity of the transmission line is sufficient to support any power transmission that may be required.

The consumers at bus B can either buy energy at 45 $/MWh from the local generator G_2 or buy energy at 20 $/MWh from the remote generator G_1 and pay for the transmission of this energy. If the cost of this transmission is less than 25 $/MWh, consumers will choose to buy their energy from generator G_1 because the overall cost would be less than the 45 $/MWh that they would have to pay to buy energy from generator G_2.

It is thus not in the best interest of the owner of the transmission line to charge more than 25 $/MWh because such a charge would discourage consumers from making use of the transmission system. In this example, the value of the transmission service is thus 25 $/MWh because, at that price, consumers are indifferent between using and not using transmission. The value of transmission is thus a function of the short-run marginal cost of generation. In this case, this function is very simple because there is no limit to the substitution between transmission and local generation.

We can also look at the problem from an investment perspective. This transmission line should be built only if its amortized cost amounts to less than 25 $/MWh.

If the maximum output of the local generators is less than 1000 MW, the transmission line must be used to supply the load. The value of transmission is then no longer determined by the price of local generation but by the consumers' willingness to pay for electrical energy. In the short term, this could be significantly higher than 25 $/MWh. Limitations on local generation place the transmission provider in a monopoly position because the consumers have a choice between using transmission and giving up consumption. This monopoly position may not be sustainable in the long run because it would encourage the development of local generation.

8.4.1.2 A second example

We will now revisit the Borduria/Syldavia example that we introduced in Chapter 6. In that chapter, we studied the effect that the operation of this interconnection has on market prices. We will now determine the optimal capacity of this interconnection.

Our model for this interconnected system is the same as the one we used in Chapter 6 and is shown in Figure 8.2. The only difference is that the capacity of the interconnection is not fixed. Our starting point is the economic characteristics of the two markets when they operate independently. We saw that the supply functions for the electricity markets in Borduria and Syldavia are respectively

$$\pi_B = MC_B = 10 + 0.01 P_B \text{ \$/MWh} \tag{8.1}$$

$$\pi_S = MC_S = 13 + 0.02 P_S \text{ \$/MWh} \tag{8.2}$$

Figure 8.2 Model of the Borduria/Syldavia interconnection

The demands in Borduria and Syldavia are respectively 500 MW and 1500 MW. We continue to assume that these demands do not vary with time and are perfectly inelastic.

In the absence of an interconnection, the two national electricity markets operate independently and the prices in Borduria and Syldavia are respectively 15 $/MWh and 43 $/MWh. The value of transporting the first megawatt-hour from Borduria to Syldavia is thus equal to the difference in price between the two countries, that is, 28 $/MWh.

We saw in Chapter 6 that when the flow through the interconnection is 400 MW, generators in Borduria produce 900 MW. 500 MW of this production is for the local load, while the remaining 400 MW is sold to consumers in Syldavia. The remaining 1100 MW of Syldavian load is produced locally. Under these conditions, the prices in Borduria and Syldavia are 19 $/MWh and 35 $/MWh respectively. The value of transporting one additional megawatt-hour from Borduria to Syldavia is thus only 16 $/MWh. This is also the maximum price that consumers in Syldavia would agree to pay for the transport of a megawatt-hour that they have bought in Borduria for 19 $/MWh. If the price of transmission were any higher, they would prefer to buy this megawatt-hour from local generators.

When the flow on the interconnection reaches 933.3 MW, the prices in Borduria and Syldavia are equal:

$$\pi = \pi_B = \pi_S = 24.30 \text{ \$/MWh} \tag{8.3}$$

At that point, the marginal value of transmission is zero because Syldavian consumers can buy one extra megawatt-hour from local generators at the same price that they would pay for a megawatt-hour purchased on the Bordurian market. They would therefore not be willing to pay anything for the transmission of this incremental energy. There is also no reason to increase the power transfer between the two countries any further because that would make the marginal value of transmission negative. Transmitting more power would require an increase in production in Borduria and would make the price of energy on that market higher than the price in Syldavia. The interconnection would then be transmitting energy from a higher-priced location to a lower-priced location. This would obviously be wasteful and economically inefficient. We can thus conclude that the marginal value of transmission is a function of the magnitude of the flow, which in turn depends on the energy prices and the capacity of the transmission network.

8.4.2 The transmission demand function

We will now formalize the observations that we made in the examples above by introducing a demand function for transmission. This function gives the value of transmission in terms of the amount of power F transmitted between Borduria and Syldavia.

$$\pi_T(F) = \pi_S(F) - \pi_B(F) \tag{8.4}$$

where $\pi_T(F)$ is the value of the transmission. The prices of electrical energy in Syldavia and Borduria, $\pi_S(F)$ and $\pi_B(F)$ are expressed in terms of the power transmitted. Substituting (8.1) and (8.2) into (8.4) we obtain

$$\begin{aligned}\pi_T(F) &= (13 + 0.02 P_S(F)) - (10 + 0.01 P_B(F)) \\ &= 3 + 0.02 P_S(F) - 0.01 P_B(F)\end{aligned} \tag{8.5}$$

The production of the generators in Borduria and Syldavia can be expressed in terms of the flow on the interconnection and the local demands as follows:

$$P_B(F) = D_B + F \tag{8.6}$$

$$P_S(F) = D_S - F \tag{8.7}$$

Equation (8.5) thus becomes

$$\pi_T(F) = 3 + 0.02(D_S - F) - 0.01(F + D_B) \tag{8.8}$$

Substituting the known values for the demands, we get

$$\pi_T(F) = 28 - 0.03F \tag{8.9}$$

Using this expression, we can check the results that we obtained above in an ad hoc manner. In particular, we see that when the flow is equal to zero, the price of transmission is 28 \$/MWh. Conversely, the transmission price drops to zero when the flow reaches 933.3 MW, which is the value of the flow for which the prices of generation in Borduria and Syldavia are equal.

Equation (8.9) can be inverted to get the demand for transmission as a function of its price:

$$F(\pi_T) = 933.3 - 33.3\pi_T \tag{8.10}$$

As Figure 8.3 shows, and as one would expect from any demand function, the demand for transmission increases when the price decreases.

It is interesting to examine the revenue that the owner of the transmission line would receive as a function of the capacity that is made available. This revenue is given by the following expression:

$$R(F) = \pi_T \cdot F = (28 - 0.03F) \cdot F \tag{8.11}$$

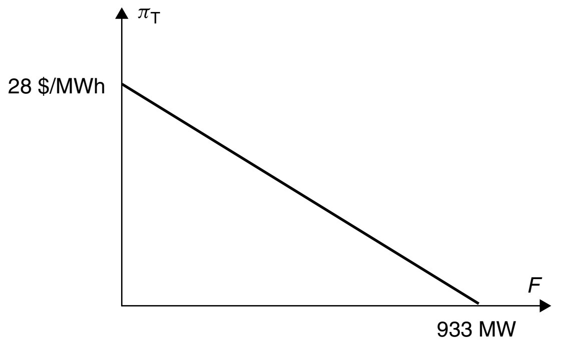


Figure 8.3 Transmission demand function for the interconnection between Borduria and Syldavia

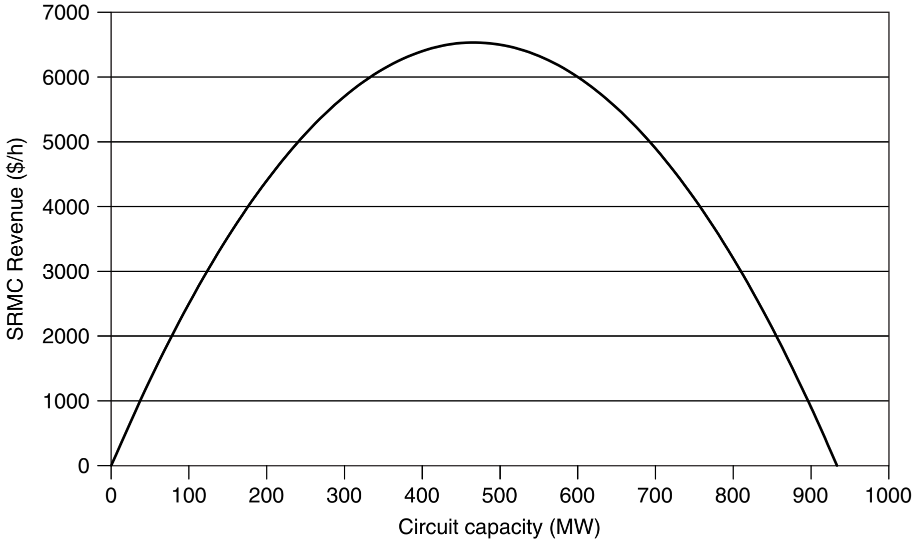


Figure 8.4 Variation of the transmission revenue as a function of the available capacity for Borduria–Syldavia

As illustrated in Figure 8.4, the revenue is thus a quadratic function of the amount of power transmitted. If no capacity is made available, the transmission owner obviously collects no revenue as no power is being transmitted. On the other hand, for a transmission capacity of 933 MW, the flow through the circuit is maximum and the nodal prices at both ends of the line are identical. We thus have $\pi_T = 0$ and the revenue is also zero. The revenue is maximum for a transmission capacity of 466 MW.

8.4.3 The transmission supply function

Let us now look at the other side of the "market" for transmission and construct a supply function for transmission. The annuitized cost of building a transmission line consists of a variable cost component, which does depend on the capacity of the line T, and a fixed-cost component, which does not depend on this capacity:

$$C_{\mathrm{T}}(T) = C_{\mathrm{F}} + C_{\mathrm{V}}(T) \tag{8.12}$$

For the sake of simplicity, we will assume that the variable component is a linear function of the capacity:

$$C_{\mathrm{V}}(T) = k \cdot l \cdot T \tag{8.13}$$

If l is the length of the line in kilometers, k is the annuitized marginal cost of building 1 km of transmission line and its dimensions are \$/(MW · km · year). The annuitized marginal cost of transmission capacity is thus:

$$\frac{\mathrm{d}C_{\mathrm{T}}}{\mathrm{d}T} = k \cdot l \tag{8.14}$$

This quantity is called the long-run marginal costs (LRMC) because it relates to the cost of investments in transmission. Dividing it by the number of hours in a year ($\tau_0 = 8760\,\mathrm{h}$), we get the hourly long-run marginal cost, which, as we need for the transmission supply function, is expressed in \$/MWh:

$$c_{\mathrm{T}}(T) = \frac{k \cdot l}{\tau_o} \tag{8.15}$$

Because of the simplifying assumptions that we made in Equation (8.13), the marginal cost of transmission is a constant that does not depend on the capacity of the line.

If we assume, for this example, that the line is 1000 km long and that

$$k = 35 \text{ \$/(MW} \cdot \text{km} \cdot \text{year)} \tag{8.16}$$

The hourly LRMC of transmission is then

$$c_{\mathrm{T}} = 4.00 \text{ \$/MWh} \tag{8.17}$$

8.4.4 Optimal transmission capacity

The optimal transmission capacity is such that the supply and the demand for transmission are in equilibrium. The price that transmission users are willing to pay should therefore be equal to the marginal cost of providing this capacity. In this case, we must thus have

$$\pi_{\mathrm{T}} = c_{\mathrm{T}} = 4.00 \text{ \$/MWh} \tag{8.18}$$

Combining Equations (8.10) and (8.18), we get the optimal capacity:

$$T^{\mathrm{OPT}} = 800 \text{ MW} \tag{8.19}$$

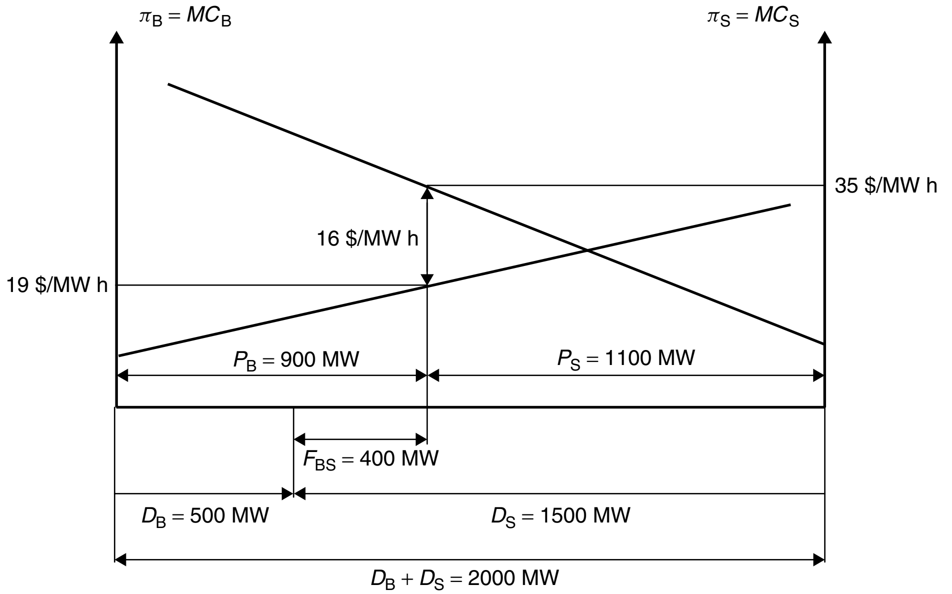


Figure 8.5 Relation between the capacity of the interconnection and the difference in nodal prices between Borduria and Syldavia

Figure 8.5, (which is identical to Figure 6.10) illustrates this optimization. It shows the nodal prices in Borduria and Syldavia as a function of the production in each country. Since we assume that the demands are constant, it also shows these nodal prices as a function of the flow on the interconnection. If this flow is limited by the transmission capacity of the interconnection, the vertical distance separating the two curves gives the difference in nodal prices that arises between the two markets. We could call this difference the SRMC of not having more transmission capacity. If this interconnection has a transmission capacity of 800 MW, the flow from Borduria to Syldavia is equal to 800 MW ($F = T$). The SRMC is then 4.00 $/MWh. This means that the SRMC is exactly equal to the LRMC of the interconnection. If the owner of the interconnection collected the difference in nodal prices between the two markets (or charged a transmission price equal to this difference), it would collect exactly enough revenue to pay for the construction of the line.

If the transmission capacity is larger than 800 MW, the operating point would move to the right in Figure 8.5 and the nodal price difference (SRMC) would be lower. Since the LRMC is constant, the value of the interconnection would be less than its cost. If the revenues of the transmission owners were proportional to the nodal price difference, they would not collect enough revenue to cover their investment costs. In other words, they would have overinvested.

On the other hand, if the transmission capacity were less than 800 MW, the operating point would move to the left in Figure 8.5. The difference in nodal prices would then be larger than the LRMC. This underinvestment is good for the owners of the interconnection because they can charge a higher price for the use of the transmission line. From a global perspective, this underinvestment is not good because it limits trading opportunities to a suboptimal level.

8.4.5 Balancing the cost of constraints and the cost of investments

From the expressions for the marginal costs of generation in Borduria and Syldavia given in Equations (8.1) and (8.2) respectively, we can deduce the variable generation costs in both countries:

$$C_B = 10P_B + \frac{1}{2}0.01P_B^2 \text{ \$/h} \tag{8.20}$$

$$C_S = 13P_S + \frac{1}{2}0.02P_S^2 \text{ \$/h} \tag{8.21}$$

In Chapter 6, we determined that the productions that minimize the total generation cost when operation is not constrained by the transmission network are

$$P_B = 1433.3 \text{ MW} \tag{8.22}$$

$$P_S = 566.7 \text{ MW} \tag{8.23}$$

The unconstrained flow in the interconnection is then

$$F = 933.33 \text{ MW} \tag{8.24}$$

The corresponding generation costs in each country and in the whole system are

$$C_B = 24\,605 \text{ \$/h} \tag{8.25}$$

$$C_S = 10\,578 \text{ \$/h} \tag{8.26}$$

$$C^U = C_B + C_S = 35\,183 \text{ \$/h} \tag{8.27}$$

This unconstrained dispatch and the associated costs are often called respectively the *merit order dispatch* and *the merit order costs*.

If the transmission capacity (and hence the flow on the interconnection) is 800 MW, the generations and the corresponding costs are

$$P_B = 1300 \text{ MW}, C_B = 21\,450 \text{ \$/h} \tag{8.28}$$

$$P_S = 700 \text{ MW}, C_S = 14\,000 \text{ \$/h} \tag{8.29}$$

The total cost of supplying the load for this constrained condition is

$$C^C = 35\,450 \text{ \$/h} \tag{8.30}$$

The difference in cost between the constrained and unconstrained conditions is called the *cost of constraints* or the *out-of-merit generation cost*:

$$\Delta C = C^C - C^U = 267 \text{ \$/h} \tag{8.31}$$

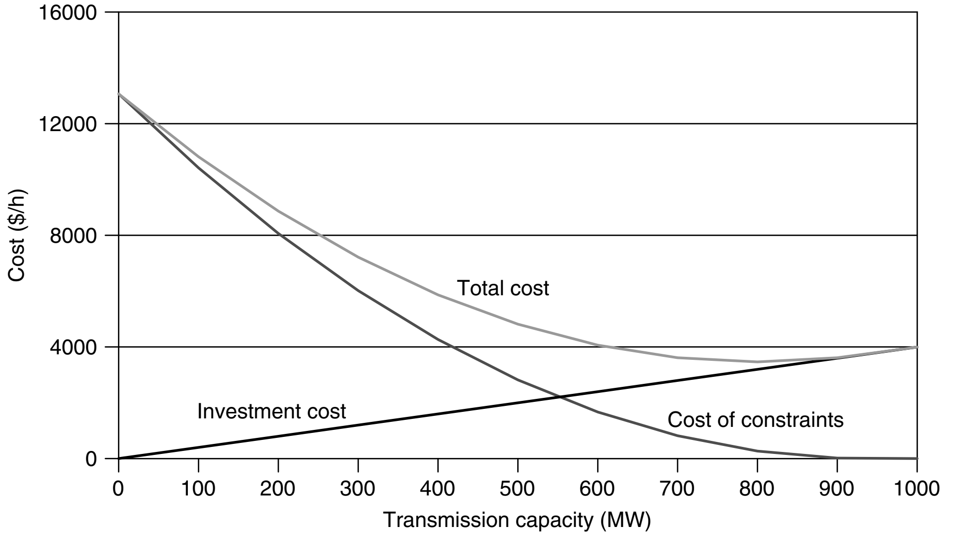


Figure 8.6 Evolution of the cost of constraints, the investment cost and the total transmission cost for the Borduria–Syldavia interconnection

The total cost of transmission is the sum of the cost of building the transmission system and the cost of constraints. As Figure 8.6 shows, the cost of building the transmission system increases with the transmission capacity while the cost of constraints decreases because the transmission network puts fewer limitations on the generation dispatch. Minimizing the total cost of transmission is thus the objective of the network development task. Figure 8.6 shows that this optimum is achieved for a transmission capacity of 800 MW. Obviously, this is consistent with the result that we obtained in Equation (8.19).

8.4.6 Effect of load fluctuations

So far, we have made the very convenient assumption that the load remains constant over time. This is obviously not realistic and we must analyze the effect that the natural fluctuations of the load with the cycle of human activities have on the value of transmission.

8.4.6.1 Load-duration curve

If we assume that the load fluctuations in the whole system follow similar patterns, we do not need to concern ourselves with the time at which the load achieves a particular value. What is important is the duration of each load level. Chronological load profiles, such as the one shown in Figure 8.7(a), show how the load varies over the course of a day. This period is divided into a number of intervals during which the load is assumed to be constant. In this case, the day has been divided into eight intervals of three hours labeled a to h. On the graph shown in Figure 8.7(b), these intervals have been sorted

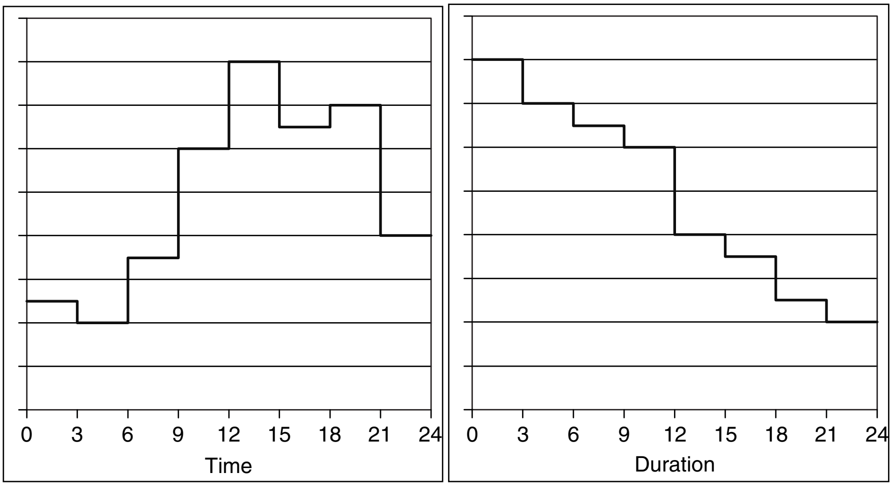


Figure 8.7 (a) Chronological load profile and (b) load-duration curve

in decreasing order of load. This graph thus shows the number of hours over the course of a day during which the load exceeded a certain value. This process could be applied over a longer period (e.g. a year) and with shorter intervals (e.g. one hour). The resulting load-duration curve then shows the number of hours over the course of a year during which the load exceeded a certain value. We have already encountered such a curve in Chapter 7.

Since handling a load-duration curve with up to 8760 hourly intervals is not practical, some aggregation is usually performed. For example, Figure 8.8 shows how the load-duration curve of Figure 8.7 has been simplified by grouping the values of the load into four groups.

8.4.6.2 A third example

Let us modify our Borduria–Syldavia example and consider the even greater simplification shown in Figure 8.9, where the load has been divided into a peak level and an off-peak level. The peak period has a duration of 3889 h while the off-peak period lasts 4871 h. For the sake of simplicity, we have assumed that the on-peak and off-peak periods are coincident in the two countries.

As discussed above, in order to determine the optimal transmission capacity, we must balance the annual saving in energy costs against the annuitized cost of transmission. While we could carry out this calculation analytically, we will instead compute the components of the cost for various values of the transmission capacity and find which one minimizes the total cost.

To calculate the hourly cost of constraints, we need to know the unconstrained cost of generation. Table 8.1 shows the unconstrained economic dispatch for the peak and the off-peak load and the corresponding generation costs as calculated using Equations (8.20) and (8.21). Tables 8.2 and 8.3 show the hourly generation costs for the off-peak and on-peak periods respectively for a range of values of the interconnection capacity.

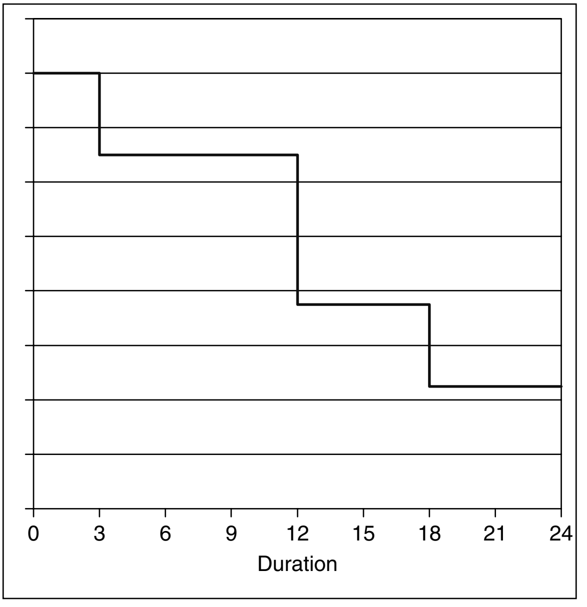


Figure 8.8 Simplified load-duration curve

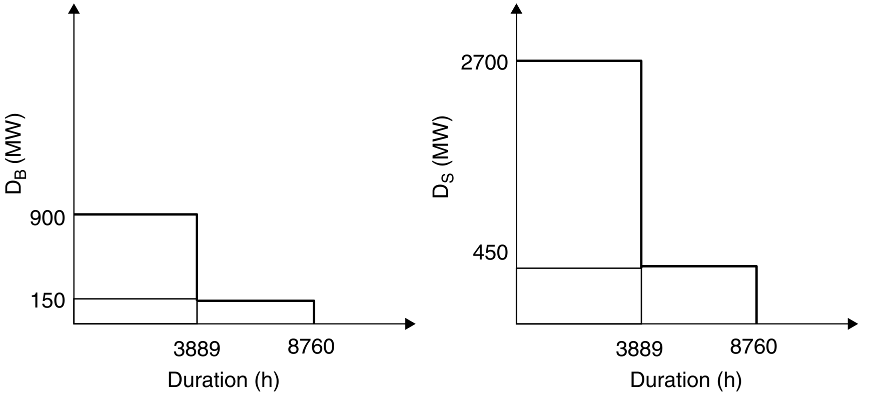


Figure 8.9 Load-duration curves for (a) Borduria and (b) Syldavia

Table 8.1 Unconstrained economic dispatch for the peak and off-peak load conditions in the Borduria–Syldavia system

Load (MW)	Generation in Borduria (MW)	Generation in Syldavia (MW)	Total hourly generation cost ($/h)
600	500	100	7650
3600	2500	1100	82 650

Table 8.2 Hourly generations, total hourly generation cost and hourly constraint cost for the Borduria–Syldavia system during off-peak loading conditions

Interconnection capacity (MW)	Generation in Borduria (MW)	Generation in Syldavia (MW)	Total hourly generation cost ($/h)	Hourly constraint cost ($/h)
0	150	450	9488	1838
100	250	350	8588	938
200	350	250	7988	338
300	450	150	7688	38
350	500	100	7650	0
400	500	100	7650	0
450	500	100	7650	0
500	500	100	7650	0
600	500	100	7650	0
700	500	100	7650	0
800	500	100	7650	0
900	500	100	7650	0

Table 8.3 Hourly generations, total hourly generation cost and hourly constraint cost for the Borduria–Syldavia system during on-peak loading conditions

Interconnection capacity (MW)	Generation in Borduria (MW)	Generation in Syldavia (MW)	Total hourly generation cost ($/h)	Hourly constraint cost ($/h)
0	900	2700	121 050	38 400
100	1000	2600	116 400	33 750
200	1100	2500	112 050	29 400
300	1200	2400	108 000	25 350
350	1250	2350	106 088	23 438
400	1300	2300	104 250	21 600
450	1350	2250	102 488	19 838
500	1400	2200	100 800	18 150
600	1500	2100	97 650	15 000
700	1600	2000	94 800	12 150
800	1700	1900	92 250	9 600
900	1800	1800	90 000	7 350

Given that the durations of the off-peak and on-peak periods are 4871 h and 3889 h respectively, we can compute the total annual cost of constraints for the values of the interconnection capacity shown in the previous two tables. In this example, we will change the annuitized marginal cost of transmission investment from 35 $/(MW · km · year) to 140 $/(MW · km · year). Table 8.4 shows these values together with the annuitized cost of transmission investments and their sum, which is the total annual transmission cost. We only consider the variable part of the cost of transmission investments and calculate it using Equation (8.14). We observed that a transmission capacity of 400 MW is optimum because it minimizes the total cost of transmission.

Table 8.4 Annual cost of constraints, annuitized cost of transmission investments and total annual cost of transmission as a function of the transmission capacity of the Borduria–Syldavia interconnection

Interconnection capacity (MW)	Annual constraint cost (k$/year)	Annuitized investment cost (k$/year)	Total annual transmission cost (k$/year)
0	158 304	0	158 304
100	135 835	14 000	149 835
200	115 993	28 000	143 993
300	98 780	42 000	140 780
350	91 159	49 000	140 159
400	**84 012**	**56 000**	**140 012**
450	77 157	63 000	140 157
500	70 593	70 000	140 593
600	58 342	84 000	142 342
700	47 257	98 000	145 257
800	37 339	112 000	149 339
900	28 587	126 000	154 587

8.4.6.3 Recovery of variable transmission investment costs

Let us now examine the effect that a transmission capacity of 400 MW has on the markets for electrical energy in Borduria and Syldavia.

During off-peak loading conditions, the capacity of the interconnection does not limit the power flow between the two countries. The two markets thus operate as a single market. Generators in Borduria and Syldavia produce 500 and 100 MW respectively. Since there is only 150 MW of load in Borduria, 350 MW flows on the interconnection to Syldavia. The marginal generation costs and hence the prices in Borduria and Syldavia are identical at 15.00 $/MWh. Therefore, during off-peak conditions, the short-run marginal value of transmission is zero. The congestion surplus or transmission revenue is thus also zero.

During peak loading conditions, Bordurian generators produce only 1300 MW because the local load is 900 MW and the transmission capacity is limited to 400 MW. The generators in Syldavia produce 2300 MW. Because of the transmission congestion, prices in the Bordurian and Syldavian markets are set by the local marginal cost of generation at 23.00 $/MWh and 59.00 $/MWh respectively. The short-run value of transmission is thus 36.00 $/MWh. During peak loading condition, the hourly congestion surplus is

$$CS_{\text{hourly}} = 400 \cdot 36 = 14\,400 \text{ \$/h} \tag{8.32}$$

Multiplying by the number of on-peak hours, we get the annual congestion surplus:

$$CS_{\text{annual}} = 14\,400 \cdot 3889 = 56\,000\,000 \text{ \$/year} \tag{8.33}$$

This amount is equal to the annuitized cost of transmission investment:

$$C_{\text{V}}(T) = k \cdot l \cdot T = 140 \cdot 1000 \cdot 400 = 56\,000\,000 \text{ \$/year} \tag{8.34}$$

For the optimal transmission capacity, the revenue earned from the congestion surplus thus covers exactly the variable part of the investment cost. However, it does not cover the fixed part of the transmission investment. Furthermore, this equality holds because we have assumed a constant value for the marginal cost of transmission capacity k. It does not hold if this marginal cost is not constant because of economies of scale.

8.4.7 Revenue recovery for suboptimal transmission capacity

In practice, the actual transmission capacity rarely coincides with its optimal value. The reasons for this discrepancy are easy to understand if we consider the uncertainties that affect the forecasts of demand and prices, the lumpiness of investments in transmission capacity and the legacy of historical investment decisions. Obviously, power system operators run the system on the basis of what the transmission capacity actually is, and not on the basis of what an optimization program says it should be. Since the nodal energy prices and the congestion surplus are determined by the actual network, it is important to study how suboptimality affects revenue recovery.

Let us consider our second example where the optimal capacity of the interconnection between Borduria and Syldavia is 800 MW and calculate what the revenue and cost would be if the transmission line was built with a capacity of 900 MW. Assuming that this capacity is made available, the flow is 900 MW. Bordurian generators increase their output to 1400 MW, while production in Syldavia drops to 600 MW. Using Equations (8.1) and (8.2), we find that energy prices in Borduria and Syldavia are 24.00 \$/MWh and 25.00 \$/MWh respectively. The short-run value of transmission has dropped from 4.00 \$/MWh for a capacity of 800 MW to 1.00 \$/MWh for a capacity of 900 MW.

The hourly and annual congestion surpluses collected are

$$CS_{\text{hourly}} = 900 \cdot 1 = 900 \text{ \$/h} \tag{8.35}$$

$$CS_{\text{annual}} = 900 \cdot 8760 = 7\,884\,000 \text{ \$/year} \tag{8.36}$$

On the other hand, the annuitized investment cost amounts to

$$C_{\text{V}}(T) = k \cdot l \cdot T = 35 \cdot 1000 \cdot 900 = 31\,500\,000 \text{ \$/year} \tag{8.37}$$

The revenue generated by the congestion surplus is smaller than it was for the optimal transmission capacity, and it is not sufficient to cover the cost of this overinvested transmission system.

Let us now examine the case of underinvestment. If the transmission capacity is only 700 MW, the flow on the interconnection is limited to this value. Generators in Borduria produce only 1200 MW (500 MW of local load and 700 MW transmitted to Syldavia) at a price of 22.00 \$/MWh. Syldavian producers generate 800 MW at a price of 29.00 \$/MWh to satisfy the remainder of the 1500 MW Syldavian load. This 7.00 \$/MWh price differential creates an hourly congestion surplus of

$$CS_{\text{hourly}} = 700 \cdot 7 = 4900 \text{ \$/h} \tag{8.38}$$

Over one year, this will generate a revenue of

$$CS_{\text{annual}} = 4900 \cdot 8760 = 42\,924\,000 \text{ \$/year} \tag{8.39}$$

On the other hand, the annuitized cost of investment for a 700-MW interconnection is

$$C_V(T) = k \cdot l \cdot T = 35 \cdot 1000 \cdot 700 = 24\,500\,000 \text{ \$/year} \tag{8.40}$$

In this case, the income generated by short-run marginal pricing of transmission is thus larger than the cost of building the transmission line. In other words, keeping the transmission capacity below the optimal value increases the revenue collected.

Let us now consider our third example where the interconnection between Borduria and Syldavia has a transmission capacity of 500 MW. During off-peak periods, this overinvestment has no effect because then even the optimal capacity does not constrain the power flow. The short-run marginal value of transmission and the transmission revenue remain at zero. On the other hand, during peak periods, the system operator makes use of all the 500 MW capacity of the interconnection. Bordurian generators can then produce 1400 MW while production in Syldavia is only 2200 MW. Equations (8.1) and (8.2) show that the energy prices in Borduria and Syldavia are 24.00 \$/MWh and 57.00 \$/MWh respectively. The short-run value of transmission is thus 33.00 \$/MWh, instead of 36.00 \$/MWh for a 400 MW transmission capacity.

The congestion surplus collected during hours of peak load is

$$CS_{\text{hourly}} = 500 \cdot 33 = 16\,500 \text{ \$/h} \tag{8.41}$$

Given the duration of the on-peak period, the annual congestion surplus is

$$CS_{\text{annual}} = 16\,500 \cdot 3889 = 64\,168\,500 \text{ \$/year} \tag{8.42}$$

On the other hand, the annuitized investment cost amounts to

$$C_V(T) = k \cdot l \cdot T = 140 \cdot 1000 \cdot 500 = 70\,000\,000 \text{ \$/year} \tag{8.43}$$

The revenue generated by the congestion surplus is larger than it was for the optimal transmission capacity, but is not sufficient to cover the cost of the overinvested transmission system.

Let us now turn our attention to the case of underinvestment. If the transmission capacity is only 300 MW, the flow on the interconnection is limited not only during peak load conditions, but also during the off-peak period.

During the off-peak periods, generators in Borduria produce 450 MW (150 MW of local load and 300 MW transmitted to Syldavia) at a price of 14.50 \$/MWh. Syldavian producers generate 150 MW at a price of 16.00 \$/MWh to satisfy the remainder of the 450 MW Syldavian load. This 1.50 \$/MWh price differential creates a congestion surplus of

$$CS_{\text{hourly}} = 300 \cdot 1.50 = 450 \text{ \$/h} \tag{8.44}$$

Over the 4871 off-peak hours, \$2 191 950 of congestion revenue is thus collected.

During peak load conditions, Bordurian generators produce 1200 MW, out of which 300 MW are transmitted through the interconnection, leaving Syldavian generators to produce 2400 MW. The marginal prices in Borduria and Syldavia are therefore 22.00 \$/MWh and 51.00 \$/MWh respectively. The hourly congestion surplus thus amounts to

$$CS_{\text{hourly}} = 300 \cdot (61.00 - 22.00) = 11\,700 \text{ \$/h} \tag{8.45}$$

Given that peak load conditions span 3889 h, \$45 501 300 is generated in congestion surplus. Considering both the off-peak and on-peak periods, the annual congestion revenue reaches \$47 693 250. On the other hand, the annuitized cost of investment for a 300 MW interconnection is

$$C_{\text{V}}(T) = k \cdot l \cdot T = 140 \cdot 1000 \cdot 300 = 42\,000\,000 \text{ \$/year} \tag{8.46}$$

In this case, the income generated by short-run marginal pricing of transmission is thus larger than the cost of building the transmission network. In other words, keeping the transmission capacity below the optimal value increases the revenue collected because congestion is more frequent.

8.4.8 Effect of economies of scale

In our discussion, we have so far assumed that the cost of investments in transmission equipment is proportional to the power transmitted. We have not taken into account the fact that a significant part of this cost might be fixed, that is, independent of the transmission capacity. Let us remove this simplifying assumption and reconsider the interconnection between Borduria and Syldavia, taking into account the component C_{F} of the total cost C_{T} of building the line:

$$C_{\text{T}}(T) = C_{\text{F}} + C_{\text{V}}(T) \tag{8.47}$$

Once we have decided to proceed with a transmission expansion project, the magnitude of the fixed cost has no influence at all on the capacity of the circuit to be built. At first glance, this may appear counter-intuitive. Consider, however, that once we have made the decision to undertake the project, we are committed to pay the fixed cost. Once this cost is paid, it has no influence on subsequent decisions, such as the capacity of the circuit.

To examine the impact of fixed costs, let us consider our second example. Assume that the fixed-cost component of the line amounts to 20 000 \$/km/year and this cost needs to be added to the total investment cost for the 1000-km long Borduria–Syldavia interconnection. This fixed component simply shifts the total cost curve upward and does not affect the location of its minimum. Let us assume that the interconnection has been built with the optimal capacity and that all this capacity is made available. As we saw in the previous section, the pattern of nodal prices is then such that the revenue derived from the price differentials covers exactly the variable part of the cost of building the transmission line. On the other hand, congestion revenues do not cover the fixed component of the cost of building the interconnection.

One way of recovering this loss in revenue would be to restrict the capacity that is made available, as suggested by Hogan (1999). Let us examine the short-run transmission revenue if, instead of offering the full 800 MW capacity of the line, its owner makes available only 650 MW to the system operator. The flow between Borduria and Syldavia is then 650 MW. Bordurian generators reduce their output to 1150 MW, while production in Syldavia increases to 850 MW. Using Equations (8.1) and (8.2), we find that energy prices in Borduria and Syldavia are 21.00 \$/MWh and 30.00 \$/MWh respectively. The short-run value of transmission increases from 4.00 \$/MWh to 8.50 \$/MWh.

The hourly and annual congestion surpluses are

$$CS_{\text{hourly}} = 650 \cdot 8.5 = 5525 \text{ \$/h} \tag{8.48}$$

$$CS_{\text{annual}} = 5525 \cdot 8760 = 48\,399\,000 \text{ \$/year} \tag{8.49}$$

On the other hand, the annuitized investment cost amounts to

$$C_{\text{V}}(T) = C_{\text{F}} + k \cdot l \cdot T = 20\,000\,000 + 35 \cdot 1000 \cdot 800 = 48\,032\,000 \text{ \$/year} \tag{8.50}$$

In this case, withdrawing 150 MW of transmission capacity generates enough additional revenue to cover both the fixed and variable costs.

Withholding some transmission capacity creates larger price differentials and increases the value of transmission. Network users may therefore be willing to pay more to buy financial transmission rights from the owners of this new line, thereby allowing them not only to cover their cost but also to make a profit.

Consider now our third example. Table 8.5, which is similar to Table 8.4, illustrates this effect on an annual basis. The optimal capacity of the interconnection remains 400 MW, independently of the fixed costs.

Let us assume that the interconnection has been built with the optimal capacity and that all this capacity is made available. As we saw in the previous section, the pattern of nodal prices is then such that the revenue derived from the price differentials covers exactly the variable part of the cost of building the transmission line. On the other hand, congestion revenues do not cover the fixed component of the cost of building the interconnection.

Let us try to recover the fixed cost by withdrawing transmission capacity. During the off-peak period, the impact of capacity that is made available on the short-run transmission revenue is shown in Table 8.6.

Reducing the available capacity from 400 MW to 200 MW increases the transmission revenue from 0 to 4 383 900 \$/year. Further reductions in capacity result in a decrease in revenue. Table 8.7 gives the same information for the on-peak period.

In this case, withdrawing available capacity reduces the revenue. This paradox is easily resolved if we consider Figure 8.4. For the on-peak period, the capacity is located to the left of the maximum, while for the off-peak period, it is located to the right of the maximum. Given that the overall contribution of the on-peak period is much greater than the contribution of the off-peak period, it is not possible to increase short-term transmission revenue by withdrawing transmission capacity.

Whenever we deal with fixed costs, it is important to examine the implications of deciding not to build the transmission line in the first place. In this case, the cost

Table 8.5 Annual cost of constraints, annuitized cost of transmission investments (including both fixed and variable costs) and total annual cost of transmission as a function of the transmission capacity of the Borduria–Syldavia interconnection

Interconnection capacity (MW)	Annual constraint cost (k$/year)	Annuitized fixed investment cost (k$/year)	Annuitized variable investment cost (k$/year)	Annuitized investment cost (k$/year)	Total annual transmission cost (k$/year)
0	158 304	20 000	0	20 000	178 304
100	135 835	20 000	14 000	34 000	169 835
200	115 993	20 000	28 000	48 000	163 993
300	98 780	20 000	42 000	62 000	160 780
350	91 159	20 000	49 000	69 000	160 159
400	**84 012**	**20 000**	**56 000**	**76 000**	**160 012**
450	77 157	20 000	63 000	83 000	160 157
500	70 593	20 000	70 000	90 000	160 593
600	58 342	20 000	84 000	104 000	162 342
700	47 257	20 000	98 000	118 000	165 257
800	37 339	20 000	112 000	132 000	169 339
900	28 587	20 000	126 000	146 000	174 587

Table 8.6 Effect of the available transmission capacity on the congestion surplus for the off-peak period

Available capacity (MW)	Generation in Borduria (MW)	Generation in Syldavia (MW)	Marginal cost in Borduria ($/MWh)	Marginal cost in Syldavia ($/MWh)	Hourly surplus ($/h)	Annual surplus ($/year)
100	250	350	12.5	20	750	3 653 250.00
200	350	250	13.5	18	900	4 383 900.00
300	450	150	14.5	16	450	2 191 950.00

Table 8.7 Effect of the available transmission capacity on the congestion surplus for the on-peak period

Available capacity (MW)	Generation in Borduria (MW)	Generation in Syldavia (MW)	Marginal cost in Borduria ($/MWh)	Marginal cost in Syldavia ($/MWh)	Hourly surplus ($/MWh)	Annual surplus ($/year)
100	1000	2600	20	65	4500	17 500 500.00
200	1100	2500	21	63	8400	32 667 600.00
300	1200	2400	22	61	11 700	45 501 300.00

of transmission will obviously be zero. However, the cost of constraints, in our third example, would be at its maximum, that is, 158 304 $/year (see Table 8.5). We therefore conclude that building transmission is justifiable as the total cost would be reduced from 158 304 000 $/year to 150 012 000 $/year.

8.4.9 A three-bus example

We must now explore the effect that Kirchhoff's voltage law has on the value of transmission and the recovery of investments in transmission capacity. To illustrate this issue, we will use the three-bus system shown in Figure 8.10. We will consider the effect of changes in demand by assuming that each year can be divided in two demand periods. Table 8.8 shows the duration of each period and the load at each bus. Note that, unlike the previous two-bus example, the load profile does not follow the same pattern at all buses.

Table 8.9 shows that the marginal cost of generation at each bus increases linearly with output. We will once again make the assumption that the electricity markets at each bus are sufficiently competitive, and that the price of energy at each bus is equal to the marginal cost.

The annuitized transmission investment cost of transmission line c is proportional to its capacity (T_c)and its length (l_c):

$$\Omega_c(T_c) = k_c \cdot l_c \cdot T_c \tag{8.51}$$

Where the marginal annuitized investment cost of the circuit per unit length k_c is 50 \$/(MW · km · year). For the sake of simplicity, we will assume that all lines in our three-bus example have the same 600-km length and thus the same reactance.

We want to determine the capacities of the transmission lines that minimize the sum of the operating cost and of the investment cost for this network. This minimization must be done over the expected life of the system. Since, in this example, we assume that the load pattern repeats itself year after year, we can carry out this optimization over an equivalent hour. This is achieved by multiplying the operating cost for each

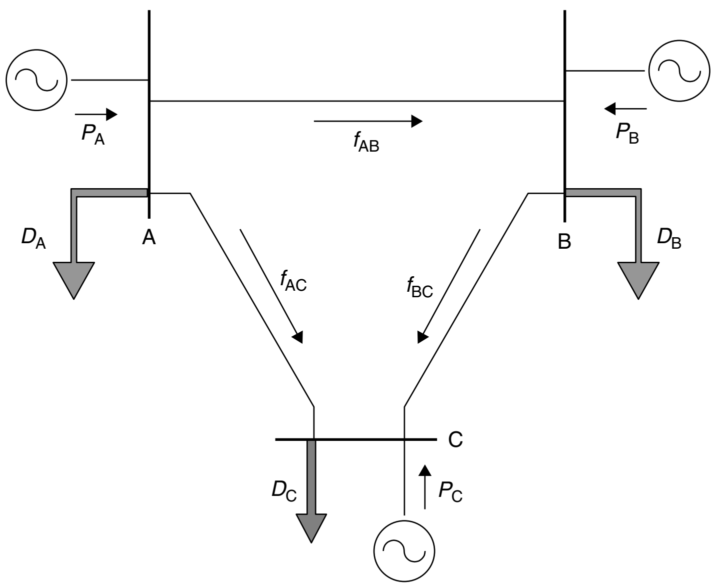


Figure 8.10 Three-bus system used to illustrate the effect of Kirchhoff's voltage law on the value of transmission and the recovery of transmission investments

Table 8.8 Variation of the load with time for the three-bus example

	Period 1	Period 2
Duration (h)	2190	6750
Load at bus A (MW)	0	0
Load at bus B (MW)	10 000	5000
Load at bus C (MW)	2500	10 000

Table 8.9 Marginal costs of electrical energy for the three-bus example

Bus	Capacity (MW)	Marginal cost ($/MWh)
A	5000	$0.003 + 2P_A$
B	7000	$0.003 + 1.35P_B$
C	8000	$0.003 + 1.75P_C$

load period by its duration ($\tau_1 = 2190$ h and $\tau_2 = 6570$ h) and dividing it by the number of hours in a year ($\tau_0 = 8760$ h). The objective function of this optimization problem is thus

$$\min_{T_{AB}, T_{AC} T_{BC}} \left[\sum_{t=1}^{t=2} \frac{\tau_p}{\tau_0} \left(\sum_{i \in \{A,B,C\}} a_i P_{it} + \frac{1}{2} b_i P_{it}^2 \right) + \frac{k_{AB} \cdot l_{AB} \cdot T_{AB}}{\tau_0} + \frac{k_{AC} \cdot l_{AC} \cdot T_{AC}}{\tau_0} + \frac{k_{BC} \cdot l_{BC} \cdot T_{BC}}{\tau_0} \right] \tag{8.52}$$

This minimization is subject to the following constraints imposed by KCL and KVL for demand period 1:

$$\begin{aligned} f_{AB1} + f_{AC1} - P_{A1} + D_{A1} &= 0 \\ -f_{AB1} + f_{BC1} - P_{B1} + D_{B1} &= 0 \\ -f_{AC1} - f_{BC1} - P_{C1} + D_{C1} &= 0 \\ f_{AB1} + f_{AC1} - f_{BC1} &= 0 \end{aligned} \tag{8.53}$$

And for demand period 2:

$$\begin{aligned} f_{AB2} + f_{AC2} - P_{A2} + D_{A2} &= 0 \\ -f_{AB2} + f_{BC2} - P_{B2} + D_{B2} &= 0 \\ -f_{AC2} - f_{BC2} - P_{C2} + D_{C2} &= 0 \\ f_{AB2} + f_{AC2} - f_{BC2} &= 0 \end{aligned} \tag{8.54}$$

Furthermore, the line flows during each period must remain below the (as yet unknown) capacity of the corresponding lines:

$$
\begin{aligned}
f_{\mathrm{AB1}}, f_{\mathrm{AB2}} &\leq T_{\mathrm{AB}} \\
f_{\mathrm{AC1}}, f_{\mathrm{AC2}} &\leq T_{\mathrm{AC}} \\
f_{\mathrm{BC1}}, f_{\mathrm{BC2}} &\leq T_{\mathrm{BC}}
\end{aligned}
\tag{8.55}
$$

Finally, the output of the generators connected to each bus must remain below their rated capacity during each demand period:

$$
\begin{aligned}
P_{\mathrm{A1}} &\leq P_{\mathrm{A}}^{\max}; P_{\mathrm{A2}} \leq P_{\mathrm{A}}^{\max} \\
P_{\mathrm{B1}} &\leq P_{\mathrm{B}}^{\max}; P_{\mathrm{B2}} \leq P_{\mathrm{B}}^{\max} \\
P_{\mathrm{C1}} &\leq P_{\mathrm{C}}^{\max}; P_{\mathrm{C2}} \leq P_{\mathrm{C}}^{\max}
\end{aligned}
\tag{8.56}
$$

This quadratic optimization problem is too complex for a manual solution but can be solved numerically using a spreadsheet. Figures 8.11 and 8.12 show the optimal generation dispatch, the line flows and the nodal prices for the two demand periods. Table 8.10 shows the detail of the operating costs. Since the duration of period 1 represents 25% of the total and the duration of period 2, the remaining 75%, the costs for each period are expressed in \$/0.25 h and \$/0.75 h respectively. The generating costs reflect the duration of each period. The operating cost for an equivalent hour is then obtained by adding the costs for each period. The annual cost is obtained by multiplying the cost for an equivalent hour by the number of hours in a year.

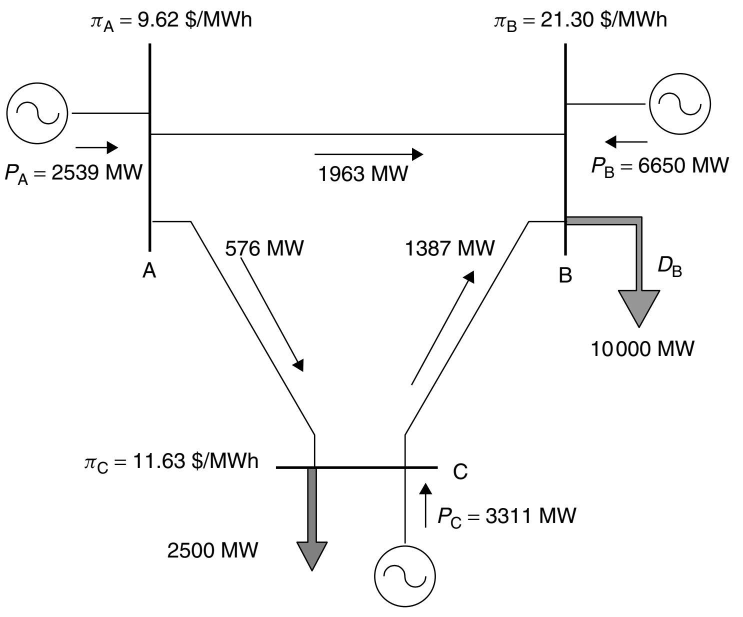


Figure 8.11 Optimal generation dispatch, line flows and nodal prices for demand period 1

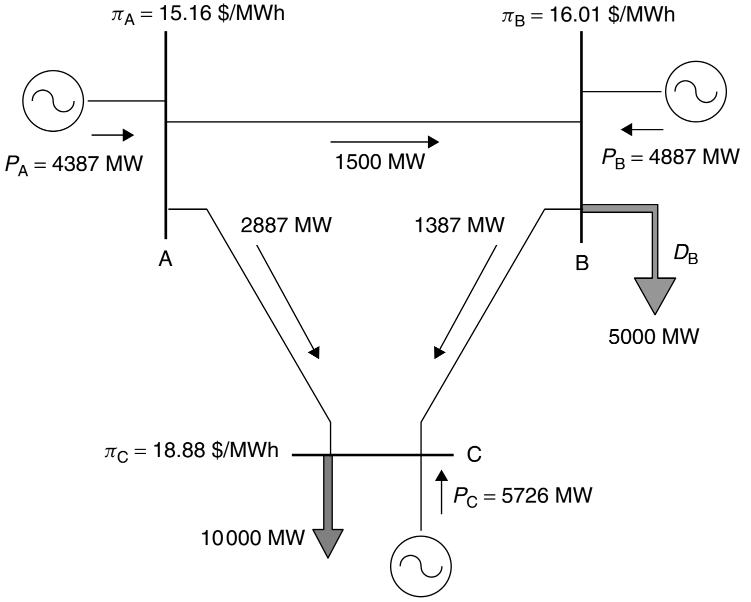


Figure 8.12 Optimal generation dispatch, line flows and nodal prices for demand period 2

Table 8.10 Optimal hourly operating cost for the three-bus example

Bus	Period 1 ($/0.25 h)	Period 2 ($/0.75 h)	Cost for an equivalent hour ($/h)	Annual cost ($/year)
A	3687	28 233	31 920	279 619 200
B	18 827	31 817	50 644	443 641 440
C	5519	44 184	49 703	435 398 280
Total	28 033	104 234	132 267	1 158 658 920

Table 8.11 shows the flows in each transmission line as well as its optimal capacity and the corresponding hourly and annual investment costs. The flow in each line reaches its maximum (and hence sets the capacity) in one of the periods. Since we minimized the total operating and investment cost, the whole capacity of the line should indeed be fully utilized during at least one period. In this particular case, the flow between buses B and C fully utilizes the capacity of that line during both periods, but in opposite directions.

Let us now turn our attention to the nodal prices and the revenues. Table 8.12 summarizes that information for each period. Negative quantities represent payments to generators, while positive quantities denote payments to loads. The revenues for each period are proportional to their duration and the revenues for an equivalent hour are a weighted average of the revenue for each period. The grand total (shown in the bottom right-hand corner of Table 8.12) represents the total congestion surplus that would be collected during an equivalent hour. This quantity is exactly equal to the

Table 8.11 Optimal line capacities and investment costs for the three-bus example

Line	Flow in period 1 (MW)	Flow in period 2 (MW)	Optimal capacity (MW)	Hourly investment cost ($/h)	Annual investment cost ($/year)
A-B	1963	1500	1963	6723	58 891 939
A-C	576	2887	2887	9887	86 612 631
B-C	−1387	1387	1387	4750	41 612 636
			Total:	21 360	187 117 206

Table 8.12 Nodal prices and revenues for the three-bus example

Bus	Nodal prices		Revenues		
	Period 1 ($/MWh)	Period 2 ($/MWh)	Period 1 ($/0.25 h)	Period 2 ($/0.75 h)	Equivalent hour ($/h)
A	9.62	15.16	−6105	−49 885	−55 990
B	21.30	16.01	17 839	1356	19 195
C	11.63	18.88	−2359	60 514	58 155
Total			9375	11 985	21 360

Table 8.13 Congestion revenues and investment costs for each line of the three-bus example

Line	Period 1			Period 2				
	Δ Price ($/MWh)	Flow (MW)	Revenue ($/0.25 h)	Δ Price ($/MWh)	Flow (MW)	Revenue ($/0.75 h)	Total revenue ($/h)	Investment costs ($/h)
A-B	11.68	1963	5732	0.85	1500	956	6688	6723
A-C	2.01	576	289	3.72	2887	8055	8344	9887
B-C	−9.67	−1387	3353	2.86	1387	2975	6339	4750
Total			9374			11 986	21 360	21 360

total hourly investment cost given in Table 8.10. This equivalence demonstrates that, in the absence of fixed costs, short-run marginal pricing generates a sufficient amount of revenue to cover the cost of transmission investments.

Table 8.13 provides the information needed to calculate the revenue "earned" by each line during each period and over the equivalent hour. As we discussed in Chapter 6, differences in nodal prices arise between two buses even when the line connecting these two buses is not congested. For example, during period 1, the flow in the line between buses A and C is 576 MW, well below its 2887 MW capacity. However, congestion in Lines A-B and B-C creates a 2.02 $/MWh price differential between nodes A and C. The flow in that line thus generates a revenue of

$$R_{\mathrm{AC},1} = 576 \times 2.02 \times 0.25 = 289 \text{ \$/0.25 h} \tag{8.57}$$

During period 2, when the flow in this line is equal to its capacity, it generates a revenue of

$$R_{AC,2} = 2887 \times 3.72 \times 0.75 = 8055 \text{ \$/0.75 h} \tag{8.58}$$

The revenue "collected" by this line during an equivalent hour is thus 8344 $/h. It is not equal to the 9887 $/h cost of this line given in Table 8.11. Similarly, the 1500 MW flow in Line A-B during period 2 is less than its 1963 MW capacity. However, the price differentials across that line generate some revenue. These results demonstrate that hourly SRMC revenues associated with individual lines do not match their hourly investment cost. However, Table 8.13 also shows that the total congestion surplus recovers exactly the investment cost of this transmission network. This result is not a coincidence and holds for all networks, no matter how complex. If the entire network is owned by the same entity, this cross-subsidization between lines is not a problem. On the other hand, it is not clear how FTRs could be sold to the network users under these conditions. For example, let us suppose that lines A-B and B-C belong to the incumbent utility and Line A-C has been developed by a merchant transmission company. If revenues are allocated on the basis of nodal price differentials, the owner of Line A-C would recover only 8344 $/h instead of its cost of 9887 $/h. On the other hand, the incumbent utility would recover more than its cost. It is also not clear on what basis negotiation between users and network owners about the purchase of FTRs could proceed.

8.4.10 Concept of reference network

In the examples that we discussed in the previous sections, we determined the optimal capacity of a new transmission line by minimizing the sum of the operating cost and the cost of investments in transmission. Maintaining that balance for the system as a whole is a major challenge for regulatory authorities in a competitive environment because generation and transmission operate as separate entities. If we assume that the transmission network operates as a monopoly, the regulator needs to devise a set of incentives that encourage the right level of transmission investments. In order to do so, the regulator needs a way to measure the overall performance of the system. This can be achieved using a reference network.

In its simplest form, a reference network is topologically identical to the existing network, and generators and loads are unchanged. On the other hand, each transmission line has an optimal capacity. Optimal capacities are determined as we did in the examples above. One important difference, however, is that instead of optimizing the capacity of one or a few new lines, the procedure is applied to the whole transmission system, including both new and existing lines.

A reference network is thus a network against which the real one can be objectively compared. Optimal investment costs and the optimal congestion costs can be quantified and compared with those found in the real system. Furthermore, by comparing the capacities of individual lines in the reference network and the real network, the needs for new investment can be identified. Stranded investments can also be detected. A comparison of optimal and actual operating costs can also be performed. Differences between actual and reference network operation and investments could be used as a measure of the performance of a transmission company. The regulator could then set financial incentives on that basis.

The concept of a reference network has a long history and a solid foundation in economic theory. See, for example, Boiteux (1949), Nelson (1967) and Farmer *et al.* (1995).

8.4.11 Generalization

In this section, we present a general formulation of the transmission expansion problem for pricing and regulatory purposes. This involves determining an optimally designed transmission network. Determining such a reference network requires the solution of a type of security constrained optimal power flow (OPF) problem. In its simplest form, this problem can be formulated using a conventional dc optimal power flow. The objective of this optimization is to minimize the sum of the annual generation cost and the annuitized cost of transmission. This optimization is constrained by Kirchhoff's current and voltage laws as well as the limits on system components. It must cover several demand levels using a yearly load-duration curve as described earlier. Finally, it must also take into account credible outages of transmission and generation facilities.

8.4.11.1 Notations

In order to state the problem mathematically, we need to introduce the following notations:

np: Number of demand periods

nb: Number of buses

ng: Number of generators

nl: Number of lines

nc: Number of contingencies

τ_p: Duration of demand period p

D_p: Nodal demand vector for period p

C_g: Operating cost of generator g

P_{pg}: Output of generator g during demand period p

P_p: Vector of nodal generations for demand period p

$P^{\max}$: Vector of maximum nodal generations

$P^{\min}$: Vector of minimum nodal generations

A^0: Node-branch incidence matrix for the intact system

A^c: Node-branch incidence matrix for contingency c

H^0: Sensitivity matrix for the intact system

H^c: Sensitivity matrix for contingency c

k_b: Annuitized investment cost for line b in \$/(MW · km · year)

l_b: Length of line b (km)

T_b: Capacity of line b

T: Vector of line capacities

F_p^0: Vector of line flows for the intact system during period p

F_p^c: Vector of line flows for contingency c during period p

The sensitivity matrix H that relates injections and power flows is defined as follows:

$$[H] = [Y_\mathrm{d}] \cdot [A^\mathrm{T}] \cdot \begin{bmatrix} 0 & 0 \\ 0 & [Y_\mathrm{bus}^\mathrm{r}]^{-1} \end{bmatrix} \tag{8.59}$$

Where Y_d is the diagonal matrix of branch admittances and Y_bus^r is obtained from the system admittance matrix Y_bus by removing the row and the column corresponding to the slack bus to make it nonsingular. The elements of the sensitivity matrix H are called sensitivity factors:

$$h_{in} = \frac{\Delta F_k}{\Delta P_n}$$

This sensitivity factor relates the change in the power flow in branch i to an increase in injection at node n. In the conventional dc power flow model, these sensitivity factors depend only on the network topology and the reactances of the network circuits and not on the loading conditions. Hence, for a network with a fixed topology, the sensitivity factors are constant and are evaluated without considering generation and demand.

Wood and Wollenberg (1996) show that if branch i connects buses u and v, the sensitivity factors relating the flow in that branch to the injection at bus n can be calculated as follows:

$$h_{in} = \frac{\Delta F_k}{\Delta P_n} = \frac{1}{x_{uv}}(X_{un} - X_{vn}) \tag{8.60}$$

Where X_{un} and X_{vn} are elements of the inverse of the reduced admittance matrix Y_bus^r. Although the values of the sensitivity factors depend on the choice of reference node, the result of the optimization problem is indifferent to this choice.

8.4.11.2 Problem formulation

The objective function of this optimization problem can be expressed as follows:

$$\min_{P_{pg},T_b}\left(\sum_{p=1}^{np}\tau_p\sum_{g=1}^{ng}C_g P_{pg}+\sum_{b=1}^{nl}k_b l_b T_b\right) \tag{8.61}$$

Since this problem covers several demand periods over a year, it must satisfy the power flow equations for the intact system and the line capacity limits for each of them. Using a dc power flow formulation neglecting losses, these constraints are

$$A^0F_p^0 - P_p + D_p = 0 \tag{8.62}$$

$$F_p^0 = H^0(P_p - D_p) \tag{8.63}$$

$$F_p^0 - T \leq 0 \tag{8.64}$$

$$-F_p^0 - T \leq 0 p = 1, np \tag{8.65}$$

Equation (8.62) is a nodal balance constraint derived from Kirchhoff's current law, which requires that the total power flowing into a node must be equal to the total power flowing out of the node. Constraint (8.63) relates flows and injections on the basis of Kirchhoff's voltage law. The last two equations represent thermal constraints on the line flows. All these constraints must also be satisfied for each contingency during each demand period:

$$A^cF_p^c - P_p + D_p = 0 \tag{8.66}$$

$$F_p^c = H^c(P_p - D_p) \tag{8.67}$$

$$F_p^c - T \leq 0 \tag{8.68}$$

$$-F_p^c - T \leq 0 \quad p = 1, np;\ c = 1, nc \tag{8.69}$$

Finally, the optimization must respect the limits on the output of the generators:

$$P^{\min} \leq P_p \leq P^{\max} \quad p = 1, np \tag{8.70}$$

Since the object of the optimization is to determine the optimal thermal capacity of the lines, this variable can take any positive value:

$$0 \leq T \leq \infty \tag{8.71}$$

8.4.11.3 Implementation

The above model calculates the vector of generation dispatch P_p in each demand period, the vector of line flows F_p^0 in each demand period and the vector of optimal capacities T valid in all demand periods. All other parameters in the above equations are either specified or determined from the network topology and data. Since we have assumed constant generation marginal costs, the optimization problem is linear. However, because of its size, this problem is usually not solved in its original form. Instead, it is solved using the iterative algorithm shown in Figure 8.13. At the start of each iteration, we establish a generation dispatch and calculate the capacity of each line in such a way that demand is met during each period and that the transmission constraints are satisfied. Note that at the beginning of the process there are no transmission constraints. The feasibility of this dispatch is then evaluated by performing a power flow analysis for all contingent networks in each demand period. If any of the line flows is greater than the proposed capacity of the line, a constraint is created and inserted in the OPF at the next iteration. For example, if line b is overloaded, the following constraint is added to the problem:

$$-T_b \leq f_b^{ps} + \sum_{j=1}^{nb} h_{jb}^S \cdot (P_j^p - P_j^{p0}) \leq T_b \tag{8.72}$$

where S represents the network topology for both the intact and contingent conditions and h_{ji}^S are the corresponding sensitivity factors. This process is repeated until all line overloads are eliminated.

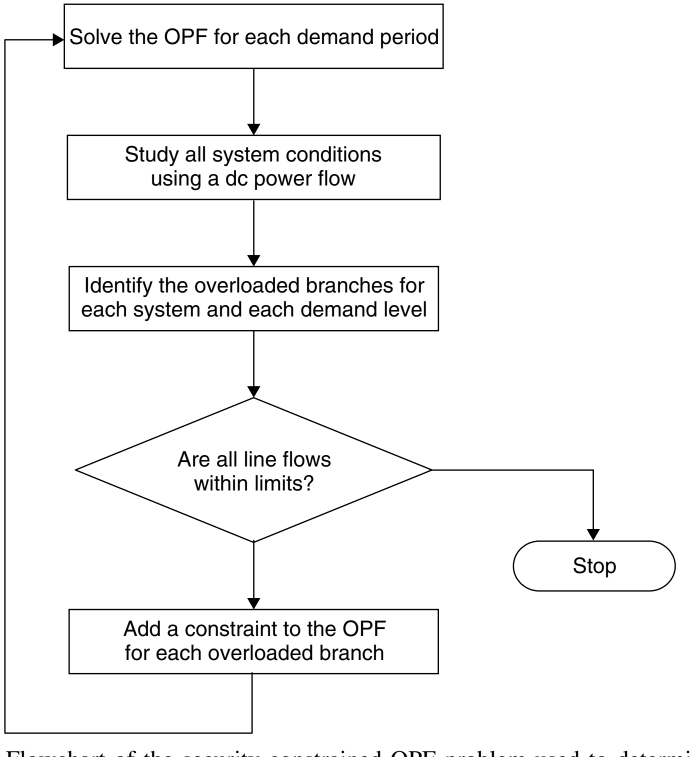


Figure 8.13 Flowchart of the security constrained OPF problem used to determine the reference network

The nodal prices can then be calculated as follows:

$$\pi_j^p = \pi^p + \sum_{s=1}^{nc} \sum_{b=1}^{nl} h_{jb}^s \cdot \mu_b^{ps} \tag{8.73}$$

where π^p is the Lagrange multiplier associated with the load balance constraint for demand period p in the intact network. This quantity is frequently called the *system marginal cost*. The variables μ_b^{ps} are the Lagrange multipliers associated with the transmission constraints (8.72) that are generated in the iterative process.

8.4.11.4 Example

This optimization procedure has been applied to the IEEE 24 bus reliability test system (RTS) depicted in Figure 8.14. For the details of this network, see IEEE (1979). Figure 8.15 shows that, except for a small number of lines, line flows are well below

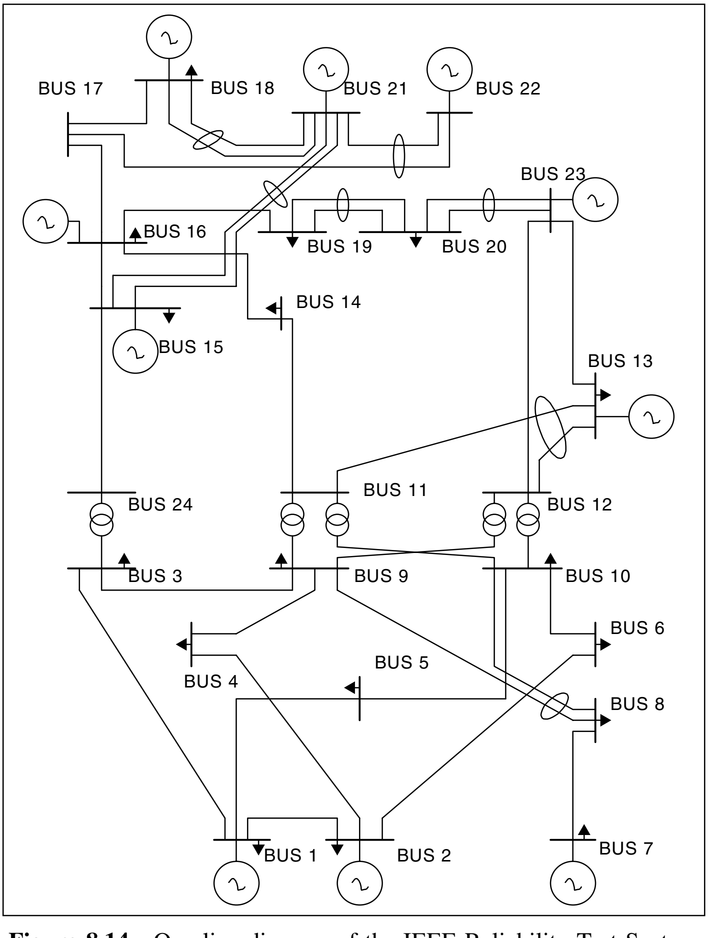


Figure 8.14 One-line diagram of the IEEE Reliability Test System

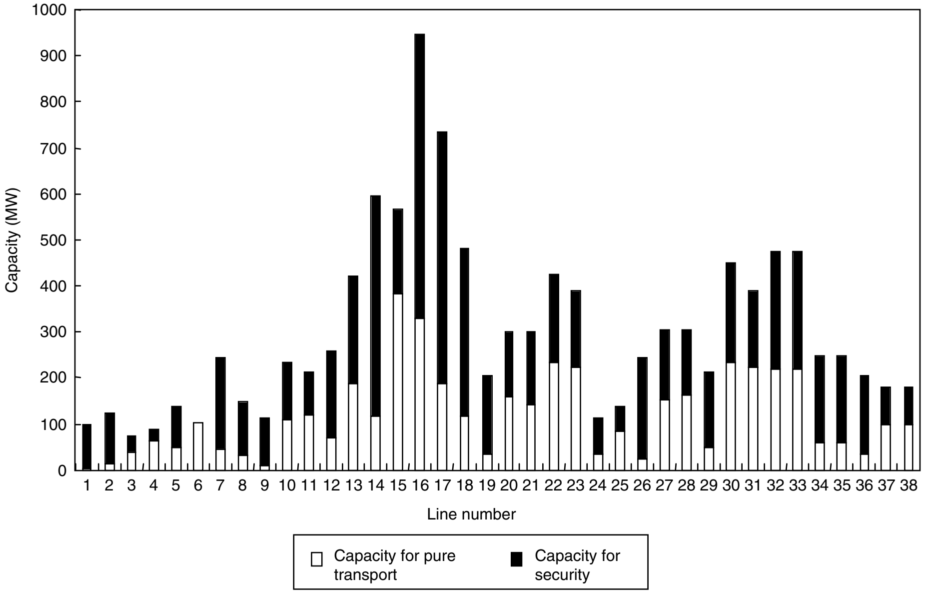


Figure 8.15 Comparison of the capacity needed for the intact network (pure transport) with the capacity needed to ensure security during the maximum demand period for the IEEE RTS system

50% of the optimal capacity even during the period of maximum demand. This observation confirms the importance of taking security into consideration when designing and pricing a transmission network.

8.4.11.5 Considering other factors

This basic algorithm for constructing the reference network becomes considerably more complex if we want to optimize the network topology and the choice of transmission voltage levels, or if we want to deal with load growth, economies of scale, new transmission technologies such as Flexible AC Transmission Systems, distributed generation, demand-side management, losses, reactive power, network stability constraints and generation reserve. The appropriate degree of complexity depends on the intended application and the specific system. However, it is important to bear in mind that the purpose of a reference network is not to replace the detailed technical design of the transmission network, but to support decisions regarding regulation, investments and pricing.

8.5 Further Reading

Boiteux, M, La tarification des demandes en pointe: application de la théorie de la vente au coût marginal, *Revue Générale de l'Electricité*, **58**, 1949, 321–340.

Farmer, E D, Cory, B J, Perera, B L P P, Optimal pricing of transmission and distribution services in electricity supply, *IEE Proceedings on Generation, Transmission and Distribution*, **142**(1), pp. 1–8, 1995.

Hogan, W W, *Market-Based Transmission Investments and Competitive Electricity Markets*, August 1999, available online from *www.ksg.harvard.edu/whogan*.
Hogan, W W, *Transmission Market Design*, April 2003, available online from *www.ksg.harvard.edu/whogan*.
IEEE Power Engineering Society Subcommittee on the Application of Probabilistic Methods, A Reliability Test System, *IEEE Transactions on Power Apparatus and Systems*, **PAS-98**(6), pp. 2047–2054, 1979.
Marangon Lima, J W, Allocation of transmission fixed charges: an overview, *IEEE Transactions on Power Systems*, **11**(3), 1996, 1409–1418.
Nelson, J R, *Marginal Cost Pricing in Practice*, Prentice Hall, Upper Saddle River, NJ, 1967.
Wood A J, Wollenberg B F, *Power Generation, Operation and Control*, Second Edition, John Wiley & Sons, 1996.
Woolf F, *Global Transmission Expansion: Recipes for Success*, Pennwell, Tulsa, OK, 2003.

8.6 Problems

8.1 Summarize the regulatory process used for transmission expansion in your region or country, or in another area for which you have access to sufficient information.

8.2 Identify the method used to allocate the cost of transmission investments in your region or country, or in another area for which you have access to sufficient information.

8.3 Consider the two-bus power system shown in Figure P8.1. Assume that the demand is constant and insensitive to price and that energy is sold at its marginal cost of production and that there are no limits on the output of the generators. What is the maximum price that could be charged for transmission if the marginal costs of generation are as follows?

$$MC_{\mathrm{A}} = 25 \text{ \$/MWh}$$

$$MC_{\mathrm{B}} = 17 \text{ \$/MWh}$$

8.4 Consider the two-bus power system shown in Figure P8.1. Assume that the demand is constant and insensitive to price and that energy is sold at its marginal cost of production and that there are no limits on the output of the generators. The marginal cost of production of the generators connected to buses A and B are given respectively by the following expressions:

$$MC_{\mathrm{A}} = 20 + 0.03 P_{\mathrm{A}} \text{ \$/MWh}$$

$$MC_{\mathrm{B}} = 15 + 0.02 P_{\mathrm{B}} \text{ \$/MWh}$$

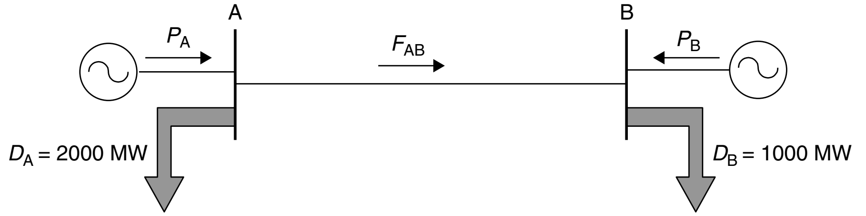


Figure P8.1 Two-bus power system for Problems 8.3, 8.4, 8.5, 8.6, 8.7, 8.8 and 8.9

Plot the marginal value of transmission as a function of the capacity of the transmission line connecting buses A and B.

8.5 Determine the transmission demand function for the system of Problem 8.4.

8.6 Calculate the hourly long-range marginal cost of the transmission line of Problem 8.4 assuming that the line is 500 km long, that the amortized variable cost of building the line is 210 \$/(MW · km · year).

8.7 Determine the optimal capacity of the transmission line of Problems 8.4, 8.5 and 8.6, assuming the loading conditions shown in Figure P8.1.

8.8 Determine the optimal capacity of the transmission line of Problems 8.4, 8.5 and 8.6, for the three-part load-duration curves summarized in the table below. Assume that the periods of high, medium and low load coincide at both buses.

Period	**Load at A (MW)**	**Load at B (MW)**	**Duration (h)**
High	4000	2000	1000
Medium	2200	1100	5000
Low	1000	500	2760

Compare the amount of congestion revenue collected annually for this optimal transmission capacity with the annuitized cost of building the transmission line.

8.9 Calculate the amount of congestion revenue collected annually for a transmission capacity 33.3% higher and 33.3% lower than the optimal transmission capacity calculated in Problem 8.8. Compare these values to the annuitized cost of building the transmission line.

Appendix
Answers to Selected Problems

Chapter 2

2.1 Revenue $= q \times \dfrac{dC}{dq} = 50q^2 + 2000q$
Profit = Revenue − Cost $= 25q^2$

2.2 a. $q^{\max} = 200$

b. $\pi^{\max} = 2000$ \$/widget

c. Maximum surplus = \$200 000. It cannot be realized because it is unlikely that producers will sell for nothing.

d. $q = 100$
Gross consumers' surplus = \$150 000
Revenue = \$100 000
Net consumers' surplus = \$50 000

e. $q = 80$
Revenue = \$96 000

f. $\varepsilon = -1$

g. Gross consumers' surplus $= 2000q - 5q^2$
Net consumers' surplus $= 5q^2$

h. Gross consumers' surplus $= 200\,000 - 0.05\pi^2$
Net consumers' surplus $= 200\,000 - 200\pi + 0.05\pi^2$

2.3 a. $q = 120$
$\pi = 800$ \$/widget

b. Consumers' gross surplus: \$168 000
Consumers' net surplus: \$72 000

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

Producers' revenue: \$96 000
Producers' profit: \$36 000
Global welfare: \$108 000

2.4 a. $\pi = 900$ \$/widget
$q = 110$
Consumers' net surplus: \$60 500
Producers' profit: \$46 750
Global welfare: \$107 250

b. $\pi = 600$ \$/widget
$q = 80$
Consumers' net surplus: \$80 000
Producers' profit: \$16 000
Global welfare: \$96 000

c. $\pi = 1100$ \$/widget
$q = 90$
Consumers' net surplus: \$40 500
Producers' profit: \$20 250
Tax revenue: \$40 500
Global welfare: \$101 250

2.5

	$q = 200 - \pi$		$q = 10\,000/\pi$	
q	π	ε	π	ε
0	200	$-\infty$	∞	-1
50	150	-3	200	-1
100	100	-1	100	-1
150	50	$-1/3$	66.6	-1
200	0	0	50	-1

2.6 $\varepsilon_{11} = -0.120$
$\varepsilon_{12} = 0.048$
$\varepsilon_{22} = -0.108$
$\varepsilon_{21} = 0.160$

2.8 a. Since the average production cost must be lower than the price, we get $65 \leq y \leq 155$; $y^{opt} = 110$

b. The fixed cost is so high that there is no range of production at which the firm would make a profit.

Chapter 3

3.2

	Pool Price ($/MWh)	NSPCo	SAlCo
a.	16	Produces 200 MWh Receives $3200 from the pool	Consumes 200 MWh Pays $3200 to the pool
	18	Produces 200 MWh Receives $3600 from the pool Pays $400 to SAlCo	Consumes 200 MWh Pays $3600 to the pool Receives $400 from NSPCo
	13	Produces 200 MWh Receives $2600 from the pool Receives $600 from SAlCo	Consumes 200 MWh Pays $2600 to the pool Pays $600 to NSPCo
b.	18	Produces 50 MWh Receives $900 from the pool Pays $400 to SAlCo	Consumes 200 MWh Pays $3600 to the pool Receives $400 from NSPCo
c.	13	Produces 200 MWh Receives $2600 from the pool Receives $600 from SAlCo	Consumes 100 MWh Pays $1300 to the pool Pays $600 to NSPCo

3.3

Company	Profit ($)
Red	650
Green	1280
Blue	1325
Yellow	515
Magenta	287.50
Purple	125

3.4 a. The supply curve is piecewise constant and is as follows in tabular form:

Company	Amount (MWh)	Cumulative Amount (MW)	Price ($/MWh)
Blue	200	0–200	10.5
Red	100	200–300	12.5
Blue	200	300–500	13
Green	50	500–550	13.5
Red	100	550–650	14
Green	50	650–700	14.5
Blue	100	700–800	15
Green	50	800–850	15.5
Red	50	850–900	18

b.

Forecast Load (MW)	Demand (MW)	Price ($/MWh)	Blue Production (MWh)	Blue Revenue ($)	Red Production (MWh)	Red Revenue ($)	Green Production (MWh)	Green Revenue ($)
400	400	13.00	300	3900	100	1300	0	0
600	600	14.00	400	5600	150	2100	50	700
875	875	18.00	500	9000	225	4050	150	2700

c.

Forecast Load (MW)	Demand (MW)	Price ($/MWh)	Blue Production (MWh)	Blue Revenue ($)	Red Production (MWh)	Red Revenue ($)	Green Production (MWh)	Green Revenue ($)
400	348	13.00	248	3224	100	1300	0	0
600	546	13.50	400	5400	100	1350	46	621
875	813	15.50	500	7750	200	3100	113	1751.50

3.5 a.

Item	Energy bought (MWh)	Energy sold (MWh)	Price ($)	Expenses ($)	Revenue ($)
Industrial customer		50	19.00		950.00
Other customers		1150	21.75		25 012.50
Future contract		200	21.00		4200.00
Put option		200	23.50		4700.00
Long-term contract	600		20.00	12 000.00	
Future contract	100		22.00	2200.00	
Call option	150		20.50	3075.00	
Generation	300		21.25	6375.00	
Spot market purchase	450		21.50	9675.00	
150 MW Call option fee			1.00	150	
200 MW Put option fee			1.00	200	
300 MW Call option fee			1.00	300	
Profit				887.50	
Balance	1600	1600		34 862.50	34 862.50

b. If the spot price increases to $23.47, the cost of the 450 MWh purchase on the spot market would offset the profit. The 20.50 $/MWh call option and the 23.50 $/MWh put option would still be in the money. The 24.00 $/MWh call option would still be out of the money.

3.6 Since the market operator accepted 175 MW of bids, using the supply curve, we determine that the spot market price was 21.00 $/MWh.

a.

Item	**Energy bought (MWh)**	**Energy sold (MWh)**	**Price ($)**	**Expenses ($)**	**Revenue ($)**
Future T4	600		20.00	12 000.00	
Nuclear unit	400		16.00	6400.00	
Gas-fired unit	200		18.00	3600.00	
Forward T1		50	21.00		1050.00
Long-term T3		350	20.00		7000.00
Forward T5		100	22.00		2200.00
Exercise Put option T6		250	23.50		5875.00
Spot sale T9		50	21.00		1050.00
Residential customers		300	25.50		7650.00
Commercial customers		200	25.00		5000.00
Balancing spot purchase	100		21.00	2100	
Fee option T6			2.00	500.00	
Fee option T7			2.00	400.00	
Fee option T8			2.00	200.00	
Profit				4625.00	
Balance	1300	1300		29 825.00	29 825.00

b. Borduria Energy's deficit for that period would increase from 100 MW to 500 MW. The spot price would increase from 21.00 to 28.00 $/MW. The cost of spot purchases would increase from $2000 to $14 000 but the cost of operating the nuclear power plant would drop to zero. Borduria Energy would therefore incur a loss of $975.00

Chapter 4

4.1 Cheapo Electrons makes a $1738.50 loss. The breakeven rate is 25.52 $/MWh.
4.2 The unit makes an operational profit of $690.07.

4.3 The unit makes an operational profit of $688.00

4.4 The unit should be brought on-line at the beginning of Period 3 and shutdown at the end of Period 5. Its operational profit would then be $976.43.

4.5 The unit should be brought on-line at the beginning of Period 3 and shutdown at the end of Period 6. Its operational profit would then be $680.43.

4.6 $P_A = 95.3$ MW; $P_B = 74.2$ MW; $P_C = 180.5$ MW. Total hourly cost = 1927.15 $/h

4.7 $P_A = 85$ MW; $P_B = 66$ MW; $P_C = 160$ MW; Market purchase: 39 MW
Total hourly cost = 1911.20 $/h

4.8 $P_A = 110$ MW; $P_B = 86$ MW; $P_C = 210$ MW; Market sale: 56 MW
Profit from the sale: $33.03

4.9 $P_A = 100$ MW; $P_B = 80$ MW; $P_C = 210$ MW; Market sale: 40 MW
Profit from the sale: $27.23

4.10 $P_A = 25$ MW; $P_B = 30$ MW; $D = 55$ MW; $\pi = 65$ $/MWh; $\Omega_A = \$725$; $\Omega_B = \$1,020$

4.11 $P_A = 26.33$ MW; $P_B = 31.33$ MW; $D = 55$ MW; $\pi = 57.66$ $/MWh; $\Omega_A = \$694$; $\Omega_B = \$982$

4.12 Profit: $5235; Efficiency that reduces profit to zero: 66.33%

Chapter 5

5.1 350 MW

5.3 a: 600 MW; b: 300 MW; c: 500 MW; d: 660 MW; e: 640 MW; f: 759 MW.

5.4 92.3 MW

5.5 106.5 MW

Chapter 6

6.1

	F_{1-2} (MW)	F_{1-3} (MW)	F_{2-3} (MW)	Feasible?
Set 1	−120	20	80	Yes
Set 2	0	400	400	No
Set 3	80	−180	−220	Yes

6.2 a. $\pi_A = 80$ $/MWh; $\pi_B = 35$ $/MWh; $P_A = 2000$ MW; $P_B = 1000$ MW; $F_{AB} = 0$

b. $\pi_A = \pi_B = 53$ $/MWh; $P_A = 1100$ MW; $P_B = 1900$ MW; $F_{AB} = -900$ MW

c. $\pi_A = \pi_B = 65$ $/MWh; $P_A = 1500$ MW; $P_B = 1500$ MW; $F_{AB} = -500$ MW

d. $\pi_A = \pi_B = 57$ \$/MWh; $P_A = 900$ MW; $P_B = 2100$ MW; $F_{AB} = -1100$ MW

e. $\pi_A = 62$ \$/MWh; $\pi_B = 47$ \$/MWh; $P_A = 1400$ MW; $P_B = 1600$ MW; $F_{AB} = -600$ MW

6.3

Case:	a	b	c	d	e
E_A (\$)	160 000	106 000	130 000	114 000	124 000
E_B (\$)	35 000	53 000	65 000	57 000	47 000
R_A (\$)	160 000	58 300	97 500	51 300	86 800
R_B (\$)	35 000	100 700	97 500	62 700	75 200

The generator at B and the demand at A benefit from the line because it increases the price at B and lowers the price at A.

6.4 \$9000. The congestion surplus is equal to zero when the flow is equal to zero and when it is equal to the unconstrained value of −900 MW.

6.5 $P_A = 0$ MW; $P_B = 0$ MW; $P_C = 120$ MW; $P_D = 400$ MW
$\pi_1 = \pi_2 = \pi_3 = 10$ \$/MWh

6.6 $F_{21} = 120$ MW; $F_{31} = 280$ MW; $F_{32} = 200$ MW

Line 1–3 is overloaded by 30 MW.

6.7 Method 1:

$P_A = 0$ MW; $P_B = 48$ MW; $P_C = 72$ MW; $P_D = 400$ MW

$F_{21} = 102$ MW; $F_{31} = 250$ MW; $F_{32} = 182$ MW

Increase in cost: \$240

Method 2:

$P_A = 80$ MW; $P_B = 0$ MW; $P_C = 40$ MW; $P_D = 400$ MW

$F_{21} = 150$ MW; $F_{31} = 250$ MW; $F_{32} = 150$ MW

Increase in cost: \$160

Method 2 is preferable because it is cheaper.

6.8 $\pi_1 = 13.33$ \$/MWh; $\pi_2 = 12.00$ \$/MWh; $\pi_3 = 10.00$ \$/MWh

6.9 $P_A = 63.33$ MW; $P_B = 10$ MW; $P_C = 6.67$ MW; $P_D = 400$ MW

$\pi_1 = 15$ \$/MWh; $\pi_2 = 12$ \$/MWh; $\pi_3 = 10$ \$/MWh

6.10 $F_{BA} = 730\,\text{MW}$; $P_A = 1270\,\text{MW}$; $P_B = 1783\,\text{MW}$; Losses $= 53\,\text{MW}$

$\pi_A = 58.10\,\$/\text{MWh}$; $\pi_B = 50.67\,\$/\text{MWh}$;

Surplus: $2727

6.11

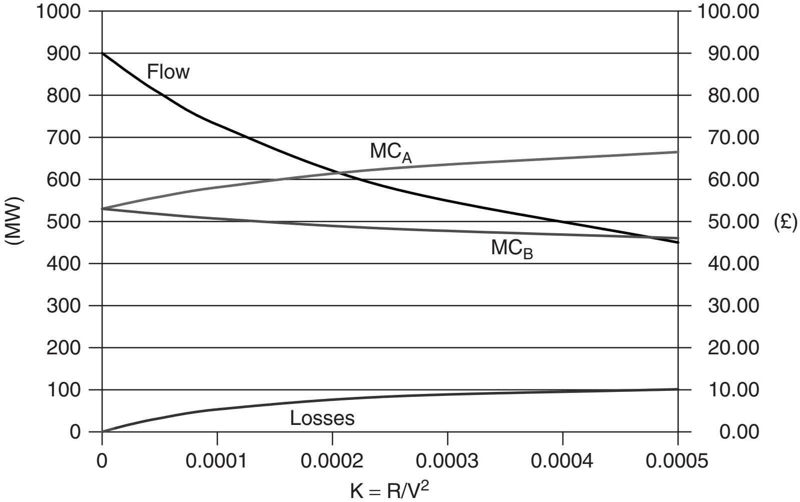


Figure P6.11 Losses and optimal flow

6.12 $\begin{cases} Y_{11}\pi_1 - y_{13}\mu_{31} = y_{12}\pi_2 + y_{13}\pi_3 \\ -y_{21}\pi_1 = -Y_{22}\pi_2 + y_{23}\pi_3 \end{cases}$

$\begin{cases} \pi_1 = 13.33\,\$/\text{MWh} \\ \mu_{31} = 5.33\,\$/\text{MWh} \end{cases}$

6.13 Slack bus at bus 1:

$\begin{cases} -y_{21}\pi_1 = -Y_{22}\pi_2 + y_{23}\pi_3 \\ -y_{31}\pi_1 + y_{31}\mu_{31} = -Y_{33}\pi_3 + y_{32}\pi_2 \end{cases}$

Slack bus at bus 2:

$\begin{cases} Y_{11}\pi_1 - y_{13}\mu_{31} = y_{12}\pi_2 + y_{13}\pi_3 \\ -y_{31}\pi_1 + y_{31}\mu_{31} = -Y_{33}\pi_3 + y_{32}\pi_2 \end{cases}$

6.14 $K = \{1, 2, 3\}$; $U = \emptyset$; μ_{31} and μ_{21} are unknown.

Choose bus 1 as the slack.

$\begin{cases} y_{21}\mu_{21} = -Y_{22}\pi_2 + y_{21}\pi_1 + y_{23}\pi_3 \\ y_{31}\mu_{31} = -Y_{33}\pi_3 + y_{31}\pi_1 + y_{32}\pi_2 \end{cases}$

$\begin{cases} \mu_{21} = 1.67\,\$/\text{MWh} \\ \mu_{31} = 7.00\,\$/\text{MWh} \end{cases}$

6.16 62.5 MW of flowgate rights on branch 3-1.

6.17 37.5 MW of flowgate rights on branch 2-1 and 62.5 MW of flowgate rights on branch 3-1.

Chapter 7

7.1 12.14%; 32.28 $/MWh

7.2 11.17%

7.3 14.13%; 12.33%;

7.4 Yes, because the IRR is 12.49%.

7.5 The investment is higher if technology A is adopted, but the Incremental Internal Rate of Return on the additional investment is 14.13%, which is higher than the Minimum Acceptable Rate of Return.

7.6 The plant should continue operating because it continues to generate a positive cash flow of $32 524 128 per year. Borduria Power would not have built the plant because it would not have achieved its MARR.

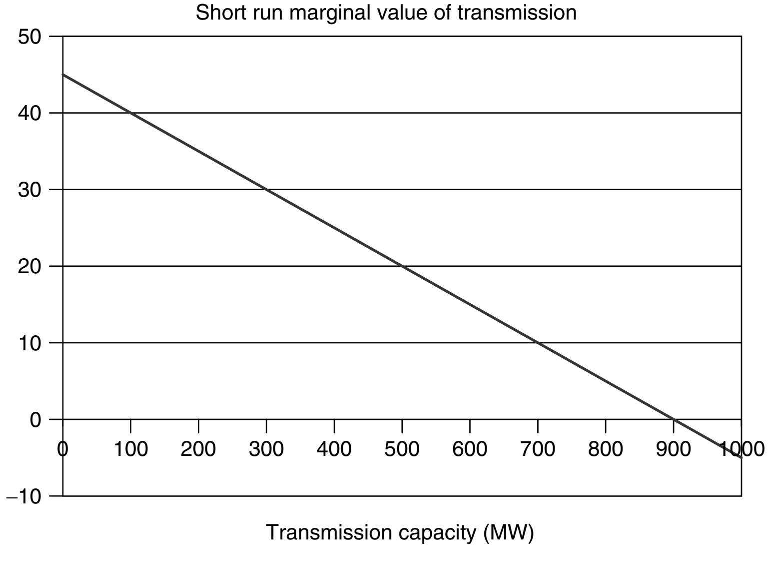
7.7 If the plant has 20 years of expected life left, Borduria Power should repair it because the Internal Rate of Return on the investment required for the repair is 12.17%, which is above the MARR used by the company. If the plant has only 15 years left, the IRR is only 10.51% and the plant should be closed down.

7.8 Minimum price: 78.80 $/MWh. Average production cost: 37.70 $/MWh.

Chapter 8

8.3 8.00 $/MWh

8.4



8.5 $\pi_T = 45 - 0.05 \cdot F_{\text{BA}}$
8.6 12.00 \$/MWh
8.7 660 MW
8.8 750 MW; \$78 750 000. The two amounts are identical.
8.9 500 MW; \$90 000 000 versus \$52 500 000
1000 MW; \$35 000 000 versus \$105 000 000

Abbreviations and acronyms

AFC	Average fixed cost
AVC	Average variable cost
CCGT	Combined cycle gas turbine
Disco (or distco)	Distribution company
FGRs	Flowgate rights
FTRs	Financial transmission rights
Genco	Generating company
IIRR	Incremental Internal Rate of Return
IPP	Independent power producer
IRR	Internal Rate of Return
ISO	Independent system operator
ITC	Independent transmission company
KCL	Kirchoff's current law
KKT	Karush-Kuhn-Tucker
KVL	Kirchhoff's voltage law
LRMC	Long run marginal cost
MARR	Minimum Acceptable Rate of Return
MES	Minimum efficient size
MO	Market operator
OCGT	Open cycle gas turbine
PTDF	Power transfer distribution factor
SMP	System marginal price
SO	System operator
SRMC	Short run marginal cost
Transco	Transmission company
VOLL	Value of lost load

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4

Index

Fundamentals of Power System Economics Daniel Kirschen and Goran Strbac
 ISBN: 0-470-84572-4