Springer Texts in Business and Economics

Christoph Weber
Dominik Möst
Wolf Fichtner

Economics of Power Systems

Fundamentals for Sustainable Energy

MOREMEDIA

Springer

Springer Texts in Business and Economics

Springer Texts in Business and Economics (STBE) delivers high-quality instructional content for undergraduates and graduates in all areas of Business/Management Science and Economics. The series is comprised of self-contained books with a broad and comprehensive coverage that are suitable for class as well as for individual self-study. All texts are authored by established experts in their fields and offer a solid methodological background, often accompanied by problems and exercises.

Christoph Weber · Dominik Möst ·
Wolf Fichtner

Economics of Power Systems

Fundamentals for Sustainable Energy

Springer

Christoph Weber
University of Duisburg-Essen
Essen, Germany

Dominik Möst
TU Dresden
Dresden, Germany

Wolf Fichtner
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany

ISSN 2192-4333 ISSN 2192-4341 (electronic)
Springer Texts in Business and Economics
ISBN 978-3-030-97769-6 ISBN 978-3-030-97770-2 (eBook)
https://doi.org/10.1007/978-3-030-97770-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Electricity and, more generally, energy systems are complex, multifaceted systems that may only be grasped and adequately analysed by combining multiple interdisciplinary skills. This has been our experience over many years of research in and teaching these topics at universities in Germany, be it in Karlsruhe, Dresden or Essen. But these interdisciplinary and intercultural challenges also popped up when giving lectures elsewhere, when discussing with practitioners and colleagues at national and international conferences and in multiple projects. So we have started at some point to think about writing a textbook to compile and consolidate our knowledge of this challenging field. Our objective has been to provide a structured guide to the subject, but this has also required that we commonly agree upon a structure and a selection of materials to be included in such a book. This has been a longer process than we initially envisaged. But the process of writing this book has also helped us deepen our shared understanding of the core concepts of this interdisciplinary research area that is labelled energy systems analysis or energy economics.

The research area is highly topical and likely to remain so over the coming decades, given the global challenge of climate change. As climate change is largely driven by CO_2 emissions related to the combustion of fossil fuels, we are convinced that the energy carrier electricity will gain more and more importance. The use of CO_2-free electricity will help to decarbonise sectors like transport and heating that so far primarily rely on fossil fuels. Environmental effects caused by energy conversion, e.g. climate change, have been a strong motivation for us to enter the field years ago, and it is a major driver for our current research and teaching. Yet at the same time, the rapid pace and shifting centres of interest in the political and academic debate pose a challenge to those trying to consolidate the knowledge in the form of a textbook. As we progressed, we concluded that it is essential to lay solid foundations, even if this means that not all subjects treated may sound topical for the reader from the outset. While finalising the book, we have experienced that the Russian war in Ukraine has lead to major impacts on the European energy markets. These tend to provide additional evidence for the principles outlined in the book. even if we were not able at this stage to consider all the effects in detail. But with a solid basis laid, we hope that our students and the readers of this book will be capable to reach out further and contribute their part to the shaping of future, more sustainable electricity systems. Based on our teaching experiences at Bachelor, Master and PhD levels, we aim to provide a balanced mix of conceptual discussions, mathematical formulations and practical applications. Hence the core target group of our textbook are students at both Bachelor and Master levels. Especially the more quantitative parts about modelling and risk management might yet also be of interest for students at PhD level and other researchers. At the same time, the authors hope that some of today's experts, including practitioners, may find this helpful book to enlarge their view on overarching questions regarding electricity systems—given that we all are experts only in limited areas.

This work would not have been possible without the continuous collaboration and exchange within our research groups and beyond. The work of many has left its imprint in this book, even if the remaining ambiguities are uniquely to be attributed to the authors. We are particularly thankful to our colleagues Russell McKenna, Michael Bucksteeg, Peter Lund and Ramteen Sioshansi for reviewing the draft of the manuscript and providing many valuable suggestions. We also thank Marvin Lepper for his detailed work on the references, Wolfdieter Fichtner for proofreading as well as Martin Lieberwirth and Dominic Rosswag for their scrupulous work on the graphs and figures. But all of this would not have been possible without the loving and enduring support of our families. Without their patience and acceptance of many hours and days spent on this and other work, this book would never have seen the light of day.

Essen, Germany — Christoph Weber
Dresden, Germany — Dominik Möst
Karlsruhe, Germany — Wolf Fichtner

Contents

Symbols

Some symbols or variables have double or multiple assignments and therefore have different meanings depending on the chapter. However, since these symbols or variables are commonly used in this form in the respective domains, we have decided to accept multiple assignments instead of introducing uncommon symbols for well-known concepts

a	Radiative forcing (Chap. 6)
a	Parameter inverse demand function (reservation price) (Chap. 9)
a	Drift rate (average rate of change over time, Chap. 11)
A	Area (Chaps. 4 and 5)
A	Coefficients in the PTDF matrix (Chap. 7)
b, B	Susceptance, i.e. imaginary part of the admittance $\underline{Y}$ (Chaps. 5 and 7)
b	Parameter inverse demand function (Chap. 9)
b	Variance rate of a stochastic process (Chap. 11)
c	Speed of light (Sect. 4.1.1)
c	Per-unit cost (Sect. 4.2.2 onwards)
$c_i(y_i)$	Marginal costs
$\bar{c}_i(y_i)$	Average costs
c^{fix}	Per-unit fixed cost (including annualised investment) (Chap. 7)
c^{inv}	Per-unit investment expenditure (Chap. 4)
c^{var}	Per-unit variable cost (Chap. 4)
C	(Total) Cost or expenditure (Chap. 4 onwards)
C	Capacitance (Chap. 5)
C'	Marginal cost (Chap. 9)
cf	Capacity factor (Chap. 4)
d	Demand function with market share as explained variable (Chap. 9)
d	Dimensionless option value parameter (Chap. 11)
dx	Increment of a stochastic process (Chap. 11)
dz	Increment of a Wiener process (Chap. 11)
D	Demand (Chaps. 4 and 7)
E	Energy (Chap. 2 onwards)
ES	Export surplus (Chap. 7)
Ex	Exergy (Chap. 2)

$\mathbb{E}$	Expectation operator (Chap. 7 onwards)
f_x	Probability density function for a stochastic variable x (Chap. 11)
F	Future price at time t for delivery in time T (Chap. 8 onwards)
F_x	Cumulative distribution function for a stochastic variable x (Chap. 11)
fh	Full-load hours (Chap. 2 onwards)
g	Gravitational acceleration (Chap. 4)
g, G	Conductance, real part of admittance $\underline{Y}$ (Chap. 5)
g	Calibration factor for future hourly prices (Chap. 11)
G	Internal energy of a thermodynamic system (Chap. 2)
H	Enthalpy of a thermodynamic system (Chap. 2)
H	Total number of hours (Chap. 4)
h	Height (e.g. head of water) (Chap. 4 onwards)
HHI	Hirshman–Herfindahl index (Chap. 9)
i	Inflow (Chap. 4)
I	Current in DC circuits or magnitude of current (Chap. 5)
$\underline{I}$	Current (complex value) in AC circuits (Chap. 5)
ID	Import demand (Chap. 7)
j	Imaginary unit for complex number calculus (Chap. 5)
k	Exponent in the Hellmann approach for wind speed estimation (Chap. 4)
k	Proportionality constant in Ramsey pricing (Chap. 5)
K	Capacity (Chap. 4 onwards)
K	Capital employed (Chap. 6)
l	Length (Chap. 5)
L	Storage filling level (Chaps. 4 and 7)
L	Inductance (Chap. 5)
$LCOE$	Levelized cost of electricity (Chap. 4)
LI	Lerner index (Chap. 9)
m	Mass (Chap. 4)
m	Market share (Chap. 9)
$\dot{m}$	Mass flow (Chap. 4)
$m(\tau)$	Mapping function (Chap. 11)
N	Number of elements (e.g. nodes) (Chap. 5 onwards)
o	Operation status variable (on/off) of a power plant (Chap. 4)
O	Operational expenditure (Chap. 6)
p	Pressure (Chap. 2 and Sect. 4.1)
p	Price (Chap. 3 and Sect. 4.2 onwards)
P	Power and power flows, in AC circuits (Chap. 5 onwards): active power
q	Quantity (Chap. 2 onwards)
Q	Heat energy (Chaps. 2 and 4)
Q	Reactive power (Chap. 5)
r	Interest rate or rate of return (Chap. 4 onwards)
R	Resistance, in AC circuits real part of impedance $\underline{Z}$ (Chap. 5)
R	(Allowable) Revenues (Chap. 6)
R	Reaction function (Chap. 9)
RPI	Retail price index (Chap. 6)

RSI	Residual supplier index (Chap. 9)
s	Plant start variable (Chap. 4)
s	Supply function with market share as explained variable (Chap. 9)
S	Entropy (Chap. 2)
$\underline{S}$	Apparent power (complex number) in AC circuits (Chap. 5)
S	Supply (Chaps. 6 and 7)
S	Stock or other spot price (Chap. 8 onwards)
t	Time (Chap. 2 onwards)
t	Transfer coefficient (Chap. 6)
T	Temperature in kelvin (Chap. 2)
T	Time of delivery or maturity (Chap. 8 onwards)
$\tilde{T}$	Time interval or set of time steps (Chap. 11)
u	Stochastic disturbance term (Chap. 9)
U	Set of units (Chap. 7)
v	(Wind) Speed (Chap. 4)
V	Volume (Chaps. 2 and 4)
V	Voltage in DC circuits or voltage magnitude (Chap. 5)
$\underline{V}$	Voltage (complex number) in AC circuits (Chap. 5)
V	Value (of a derivative product, Chap. 7 onwards)
w	Infeed relative to installed capacity for variable renewables (Chap. 4)
W	Work (Chap. 2)
W	Welfare (Chap. 6 onwards)
x	General variable or observation (Chap. 4, Chap. 7)
x	Stochastic process variable (Chap. 11)
$\widehat{x}$	General model output (Chap. 7)
X	Reactance, i.e. imaginary part of impedance $\underline{Z}$ (Chap. 5)
X	Expected annual productivity gain (Chap. 6)
X	Strike price of an option (Chap. 8 onwards)
y	Power plant output (Chap. 4 and 7)
y	Retail product supply (Chap. 9)
$\underline{Y}$	Admittance (complex number), inverse of the impedance (Chap. 5)
Y	Total (fixed) supply of retail products (Chap. 9)
z	Stochastic process variable (Chap. 11)
$\underline{Z}$	Impedance, also called complex resistance (Chap. 5)

Indices

c	Countries (Chap. 6)
i	Products (Chap. 6) or agents (Chap. 9)
l	Line (Chap. 5)
m	Node (Chap. 5)
n	Node (Chap. 5)
r	Region (Chap. 6 onwards)

s Market segment (Chap. 9)
s Time segment (Chap. 11)
t Time
T Delivery time (Chap. 8 onwards)
u Power plant (Chap. 4 onwards)

Greek Letters

α Weight of the incentive-based scheme in sliding-scale regulation (Chap. 6)
α Measure of relative attractiveness of retail product (Chap. 9)
β Measure of price sensitivity of customers (Chap. 9)
γ (Additional) Concentration of air pollutant or greenhouse gas (Chap. 6)
Γ Emission or immission threshold (Chap. 6)
Γ Nodal generation surplus (Chap. 7)
Δ Difference operator
ΔF Radiative forcing (Chap. 6)
Δt Time interval (Chap. 4 onwards)
Δx_t Sequence of steps
ε Price elasticity of demand (Chap. 3 onwards)
ε Emission intensity (Chap. 4 onwards)
ε Random variable (Chap. 11)
η (Conversion) Efficiency (Chap. 4)
Θ Voltage angle (Chap. 5)
κ Mean-reversion rate (Chap. 11)
λ Risk premium for avoiding spot market risk (Chap. 8)
Λ Set of lines (Chap. 7)
μ Mean value (Chap. 11)
ν Emissions (Chap. 6)
ξ Emission reduction rate (Chap. 6)
Ξ Absolute emission reduction (Chap. 6)
π Profit (Chap. 9)
ρ Density (Chap. 4)
ρ Electric resistivity of a conductor (Chap. 5)
σ Standard deviation (Chap. 11)
σ^2 Variance (Chap. 11)
τ Time
φ Phase angle (Chap. 5)
φ Availability (Chap. 9)
ϕ Probability density function of the standard normal distribution (Chap. 11)
Φ Cumulative distribution function of the standard normal distribution (Chap. 11)
Ψ Power flow limit (Chap. 7)
ω Angular velocity equal to $2\pi f$, with f frequency (Chap. 5)
Ω Information set (Chap. 8 onwards)

List of Figures

List of Tables

1 Introduction

Electricity and, more generally, all forms of energy are essential resources for modern life. Almost all our daily activities rely on the utilisation of energy. At the same time, energy transformation and usage are linked to some of the most challenging issues of modern societies. Notably, the continued combustion of fossil fuels, frequently for electricity production, is responsible for the largest part of so-called greenhouse gas emissions, which cause climate change. This contrasts with the political objectives as stated in the Paris Agreement within the United Nations Framework Convention on Climate Change (UNFCCC). In this agreement, the central objective has been set to combat climate change and limit the rise of global warming to well below 2 °C, preferably to at most 1.5 °C. To reach this objective, rapid and massive reductions of greenhouse gas emissions have to be realised, and the current energy systems must be transformed to (almost) carbon neutrality. More and more countries as well as, e.g. Europe as a whole strive to become climate neutral by 2050 – illustrating the challenges especially the electricity systems worldwide are facing.

In a nutshell, these statements comprise the motivation for writing this book and at the same time the challenge faced when conceiving and writing it: energy issues are intertwined with a multitude of topics and are interdisciplinary by nature. And at the same time, countries worldwide are not only facing issues in the gradual further development of their electricity and energy systems. But rather the conventional electricity systems based on fossil-fired thermal power plants are challenged by the threat of global warming, and all supply alternatives—being renewables, nuclear or carbon capture and storage—are confronted with specific challenges and risks. Moreover, the state of knowledge both on conventional and new electricity systems is continuously evolving while citizens and stakeholders in modern societies set varying priorities and weight objectives differently. Experts therefore have to continuously review their assessments and place them in the context of the broader political and societal debates.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_1

In this context, this textbook aims to provide suitable foundations for analyses of upcoming electricity systems and for developing adequate solutions to the current challenges.

But what are then the essentials of energy and electricity economics to be included in just a few hundred pages? And how avoid compiling a heap of material which is both difficult to digest and prone to rapid obsolescence? Answering these questions required tough choices. The primary guideline for selecting the material has been to bring together the first principles of various disciplines with the actual questions in the field. The authors are convinced that only a sound knowledge of the basic concepts across different disciplines will provide scholars and practitioners in the field the necessary background to successfully tackle the challenges of the ongoing electricity system transformation. For example in view of assessing the role of wind energy in future electricity systems, key physical and meteorological aspects of wind energy have to be considered along with the physical and engineering characteristics of wind turbines. But also the cost of those turbines and their operation principles and profitability have to be analysed in the context of the overall electricity system, including competing and complementary technologies, grids and possibly storage. Operation and investment decisions will be influenced by regulatory settings and market structures – where economics provides a broad range of basic and more advanced concepts. Environmental and social sciences offer further relevant concepts for analysing impacts on ecosystems and humans as well as attitudes of stakeholders.

Yet when it comes to investigating the envisaged transformation of electricity systems, models of the electricity system capturing key economic and engineering characteristics and also major environmental and societal aspects provide a centrepiece for analysis. Therefore, a range of technoeconomic bottom-up models for investigating the economics of power systems and markets form the backbone of this book. Yet, they are complemented by an overview of the main technological, economic and environmental drivers for the development of power systems. And these models provide the basis to discuss and highlight the implications of policy and regulatory design choices for future sustainable power systems (cf. Fig. 1.1).

Hence, two main parts may be distinguished in the book: Chaps. 2–5 take a detailed look at the power system and its elements from a mostly engineering–economics perspective, i.e. focusing on system characteristics and costs. Chapters 6–11 complement this view from an economic perspective by addressing more the layers of regulation and markets and the resulting incentives for market participants. The present chapter and Chap. 12 provide the frame that hopefully contributes to understanding how the pieces of the puzzle play together.

To provide the general context for these analyses, Chap. 2, entitled **Fundamentals of Energy and Power Systems**, discusses physical and engineering basics of power and energy as well as the current economic and societal context. It also introduces to the reading of energy statistics by presenting the concept of the energy transformation chain and highlights what makes electricity different from other goods and services.

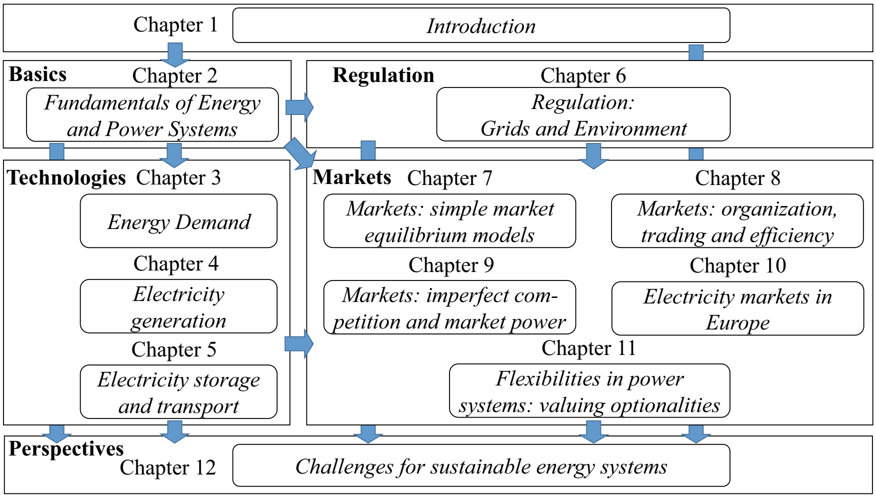


Fig. 1.1 Structure of the book

Chapter 3, **Energy Demand**, discusses an essential driver of any (energy) market – the importance of which was already pointed out by Adam Smith in his famous quote "Consumption is the sole end and purpose of all production" (Smith 1776, p. 660). The chapter discusses both decomposition of demand by applications and time (so-called load profiles) as well as demand-side management as a concept to overcome the traditional power system planning view of demand as a fixed "load". Besides electricity demand also heat demand is considered, as it is supposed to play an important role in strategies for decarbonisation. Demand from electric vehicles is also touched upon, although a full treatment of the link between transportation and energy usage is beyond the scope of this book.

The supply side of the electricity system is subsequently discussed in Chap. 4, **Electricity Generation and Operational Planning**. Both conventional and renewable technologies are discussed as supply options with their main technological, environmental and economic characteristics. This chapter also contains the formulation of a first power system model, which solves the problem of optimally scheduling several generation units to meet a pre-specified demand. This is also known as the unit commitment and dispatch model.

In Chap. 5, **Electricity Transport and Storage**, the so-far missing links between demand and supply are introduced: networks as means to transport electricity between locations in space and storage as means to transfer electrical energy from one point in time to another. This abstract functional characterisation is complemented in the chapter by detailed technological considerations, also including the principles of power flows through alternating current networks and principles of power system operation.

Chapter 6, **Regulations: Grids and Environment**, discusses those parts of governmental regulations of the power industry that are clearly complementary to electricity markets. So the focus here is on grid regulation and environmental regulation as domains, on which also mainstream economists agree that markets without government intervention are almost certain to provide inefficient results.

The subsequent chapters are devoted to electricity markets. In Chap. 7, **Simple Electricity Market Equilibrium Models**, the focus is on graphical and formal approaches to describe the matching of supply and demand in electricity markets. Thereby both short-term and long-term market equilibria are discussed, and the impacts of storage and transmission capacities on market outcomes are addressed.

Chapter 8, **Markets: Organisation, Trading and Efficiency**, deals first with institutional aspects and key design choices for electricity trading. Also, the link between the electronic trading business and the actual operation of the system is highlighted. Then the implications of well-functioning markets in terms of economic efficiency are discussed in a formalised way.

In Chap. 9, **Imperfect Competition and Market Power**, the focus is on the contrary on possible market imperfections. Thereby the monitoring of non-competitive market structures and corresponding behaviour as well as several modelling approaches are discussed.

As the last chapter focusing on market design, Chap. 10, **Electricity Markets in Europe**, deals with further details regarding power market architectures. These include notably the markets for so-called ancillary services like frequency regulation and the available approaches for grid congestion management. Also, a comparison between European and North American market designs is provided.

Chapter 11, **Valuing Flexibilities in Power Systems as Optionalities**, takes a somewhat different perspective on the interplay between generation as a core competitive business segment and the market. It considers dispatchable generation as a provider of flexibility in the market equilibrium. It then discusses approaches to determine the economic value of such flexibilities based on appropriate market price models, especially from finance and real options theory – which may obviously also be extended to storage and demand-side flexibilities.

Chapter 12, **Challenges for Sustainable Electricity Systems**, finally provides an overview on how the material and methods presented in the previous chapters may be used to answer major challenges of the ongoing transformation of electricity systems.

In all chapters, key learning objectives are defined at the beginning and questions for self-check as well as exercises are provided at the end to support self-study and application of the material. For selected models, also worksheets are provided in an electronic supplement to the book. We also provide suggestions for further reading on both basic concepts and more advanced topics at the end of each chapter. Within the text itself, we add references when specific issues are treated. But in line with the ambition to provide a readable, self-contained textbook, we do not refer in the text to other textbooks when basic concepts of economics, engineering or operations research are explained.

This volume may be used as a conventional textbook for studying the material consecutively in the presented order. In a problem-oriented learning approach, the reader may yet also start at the end and contemplate what elements are needed to analyse different problems faced in the transition towards more sustainable electricity systems. She or he may then take a more detailed look at those parts corresponding best to current personal interests.

The textbook as a whole hence provides an integrated, interdisciplinary perspective on electricity systems. For example, wind turbines and gas-fired generation units are dealt with their technological characteristics in Chap. 4, where some key cost and environmental impacts are compiled. Environmental aspects are further deepened in Chap. 6, whereas the economics of operation and investment are discussed in Chap. 7, and the related key challenges for future sustainable electricity systems are pointed out in Chap. 12.

The textbook may be used in a broad variety of courses for students with different disciplinary backgrounds, although it presumes knowledge of basic economic concepts like market equilibrium or investment calculus. We believe there are many valuable textbooks available for these topics so that we have not seen an added value in discussing these concepts in detail. Also basic knowledge of energy concepts from physics is advantageous, even though a summary of fundamental aspects is given in Chap. 2. Moreover, the book recurs to mathematical concepts of optimisation and statistics, where some background knowledge or some parallel reading is again advantageous.

The book has been designed for use in advanced Bachelor (undergraduate) and Master (graduate) courses. The material is diverse in multiple respects, including not only the disciplinary background from which it is taken but also the mathematical complexity. We believe that most of the material contained in Chaps. 2–5 may be used in an advanced Bachelor course except for Sect. 4.4.1.3 on optimal power plant scheduling and Sect. 5.1.3 on load flow in AC power grids. Chapter 6 may also be suitable for Bachelor courses except for the discussion of network pricing in Sect. 6.1.4. Chapters 7 and 8 may be used already at (advanced) Bachelor level, except for the section on information efficiency (Sect. 8.5). Chapter 9 dealing with imperfect competition is rather advanced material, except for the introductory Sect. 8.1 on indicators for market power. In Chap. 10, we would suggest leaving out the more detailed discussions of market aspects from Sect. 10.3 onwards of introductory courses at the Bachelor level. For Chaps. 11 and 12, we would in general recommend the use rather at Master or even at Ph.D. level, as current and new research trends are addressed, although a flavour of it may be given at Bachelor level, too. But obviously, the selection of material will depend on the background and interests of students as well as professors.

Reference

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations, 2 vol. R. H. Campbell & A. S. Skinner (Eds.), W. B. Todd (Textual Ed.) (1976). Oxford University Press.

2 Fundamentals of Energy and Power Systems

To understand energy economics and the fundamentals of energy and power systems, some basics are required. In this chapter, physical and engineering basics, including the laws of conservation and thermodynamics, the role of energy in economics and society, the energy transformation chain, aspects of resource availability as well as particularities of the electricity sector, are discussed. This introductory chapter aims at providing answers notably to the questions:

- What are the physical laws governing energy use?
- What are the critical societal and economic challenges regarding energy?
- What are key indicators for the energy system?

The chapter is primarily intended for readers who previously have had little contact with energy topics and are particularly interested in energy and power systems fundamentals. Subsequently, we first provide an overview of physical and engineering basics in Sect. 2.1 before discussing the role of energy for the economy and society as a whole in Sect. 2.2. We then discuss key challenges regarding energy use, namely resource availability and potential environmental damages in Sect. 2.3. Section 2.4 then introduces key concepts to measure the use of energy and its transformation in the context of national and regional energy balances. Finally, Sect. 2.5 highlights the specificities of electricity compared to other energy carriers.

Key Learning Objectives

After having gone through this chapter, you will be able to

- Describe key physical and engineering concepts for energy systems including the notions of energy and power and the fundamental laws of thermodynamics.
- Discuss key societal aspects of energy use and current challenges.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_2

- Describe the elements of the energy transformation chain and use energy balances.
- Point at key peculiarities of electricity compared to other energy carriers.

2.1 Physical and Engineering Basics

Phenomena related to energy are frequently observable in everyday life, be it natural phenomena like lightning or solar radiation or results of human engineering like the traction developed by a motor engine or the warmth provided by a gas heating system. Subsequently, we first look at the physical concept of energy and the related general physical concept of power in Sect. 2.1.1. We thereby also introduce thermodynamical systems as a useful scheme to analyse energy transfers and power flows. Building on these foundations, we give a non-formal introduction to the fundamental laws of thermodynamics in Sect. 2.1.2 as they are also a key for understanding energy systems. Finally, we address the limits of energy conversion in Sect. 2.1.3.

2.1.1 Energy and Power and Thermodynamic Systems

When investigating the physical concept of energy, one may note that energy is not directly observable or measurable. Instead, it is usually introduced by deriving it from the concept of **work**.

In mechanics as part of classical physics, **work** is **defined as the outcome of a force acting along the direction of displacement of an object**. The amount of work done on an object depends on the strength of the force exerted on the object and the distance the object moves. Consequently, work is the (vector) product of force and distance and is measured in joule in the International System of Units.

Energy is the **ability to do work (mechanical energy) or to provide heat (thermal energy).** Any time an object does work on another object, some of the energy of the working object is transferred to that object, raising its energy state. Like for work, the unit of measurement for energy is joule (abbreviated J).

But besides this mechanical energy, different forms of energy have been identified in physics. Without being exhaustive, we may name:

- **Mechanical (or kinetic) energy** is energy associated with the movement of an object, so mechanical energy may for example be contained in the movement of a car.
- **Chemical energy** is energy stored in the bonds linking atoms together in molecules and other chemical compounds. Examples are fossil fuels, food energy or the energy contained in plants/crops.

- **Thermal energy** is the internal energy of substances corresponding to the vibration and movement of atoms, molecules and ions within the substance. It becomes perceptible, e.g. in hot water, steam, etc., due to the corresponding temperature difference with the surroundings.
- **Electrical energy** is energy derived from electric potentials or contained in moving electrical charges, hence, electricity.
- **Radiant energy** is energy included in electromagnetic waves and can occur as visible light, ultraviolet, infrared, radio waves, microwave radiation, etc. An example is sunlight.
- **Nuclear energy** is energy stored in the nuclei of atoms, e.g. of uranium. Part of it may be released through nuclear fission in the case of uranium.
- **Gravitational (or potential) energy** is the potential energy associated with the gravitation force and hence depends on the position of a mass, such as stored water in a reservoir.

As energy can occur in different forms and is used in multiple contexts, there exist plenty of units (besides joule) to measure energy (see Table 2.1), e.g.:

- A calorie (cal) is defined as the energy needed to raise the temperature of 1 g of water by 1 °C (in the SI system, this is 1 K, cf. below), and 1 cal corresponds to 4.184 J.
- A British thermal unit (BTU) is defined as the energy needed to raise the temperature of 1 pound of water by 1 °F, and 1 BTU corresponds to 1.055 J.

Most other units for energy are similarly defined and take the particular characteristics of the object of interest into account. Table 2.1 gives an overview of different energy units and their conversion factors.

As stated before, energy is a measure of how much work can be done. This is valid without any consideration of how long it takes to accomplish the work. **Power** measures how rapidly the work is performed, so it relates work to time. Or, more precisely, **power** is **defined as work per time**. This is the general physical concept of power which is much broader than the frequent use of power as a synonym for electricity in the English language suggests.

The unit of power is joules per second or J/s, which corresponds to watt (W). The capacity of an electrical power plant is measured in power units, in general watt. Since watt is a relatively small unit, power is often measured with a SI prefix[1] in kilo-, mega- or giga-watt (kW, MW or GW). The output of a power plant is measured in energy units, in general joule, but in an electrical engineering context also often in watt-hours (Wh). This unit directly illustrates that energy is the product of power and time – or more generally, the integral of power over time. Since Wh

[1] The International System of Units (SI) also describes prefixes beside the description of coherent systems of units. Prefixes are added to unit names to produce multiples and sub-multiples of the original unit.

Table 2.1 Energy and power units and conversions

Symbol	Unit name	Usage	Conversion factor
J	Joule	SI unit for energy	1 J = 1 Nm = 1 W·s
kWh	Kilowatt-hour	Frequently used energy unit for electricity	1 kWh = 3600 kJ
toe	Tonne of oil equivalent	Commonly used units for energy in energy balances	1 toe = 41.868 GJ
tce	Tonne of coal equivalent		1 tce = 29.308 GJ
kcal	Kilo-calorie	Often used for energy content in food	1 kcal = 4.18 kJ
BTU	British thermal unit	Anglo-Saxon unit for energy	1 BTU = 1.055 J
W	Watt	SI unit of power, i.e. energy per time	1 W = 1 J/s

is also a relatively small unit concerning typical sizes, energy is also often measured in kilo-, mega-, giga- or even tera-watt-hours [kWh, MWh, GWh, TWh].

Before continuing, we consider two simple examples:

Example 1

The (minimum) energy required for driving a car up to a mountain top can be calculated based on the gravity force and the difference in altitude between the starting point and the top of the mountain. If the car has a mass of 1500 kg, it is subject to a gravity force of 1500 kg · 9.81 m/s^2 = 14,715 N. The value 9.81 m/s^2 thereby corresponds to the gravity constant at the earth's surface (it would be only one sixth on the surface of the moon…). The work performed by the motor engine to get the car on the top of the mountain is then the product of the gravity force and the (vertical) distance travelled, e.g. 14,715 N · 1000 m = 14,715,000 Nm = 14.715 MJ. Hence, the car will need a minimum of about 15 MJ of energy to get on top. In reality, it requires considerably more, since there are losses in the car engine, losses related to the wheels, the air resistance, etc.

Example 2

Heating one litre of water (approximately 1 kg) from 20 to 70 °C requires energy. It is known that raising the temperature of one kilogram of water by one degree (1 K in the SI system[2]) requires 4.18 kJ/(kg·K). This is also known as the specific heat capacity of water. Raising the temperature of one litre of water by 50 K, hence

[2] K stands for the unit kelvin. 0 °C corresponds to 273.16 K, 100 °C to 373.16 K. The temperature intervals on the kelvin scale are hence identical to those on the Celsius scale, yet the zero point is the absolute zero (−273.16 °C), cf. Sect. 2.1.2. By convention, temperature intervals should always be indicated in kelvin.

requires 1 kg · 50 K · 4.18 kJ/(kg·K) = 209 kJ of energy. This energy may be provided, e.g. by a gas stove. And again, the energy provided by the fuel will be larger than the useful energy due to losses in conversion and during heating. If the heating is done using a burner with a nominal power of 1 kW, a lossless heating up will require t seconds, with t given through the equality: 1 kW t = 209 kJ. Solving for t, we obtain t = 209 s ≈ 3.5 min.

The second example may serve as a starting point to introduce the concept of **thermodynamic systems**.[3] In general, a thermodynamic system may be any part of the universe that is subject to scrutiny. It is defined through its system boundary, which separates it from the surroundings (cf. Fig. 2.1). The system boundary may

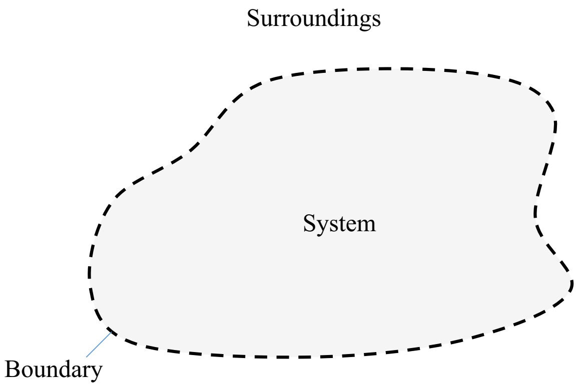


Fig. 2.1 Thermodynamic system

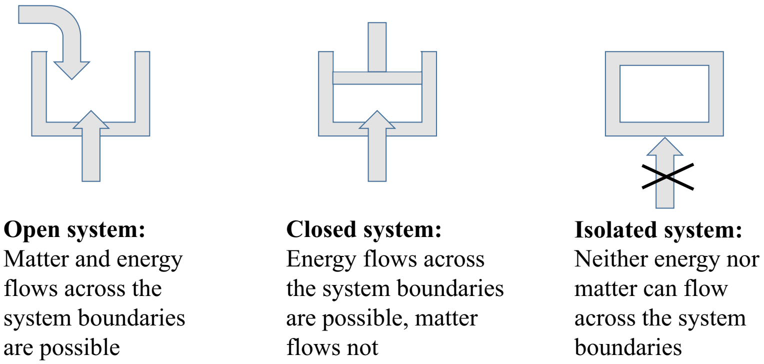


Fig. 2.2 Types of thermodynamic systems

[3] Thermodynamics is the term used to designate the branch of physics that deals with heat phenomena and in particular with all types of heat engines.

consist of solid walls, or it may just be described by a virtual boundary in space. Hence, the system boundary may be permeable to material flows and energy transfers or not. We thereby distinguish three types of systems (cf. Fig. 2.2):

- **Open systems**: both material and energy flows may cross the boundary.
- **Closed systems**: the system boundary prevents the exchange of matter yet provides possibilities to exchange heat, work or both.
- **Isolated systems**: these are characterised by the fact that neither material nor energy flows can cross the system boundary.[4]

A kettle filled with water is an example of an open system since the heat coming from a stove will penetrate the kettle and the water may leave the kettle as steam. The water circulating in the pipes of a heating system is yet rather a closed system – it will be heated in the boiler and transfer the heat through the radiators to different rooms. Yet, usually, water or other substances will neither enter nor leave the system. A well-insulated box without openings is finally an example of an (almost) isolated system.

2.1.2 Laws of Thermodynamics

The concept of energy is closely linked to one basic rule of thermodynamics, the **law of conservation of energy**, also called the first law of thermodynamics. It is part of a series of basic physical laws that describe the behaviour of energy systems.

Zeroth law of thermodynamics: If two systems are each in thermal equilibrium with a third, they are also in thermal equilibrium with each other.

This law justifies the use of thermometers: if two systems are at the same temperature, they will heat the thermometer (which is the third system) to the same temperature that hence may be measured.

First law of thermodynamics: The total energy of an isolated system is constant.

This **law of conservation of energy** expresses the theoretical postulate that energy is a conserved quantity. One form of energy may be transformed into another one, for example, kinetic energy into thermal energy. In addition, energy can be transported out of a system or into a system, but it is not possible to generate or destroy energy. In the physical sense of the law of conservation of energy, a "loss" of energy is not possible.

Second law of thermodynamics: Heat cannot spontaneously flow from a colder location to a hotter one.

The second law of thermodynamics makes statements about the direction of processes and the principle of irreversibility (see next section). In consequence, all forms of energy lose their ability to provide high-quality (mechanical) work. In this

[4] Note that sometimes the term "closed system" is used for isolated systems as defined here. As this lends to confusion, we prefer the definitions used here.

sense, from an economic-technical point of view, there is energy consumption (but, it should be correctly called "exergy consumption", based on the concept of exergy explained in the next section).

Third law of thermodynamics: As a system approaches absolute zero temperature (0 K), all processes cease.

The so-called entropy of the system, a measure of its disorder, then approaches a minimum value. This law hence provides an absolute reference point for the determination of entropy.

As the energy of an isolated system, i.e. a system without mass and heat transfers across its boundaries, remains constant, the **energy balance** of an open system may be described as difference of its input and output energy and material flows. For a heat engine that transforms high-temperature heat Q_I into mechanical work W and low-temperature heat Q_O, this may be written in stationary operation (cf. Fig. 2.3):

$$Q_I = W + Q_O \tag{2.1}$$

So-called state variables may describe the state of a thermodynamic system. Temperature is one of them, as are mass, pressure and volume. In terms of energy conversion, enthalpy and entropy are the most important state variables and will be discussed in the next section along with key implications of the laws of thermodynamics.

2.1.3 Thermodynamic State Variables, Energy Transformation and Carnot Efficiency

The term **enthalpy** describes a property of a thermodynamic system, which is equal to the system's internal energy plus the product of its pressure and volume. The enthalpy of a thermodynamic system is defined as

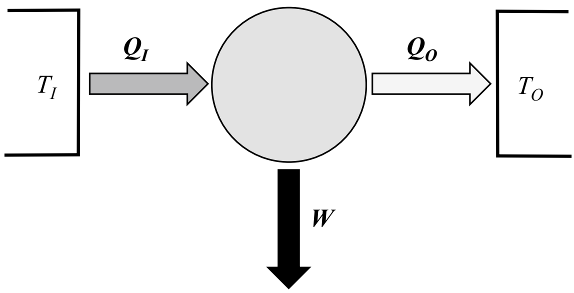


Fig. 2.3 Schematic representation of a heat engine with heat and work flows

$$H = G + pV \tag{2.2}$$

with H as enthalpy, G as internal energy of the system, p as pressure and V as volume of the system. The internal energy of the system G is not directly observable. But based on the 1st law of thermodynamics, it may be derived from the thermodynamical processes required to move the system from a reference state to the current state of interest.

The term **entropy** is closely linked to the 2nd law of thermodynamics. It is defined as a thermodynamic quantity representing the unavailability of a system's thermal energy for conversion into mechanical work. It is often interpreted as the degree of disorder or randomness in the system. This state variable entropy describes an extensive property of a thermodynamic system and is never decreasing in an isolated system. As stated above, an isolated system is a system where there is no energy or matter exchange with the environment.

In consequence, the entropy of an isolated system will tend to increase over time, approaching a maximum value at equilibrium. Supply of heat or matter causes an increase in the entropy present in a system, as well as any spontaneous processes within the system, such as mixing, thermal conduction or chemical reaction or conversion of mechanical energy into thermal energy by friction. Processes that produce entropy are therefore termed irreversible. The 2nd law of thermodynamics is hence a law of "entropy increase". It enables quantifying the reduction in the capacity of a system for change and determines whether a thermodynamic process may occur. "Irreversibility" is central to the understanding of entropy. Irreversibility can easily be illustrated by the example of putting milk into coffee. There is no way (without using machinery and a lot of energy) to get the milk to jump out of the coffee back into the milk bottle. Pouring milk into the coffee is thus an irreversible process that increases the entropy in the system. In this sense, irreversible means that a process cannot happen in reverse. For any process in an isolated system, the thermodynamic state variable entropy is never decreasing; for an irreversible process, it is strictly increasing.

Hence, entropy is a non-decreasing state variable in an isolated system, whereas energy is conserved. Accordingly, energy cannot be consumed. Similarly, it is impossible to produce or generate energy out of nothing. Only one form of energy may be transformed into one (or several) others. E.g. the chemical energy contained in natural gas may be transformed in a power plant into electricity and heat. Nevertheless, "energy consumption", "energy waste", "energy saving" and "energy loss" are typical terms used when speaking about energy.

When economists and engineers recurrently use terms like "energy consumption", they implicitly refer to the fact that only some forms of energy are entirely usable, i.e. can deliver work, while other forms cannot. The terms "energy consumption", etc. thus refer to the transition from technically or biologically usable forms of energy (exergy) into unusable forms (anergy), which are related to the 2nd law of thermodynamics.

Notably, **exergy** is defined as the maximum useful work obtainable from a source of energy through a process interacting with the environment. It is sometimes also labelled by the synonym technical working capacity. Correspondingly, **anergy** is the part of an energy potential that cannot deliver any work in a process. Basically, the following relation energy = exergy + anergy applies.

Exergy can be calculated at given ambient conditions U from the thermodynamic quantities of enthalpy H, entropy S and temperature T (on the absolute Kelvin scale)[5] with the help of the following formula:

$$Ex = H - H_U - T_U(S - S_U) \tag{2.3}$$

Consequently, different forms of energy can have a different potential to perform work. Thermal energy at ambient temperature level cannot deliver any work and is thus 100% anergy. The transformation of heat into work is not entirely possible and is in general associated with losses. Finally, note that exergy is not a state variable but depends on the ambient state.

To illustrate the concepts of exergy and anergy, the example of the heat engine introduced in Fig. 2.3 is elaborated on subsequently. Thereby, the following values are used to characterise the ambient condition: T_U = 288 K (corresponding to 15 °C), p_U = 1.013 bar; H_U = 63 kJ/kg; S_U = 0.224 kJ/kg K. The values in Table 2.2 show the different exergy content of water and steam at different temperatures and pressures. As can be seen from the numbers, water at ambient temperature has an exergy of 0 kJ/kg, while the highest pressure and temperature show the highest exergetic values in the table.

While energy cannot be generated or produced, the generation of electrical energy is possible as it is a particular form of energy and may hence be obtained by transforming other energy forms. Electrical and mechanical energy are pure exergy, while the exergy of heat energy depends on the temperature level. High-temperature heat at temperatures well above the ambient temperature can be converted with relatively good efficiency by suitable heat engines into mechanical energy. The **efficiency** of a heat engine is thereby generally defined as the ratio of the usable mechanical output W to the heat input Q_I:

$$\eta = \frac{W}{Q_I}. \tag{2.4}$$

The maximum physical efficiency of this conversion is called **Carnot efficiency**. Carnot's efficiency, often also called Carnot's rule, is a principle that refers to an idealised heat engine operating with a so-called ideal gas and specifies limits on the maximum efficiency of this idealised process. The efficiency of such a Carnot engine is a function of the temperatures of its hot (input) and cold (output) reservoirs. The formula for this maximum efficiency is

[5] Cf. footnote 2, p. 10.

Table 2.2 Enthalpy, entropy and exergy for water and steam at different pressures and temperatures[a]

Pressure (bar)	Temperature (°C)	Temperature (K)	Enthalpy (kJ/kg)	Entropy (kJ/kg K)	Exergy (calculated) (kJ/kg)
250	500	773	3166	5.96	1450
30	370	643	3163	6.82	1200
12	330	603	3111	7.14	1055
1.013	90	363	377	1.19	36
1.013	15	288	63	0.22	0

[a] Steam values for enthalpy and entropy are taken from steam tables.

$$\eta_{\max} = \eta_{\text{Carnot}} = 1 - \frac{T_U}{T_I} = \frac{T_I - T_U}{T_I}. \tag{2.5}$$

where T_U is the absolute temperature of the cold reservoir (ambient temperature) and T_I is the absolute temperature of the hot reservoir. From the formula, it is obvious that a higher absolute temperature of the hot reservoir increases the Carnot efficiency (when keeping the absolute temperature of the cold reservoir constant).

Note that the concept of efficiency may also be generalised to further types of processes or even entire technical systems such as power plants or motor engines. **Efficiency** then refers to the ratio of the useful output power flow to the corresponding input power flow:

$$\eta = \frac{P_O}{P_I}. \tag{2.6}$$

The efficiency in practical operation depends on the actual operation setpoint. Notably, part-load operation often comes along with lower efficiency (cf. Chap. 4). Additionally, it may also be helpful to define annual efficiencies (or efficiencies for other time periods) based on the output and input energy (work or heat) quantities:

$$\eta_{\text{ann}} = \frac{E_O}{E_I}. \tag{2.7}$$

Another issue arises when defining efficiencies related to combustion processes. Here, the energy content of the corresponding fuel has to be appropriately specified. This is done using a substance's heating value, which indicates the amount of heat (energy) per mass unit released during the combustion. However, the heat released during combustion may be measured in two ways. Depending on how much the combustion products are cooled down and whether compounds like water (H_2O) are condensed, two values are distinguished:

- the **lower** heating value, and
- the **upper heating value**.

Both measure the chemical energy that is set free in the combustion process. Yet, the combustion process also leads to the production of water (cf. Chap. 4), and the two indicators differ in their consideration of the condensation of vapour to liquid water. The condensing energy is included in the upper heating value, whereas it is not part of the lower heating value. The difference between the two values is particularly relevant in the case of natural gas. As it consists primarily of methane (CH_4), the proportion of formed water is exceptionally high and the upper heating value is approximately 10% higher than the lower.

Note that efficiencies of thermal processes are usually given relative to the lower heating value, whereas prices are mostly given relative to the upper heating value. Heating values of selected fuels are depicted in Table 2.3.

When it comes to the usage of a technical device related to energy, there is another important concept: the **full-load hours** are obtained by dividing the produced or consumed energy of the device by its rated capacity. It corresponds to the number of hours that the generation unit in question theoretically has to be operated at full capacity to achieve this annual energy yield. While the full-load hours are often used in Europe to characterise generation or consumption technologies, in the US, typically, the **capacity factor** is specified. The annual capacity factor is equal to the full-load hours divided by 8760 h (the number of hours per year).

The concept of full-load hours (or capacity factors) makes it possible to compare different technologies or also different locations or operation years for a technology under consideration (e.g. wind turbines) by standardising the capacity. The case of a

Table 2.3 Upper and lower heating values of selected fuels at 25 °C

Fuel	Upper heating value (MJ/kg)	Lower heating value (MJ/kg)
Hydrogen	141.7	120.0
Methane	55.5	50.0
Ethane	51.9	47.8
Propane	50.4	46.4
Butane	49.1	45.3
Paraffin wax	46.0	41.5
Kerosene	46.2	43.0
Diesel	45.6	42.6
Biodiesel (methyl ester)	40.2	37.5
Petrol (gasoline)	46.4	43.4
Coal	32.5	31.0
Lignite	15.0	
Wood (depends on drying and type)	16.2	15.4
Peat (dry)	17.0	

Note Variations in quality may induce heating values changes within a range of 5–10% around the given value.
Source McAllister (2011) and TheEngineeringToolbox (2021)

wind turbine may serve as illustration: the annual average full-load hours of onshore plants are between 1500 and 3500 h depending on the location of the wind turbine and the corresponding wind speed distribution. Average full-load hours of all land-based wind turbines in Europe are roughly 2000 h. At sea, full-load hours range typically between 3500 and 4000 h. For heating systems, also full-load hours may be defined: as the heating is operating at full power only during periods of very harsh cold, the annual full-load hours will be in the range of 1000 h even in moderate European climates. The system will be in operation during most of autumn, winter and spring. Yet, the delivered heat will during most periods correspond only to a fraction of the maximum output.

2.2 Energy, Economy and Society

Energy is not only a physical concept and an engineering discipline. It is an ingredient of everyday life, essential to any form of human activity. This is true in the literal sense: human life is based on the transformation of energy contained in nutrition (i.e. carbohydrates, fats and proteins) into storable and usable forms of energy in our body. And, life makes use of this energy for both physical and mental activities. But, energy is essential for modern life also in a more general sense: most of our daily activities do not only involve the use of our own energy but also the use of additional energy provided by techno-economic-societal systems. Think of reading a book in the evening with artificial light, heating and cooling of our homes, commuting in cars and trains, using smartphones and other digital devices, etc. Also, the production of goods and services involves the use of energy, whether for the manufacturing of a car or the provision of internet and cloud storage services. Hence, energy is a consumption good and a production factor, which may or may not be substituted by other production factors like capital or labour. Similar to labour and raw materials, energy is a consumable factor of production – unlike machines and buildings, which are potential factors that are not immediately consumed in the production process.

In a historical context, one may even argue that the development of modern societies is closely linked to increasing use of natural energy resources for human purposes, as illustrated in Fig. 2.4. Obviously, the development of human societies has come along with an increased use of energy sources beyond food, starting with firewood and moving along with coal, oil, natural gas, nuclear energy and renewables. Over time, the efficiency of conversion processes has increased considerably, but the amount of "energy services" (cf. Sect. 2.4) has augmented even more substantially.

Even if we limit ourselves to a comparison of modern societies, we see striking differences in the per capita consumption of energy as illustrated in Fig. 2.5. There is a substantial correlation between energy consumption per capita and economic development as measured by gross domestic product (GDP) per capita. But, a comparison among industrialised countries (OECD members) reveals that energy use may vary considerably even at similar GDP levels.

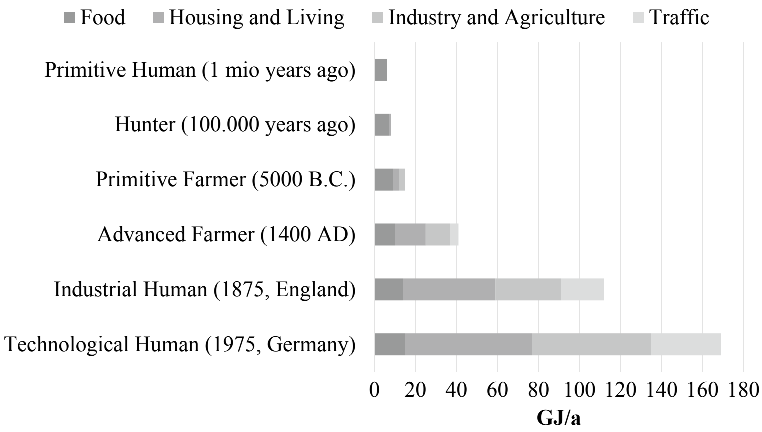


Fig. 2.4 Use of energy at different development stages of human societies. *Source* Own illustration based on Voß (2003)

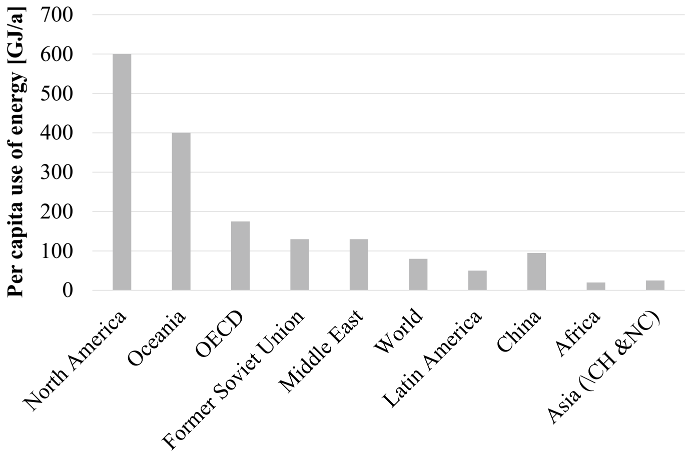


Fig. 2.5 Per capita use of energy in different world regions. *Source* Own illustration based on data from IEA (2020)

A look back over more than four decades shows that primary energy supply in the industrialised countries (OECD countries) has increased by about 41% in the period from 1973 to 2016 (cf. Fig. 2.6). Note that the concept of primary energy and the corresponding measurement conventions are discussed in detail in Sect. 2.4.1. Over the same period, the GDP in the OECD countries has increased by almost 180% in real (deflated) terms – i.e. the use of energy has increased much slower than economic wealth, or put differently: energy consumption almost decoupled from economic growth.

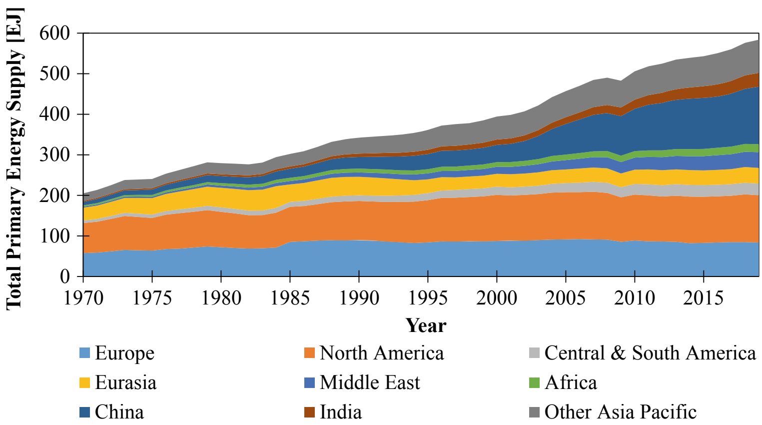


Fig. 2.6 Development of world total primary energy supply by regions from 1971 to 2019. *Source* Own illustration based on data from bp (2020)

This decoupling from economic growth is not true for other parts of the world. During the same period, energy use in the other countries of the world has risen by about 280%. As Fig. 2.6 illustrates, a large part of that increase has occurred in China. But also, other emerging economies, notably in Asia, have contributed to that growth. China has now a per capita energy consumption that exceeds the world average.

To interpret these numbers, it is noteworthy that the world population has also grown over the same period from 3.8 billion to 7.3 billion persons, again mostly in non-industrialised countries. Expressed in per capita terms, the growth in energy consumption is hence less pronounced. The OECD countries only saw their per capita energy consumption grow by three per cent from 1973 to 2016, from 170 to 176 GJ/cap. In the non-OECD countries, the average increased from 33 to 58 GJ/cap over the same period, i.e. an increase of about 80%.

IEA (2017, p. 21) provides a decomposition of the changes in energy consumption between the years 2000 and 2016. Accordingly, energy efficiency improvements in the order of 13% have offset the effect of economic activity growth during that period in the OECD countries. Effects of economic and other structural changes have been of minor importance. For non-OECD countries, improvements in energy efficiency have been of similar magnitude, yet the growth effect has been much higher, giving rise to a substantial increase in energy use. Hence, we have seen a decoupling of energy demand growth from economic growth in industrial countries since the year 2000, yet energy efficiency improvements have not been sufficient to compensate for the higher economic growth in developing countries and emerging economies.

Focusing on electricity, it is first to be noted that electricity is not a primary energy source but a secondary energy carrier (cf. Sect. 2.4). Its use has grown more rapidly than overall energy use over the past decades. In the industrial countries,

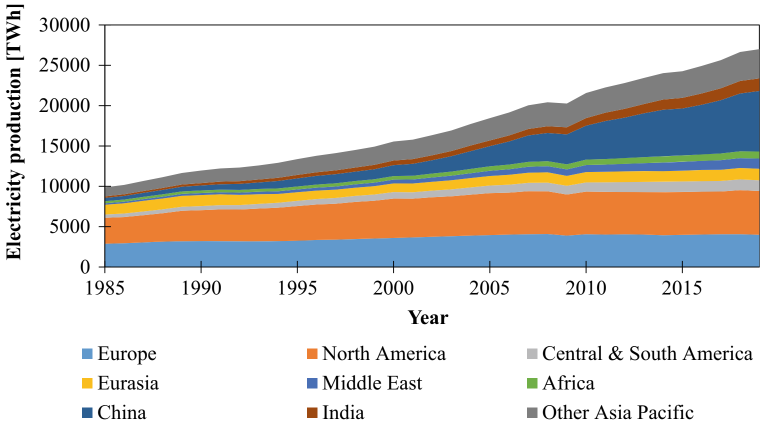


Fig. 2.7 Development of world electricity production by regions from 1985 to 2019. *Source* Own illustration based on data from bp (2020)

electricity generation surged from 4.5 PWh in 1973 to 11.0 PWh in 2016; in the other countries, the increase was even from 1.7 to 14.0 PWh, i.e. by more than a factor 8 (cf. Fig. 2.7). The development of modern societies is hence closely linked to the use of electricity. This is a consequence of multiple electricity applications like household appliances, electric drives, information and communication technologies, etc. and their increased penetration around the world.

For overall energy use, we may conclude that a decoupling of economic growth and energy use is possible – at least on a per capita basis. Countries like Germany have seen their primary energy supply remain nearly unchanged over more than 40 years. For electricity, growth has been substantial even in industrialised countries. Only in the last decade, growth rates fell below the rate of population increase. And for countries outside Europe and North America, an extrapolation of past developments suggests that there is still substantial growth in energy consumption but especially in electricity use ahead. This is at least true if we assume that the developing world, including the emerging economies, aims to catch up with Western economies and tend to replicate - at least to a considerable degree – the Western way of life and its energy consumption patterns. Figure 2.5 then suggests that there is still a substantial increase of energy use to be expected in future years.

Figure 2.8 gives an overview of the type of primary energy carriers used at a global scale – both for total energy supply and electricity production. It is evident that to date, oil is the most important primary energy source when it comes to total supply – almost one-third of global primary energy use is based on oil. Coal with about 27% and natural gas with 22% come next – meaning that more than 80% of the global energy usage is based on fossil fuels. The implications of this, both in terms of resource availability and environmental problems, are discussed in the following subsection.

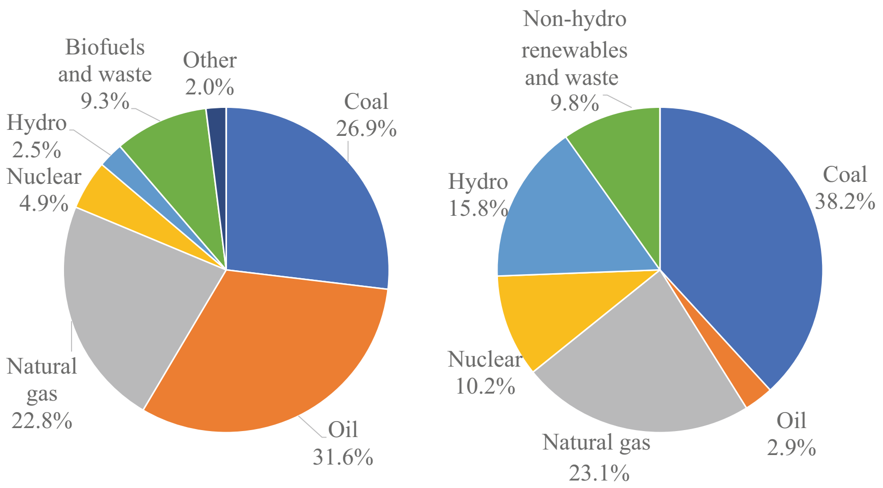


Fig. 2.8 World primary energy supply (left) and world electricity production (right) by primary energy source in 2018. *Source* Own illustration based on data from IEA (2020)

For electricity, the situation is somewhat different: here, oil plays only a very minor role – oil mostly goes into the transport sector and is also partly used in the residential, commercial and industrial sectors. Yet, coal is used to produce almost 40% of world electricity, followed by natural gas and hydro. Both nuclear and non-hydro renewables only account for about 10% of the global electricity consumption.

2.3 Challenges of Resource Availability and Environmental Damage

The previously described developments in energy use lead to two challenges: resource availability and environmental problems. These are not discussed in full depth here. Yet, some key elements of analysis are given – in particular, environmental issues and corresponding technological and political responses are discussed in more detail in Sect. 6.2.

2.3.1 Resource Availability

In terms of resource availability, it is crucial to distinguish between exhaustible and renewable resources.

Exhaustible or **finite resources** are resource stocks that are depleted by human activity and are not replaced by natural means at a pace that keeps up with consumption. **Renewable resources** are by contrast available over periods beyond human time scale (e.g. solar radiation) or replenished continuously by natural processes (e.g. biomass).

Exhaustible resources include fossil fuels and also nuclear resources. Given the finiteness of the planet earth, there is an upper limit to the availability of these resources. Yet, this ultimate resource constraint is unfortunately not known. On the one hand, human knowledge still has limitations regarding the details of the deposits available underneath the earth surface. On the other hand, human ingenuity itself moves the borders of the resource constraint. This has become very prominent with the so-called shale gas revolution. Through several advanced technologies, notably 3D seismology, fracturing and horizontal drilling, previously non-recoverable resources have become commercially recoverable from roughly 2009 onwards.

Therefore, several indicators for the availability of fossil and nuclear energy resources have been established:

Reserves refer to amounts of energy carriers where the corresponding geological sites have been proven and which can be extracted with today's technology and at current prices. This implies that a price increase induces a rise in identified reserves since higher prices allow a profitable extraction even of less favourable sites. And conversely, price decreases imply also a reduction in reserves.

Resources include sites that have been proven but are today either technologically or economically not extractable and sites that are considered as geologically possible and extractable in future.[6] So, this may be considered as the current knowledge of what may be ultimately available. The study underlying the data in Table 2.4 yet takes a rather conservative view and considers the potential economics of extraction when estimating the resources.

Production per year is given as a reference quantity to put reserves and resources into perspective.

The reserve-to-production ratio (short **R/P ratio**) is most commonly formed by dividing the reserves by the production of the latest available year. It provides a relative measure of the scarcity. Yet, interpretation has to be done with care. E.g. the R/P ratio of 55 years for oil does not necessarily mean that oil will be exhausted in 55 years from now. On the one hand, oil consumption and consequently also production may change in the years to come. On the other hand, resources may turn to reserves due to price changes, technological development or even by new discovery (in the case of unknown resources) (cf. above). Unproven or uneconomical resources are continuously turned into reserves, and this has led to the quip "since forty years, oil lasts for forty years". Alternative formulations of the **R/P ratio** therefore use the resources instead of the reserves in the numerator and the expected future consumption in the denominator in an attempt to provide more realistic

[6] Note that according to this definition, reserves are not included in the resources. Sometimes, the term resources is yet also used in a broader sense, encompassing also the reserves. This is particularly the case in general or qualitative statements on the "available resources".

Table 2.4 Key indicators of the global resource availability for exhaustible energy carriers

Primary energy carrier		Reserves	Resources	Production in 2016	R/P ratio[a]
Unit		EJ	EJ	EJ/a	a
Oil	Conventional	**7,155**	**14,183**	183	**55**
	Unconventional	**2,919**	**14,612**		
Gas	Conventional	**7,202**	**19,492**	138	**54**
	Unconventional	**269**	**20,323**		
Coal		**21,374**	**512,033**	163	**131**
Uranium and thorium		**612**	**9578**	31	**19**
Total		**39,531**	**590,221**	515	**77**

[a] Reserve-to-production ratio
Source BGR (2017), own calculations

estimates of the time until resource exhaustion. Yet, these modifications introduce additional uncertainty: resources are generally less precisely estimated than reserves, and forecasting of future consumption levels is also rather error-prone.

Overall, the data given in Table 2.4 illustrate that resource availability is not a matter of significant concern at a global scale for the decades to come. Obviously, the resources are unevenly distributed over world regions, especially Europe is (increasingly) relying on imports from other regions. But, transportation costs are affordable, and long-distance transport is current practice for all these energy carriers, except low-calorific coal resources like lignite (subsumed under coal in Table 2.4). Moreover, given the imminent challenge of global warming, global fossil fuel resource limitations are not a binding constraint for future energy and electricity systems (cf. Sect. 2.3.2). This is in contrast to traditional views on exhaustible resources. Based on Hotelling (1931), economists have developed models of price formation for exhaustible resources like fossil energy carriers for decades. These include a scarcity rent (sometimes also called **Hotelling** rent) that increases exponentially over time. In its simplest version, extraction costs for the exhaustible resource are set to zero and then the price of the resource itself grows exponentially with the general interest rate over time (cf. e.g. Zweifel et al. 2017).

For **renewable resources**, the exhaustion of reserves is, at first sight, no issue. Yet, a closer look reveals that two types of renewables have to be distinguished: those based on a stock of energy that is (more or less) continuously renewed and those where a (possibly time variable) flow of energy is exploited. Biomass is a prime example of the former case, and wind and solar energy are typical for the latter.

If there is a large stock – as is also the case for geothermal and hydro – then the critical resource constraint is the rate of renewal of the stock per year, i.e. the **potential** in primary energy newly available per year. These resources are then also dispatchable like conventional energy resources (cf. Sect. 3.3). For energy flows, the annual energy potential is also important, yet also the variability of the energy flow. Key indicators for the variability are the **full-load hours**, respectively the **capacity factor** which will be discussed in more detail in Sect. 4.3.1.

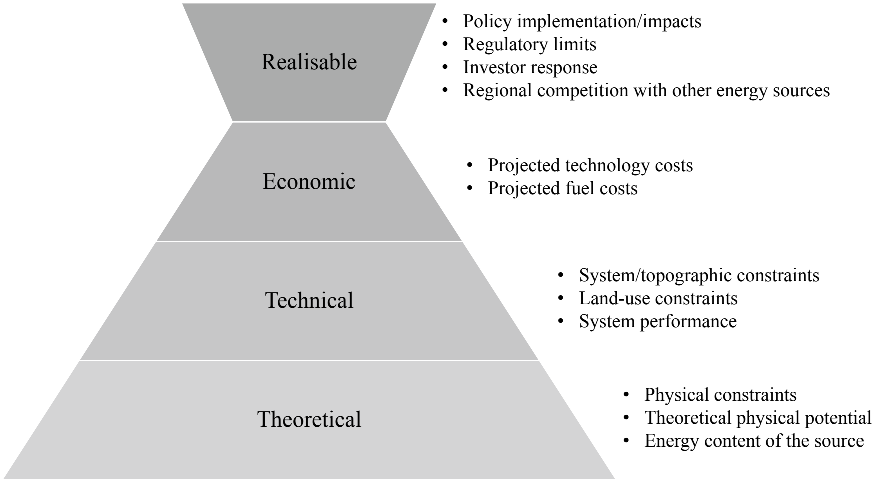


Fig. 2.9 Renewable resource potentials

Moreover, the geographical distribution of the resource availability is much more important for renewable resources – the energy flows are by nature not transportable before conversion, and even for biomass and a fortiori hydro and geothermal energy, transport is much more costly than for conventional energy carriers – basically as the energy density per unit of mass is much lower than for oil or coal.[7]

More detailed considerations of the potentials for renewable energies reveal that it is helpful to distinguish different layers of potentials (cf. Fig. 2.9)[8]:

The **theoretical potential** (sometimes also called **resource potential**) is derived based on the general physical properties of the energy source in the considered region and the conversion process (e.g. "Betz" law for wind energy, see Sect. 4.2.2). It represents the upper bound of the possible use of the energy resource. Typically, it is expressed in energy units per year, e.g. EJ/a, for a specific region – the same is valid for the other potentials.

The **technical potential** considers actual efficiencies of conversion technologies (i.e. current wind turbines). Also, land availability and topographic constraints are usually considered when the technical potential is derived. In contrast to the theoretical potential, the technical potential may change over time as technologies evolve. Also, further potentials are subject to changes over time; i.e. they are rather dynamic than static quantities.

[7] In case renewable resources are rather evenly distributed over a country, this is an advantage compared to many fossil fuels with often locally concentrated deposits. Yet, if the renewable potentials are unevenly distributed, additional challenges for energy transport and electricity grid extension arise (cf. Chap. 12).

[8] Note that partly diverging terminologies and concepts are used when renewable energy potentials are delimited, cf. e.g. NREL (2019), Resch et al. (2008). Also similar concepts of "potentials" may be applied when it comes to energy efficiency improvements.

The **economic potential** aims at identifying the share of the technical potential that is economically viable under current or future technology cost and energy prices, respectively, cost for competing generation technologies. The assessment is typically done in a system or social welfare perspective, i.e. disregarding specific policy instruments that support renewable installations. Yet, this assessment may differ considerably from potentials realised under current policy settings, which impact the profitability for economic agents, be they households or firms. E.g. grid tariffs may provide incentives to install rooftop PV systems that are not economical from a system perspective (cf. Sect. 8.2 for a more detailed discussion).

The **realisable potential** describes an upper limit to what may be implemented when policies are in place to foster renewable development and existing barriers are overcome. It usually depends on the considered time horizon and considered competing uses of land and other resources (e.g. wood). Obviously, the technical and economic potential may vary over time due to technical progress or changing prices, yet the dynamics are more pronounced for the realisable potential.

Nevertheless, the technical potential is a subset of the theoretical potential, and the economic potential is a subset of the technical potential. Yet, the realisable potential may exceed the economic potential or fall short of, dependent on whether specific policy support mechanisms also induce the realisation of installations that are non-economic according to the previously defined terms.

Potentials for renewables have been assessed at very different spatial scales over the past decades – from the level of single municipalities up to the global scale. Results yet exhibit a considerable bandwidth as shown in Table 2.5, notably when it comes to realisable potentials at larger scales. The potentials indicated in Table 2.5 are derived at a very disaggregated level and then primarily used as input data for models aiming at identifying optimal energy strategies for Europe. They are correspondingly derived as an upper limit to installations in a mid- to long-term perspective and correspond to estimates of realisable potentials. Yet, the estimates obviously differ by the number and severity of the constraints they include in the derivation, e.g. with respect to land use restrictions or trade-offs. Also, some studies such as Teske et al. (2019) explicitly include minimum requirements on resource quality when deriving the potentials. The lowest data mostly result from Hoogwijk

Table 2.5 Key indicators of the potential for renewable energy resources for Europe

	Energy potential (TWh/a)	Potential installed capacity (GW)	Typical full-load hours (h)
PV	3,000–97,400	2,400–77,500	1,300
Wind onshore	3,000–21,300	1,200–8,700	2,400
Wind offshore	1,300–24,400	300–6,600	3,700
Biomass	1,700–4,700		
Hydro (excl. pumped hydro)	600–900	130–190	4,500

Sources Hoogwijk and Graus (2008), Stetter (2014), Gils et al. (2017), Teske et al. (2019); own calculations and Graus (2008), who not only provide the earliest data reported here but also explicitly assess the future expected cost of generation. Thereby, only potentials with generation cost below 10 $ct/kWh have been retained.

Despite the broad range of estimates, the compiled data clearly show that renewable potentials in combination vastly surpass annual electricity demand (cf. Fig. 2.7). It is also apparent that hydro is most constrained, whereas the upper limit is the highest for PV. This is notably related to more specific site constraints for hydro than for the other technologies, yet it also reflects different energy yields per area. These aggregate indicators hence clearly mask regional disparities and also the time variability of many renewable generation.

2.3.2 Environmental Damage

A further key challenge related to energy use and electricity systems is the caused environmental damage. We will come back to environmental issues and how they may be coped within Sect. 6.2. Yet, the observed increase in energy supply (cf. Sect. 2.2) implies that the environmental impact of energy use raises over time – unless more efficient and environmentally friendly technologies help to curb environmental impact down. The development of global **CO_2 emissions** shown in Fig. 2.10 readily illustrates this increasing environmental impact over time. These emissions have more than doubled between 1973 and 2018, rising from 15.4 to 33.5 Gt. And be aware: for the environmental damage, it is not the per capita emissions that matter but the absolute levels as those trigger **changes in climate** and ecosystems.

This is a major challenge to mankind, and it requires enormous efforts in all parts of the world. It also sheds a different light on the resource limitations on our planet: given the requirement to limit **global warming** to a maximum of 2.0 °C above

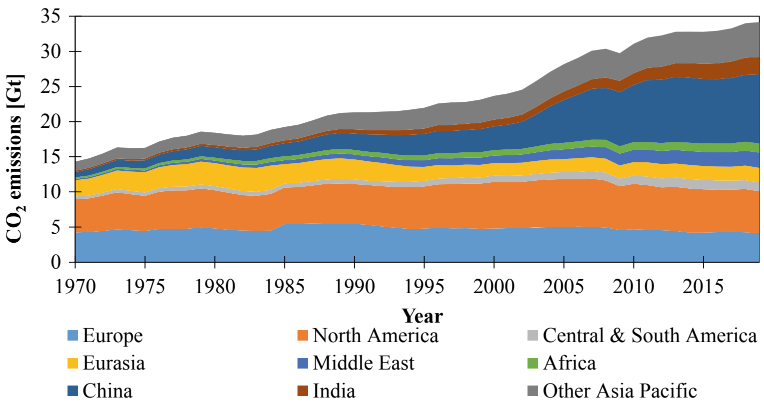


Fig. 2.10 Development of world CO_2 emissions from fuel combustion by regions from 1971 to 2018. *Source* Own illustration based on data from bp (2020)

pre-industrial levels (cf. UN 2015a, b), more than two-third of all fossil fuel resources available on earth have to remain in the ground (cf. e.g. McGlade and Ekins 2015). Or put differently: scarcity of fossil fuels is not the central issue for future energy and electricity systems. Rather, it is the limited absorption capacity of our natural environment for CO_2, other energy-related emissions, land usage restrictions and possibly metal depletion.

The link between CO_2 emissions and global warming and the role of further greenhouse gas emissions is discussed further in Sect. 6.2. Yet besides global warming, there are also other environmental issues strongly linked to energy use. These include acidification of soils and resulting stress for forests, ground-level ozone formation and resulting health impacts and further air quality deteriorations related, among other things, to the emissions of sulphur oxides (SO_x), nitrous oxides (NO_x) and particulate matters (PM).

However, there are at least three significant differences between these environmental problems and global warming: first, they are throughout of local or regional nature; i.e. the concentration of pollutants varies enormously between locations. Correspondingly, an analysis of emission developments at a global level may be misleading, even though similar problems arise in many regions of the world – frequently in densely populated urban areas.

Second, the link between the emissions related to energy and the corresponding environmental and health impacts often involves several steps of chemical reactions and air-borne transportation. This will be discussed in some more detail in Sect. 6.2. One point worth noting here is that the relevant pollutant emissions may not only arise when fossil fuels are burnt, but also in combustion processes for biomass (e.g. wood).

Finally, there is a technological difference regarding the treatment of these pollutant emissions. For many air pollutants and the corresponding human activities, specific emission reduction technologies are available (cf. Sect. 6.2.2.3). For CO_2 emissions,[9] such technologies are still under development and both their economic viability and their environmental benefits are questioned.

These fundamental issues may be illustrated by considering the past developments for three major air pollutants strongly linked to energy use, namely SO_x, NO_x and particulate emissions.

Figure 2.11 shows the development of **sulphur oxide (SO_x) emissions** in the European Union over almost three decades. Sulphur is contained to a varying degree in most fossil fuels but especially in coal. Correspondingly, the energy sectors, namely the power plants, have historically been a major source of SO_x emissions. In the power plant combustion process, the sulphur in the fuel reacts with oxygen to form mostly sulphur dioxide (SO_2) and sulphur trioxide (SO_3). Yet from the mid-1980s, power plant operators have been successively obliged to implement flue gas desulphurisation (FGD) technologies. With other measures such as fuel switch in the industry from heavy fuel oil with high sulphur content to

[9] Note that CO_2 is often not classified as an air pollutant, since it is a component of the earth atmosphere even in the absence of human activities.

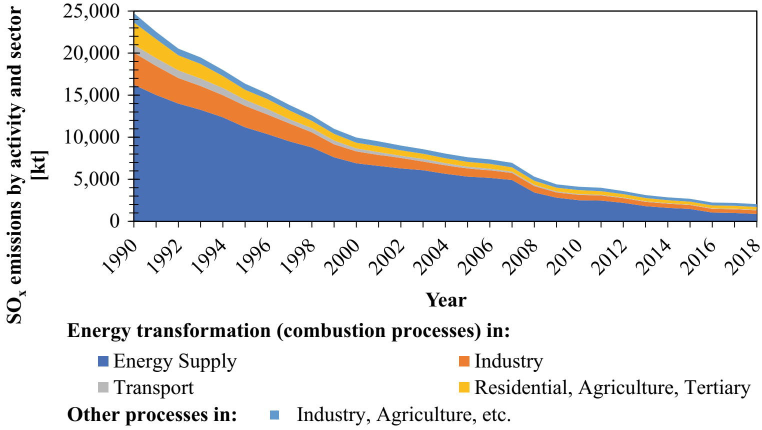


Fig. 2.11 Development of SO_x emissions in the EU-28 by activities and sector from 1990 to 2018. *Source* Own illustration based on data from EEA (2020)

natural gas, the emissions have decreased over the considered period by more than a factor of 10 (cf. Fig. 2.11).

For **nitrogen oxides (NO_x)**, Fig. 2.12 illustrates that there has also been a substantial reduction in overall emission levels. Yet, the observed reduction factor is only 2.5; i.e. the emission level in 2018 is 60% lower than in 1990. This is a consequence of several effects. Notably, the transport sector, i.e. mostly passenger cars and trucks, is the most important source of NO_x emissions as shown in Fig. 2.12. Four factors may be named why emission reduction has been less effective in this field: (1) transport demand (i.e. the “energy service” in the terminology of Sect. 2.4.1) has substantially increased in Europe over the considered period. (2) The nitrogen oxides are formed during the combustion process. As combustion temperatures are raised to obtain higher efficiencies (cf. Carnot efficiency in Sect. 2.1.3), the production of NO_x increased in the combustion due to higher temperatures (cf. Sect. 6.2.2.1). (3) A retrofit on existing vehicles is complicated and expensive and thus a seldom-used option compared to retrofits on power plants. New abatement technologies penetrate only at a slower pace in the vehicle stock. E.g. the average age of the vehicles currently registered in Germany is almost 10 years. (4) Both legal and illegal[10] practices of car manufacturers have led to a divergence between the actual on-road emissions and the technology specifications imposed by regulations, notably the so-called Euro 6 norm.

As the last case, we consider **particulate matter (PM) emissions**, i.e. small solid particles (or liquid droplets) suspended in the air. Different size classes are

[10] The most publicized case has been the manipulations performed by Volkswagen that were revealed by the U.S. Environmental Protection Agency (EPA) in 2015. But other car manufacturers have also been accused of implementing so-called defeat devices, i.e., installations that intentionally reduce the effectiveness of emission controls under real-world driving conditions, cf. Contag et al. (2017).

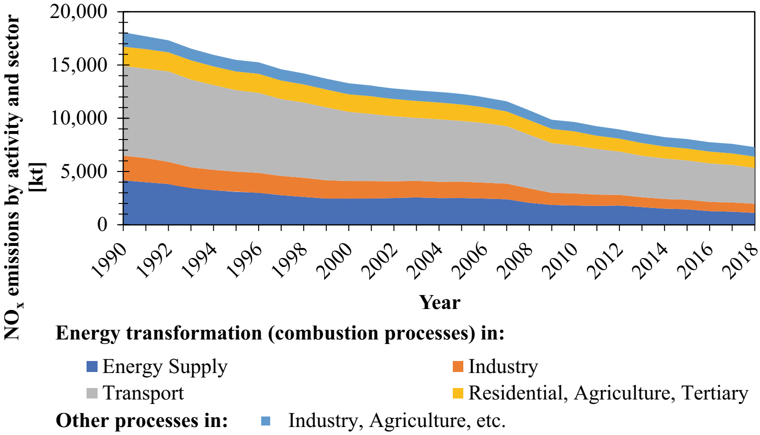


Fig. 2.12 Development of NO_x emissions in the EU-28 by activities and sector from 1990 to 2018. *Source* Own illustration based on data from EEA (2020)

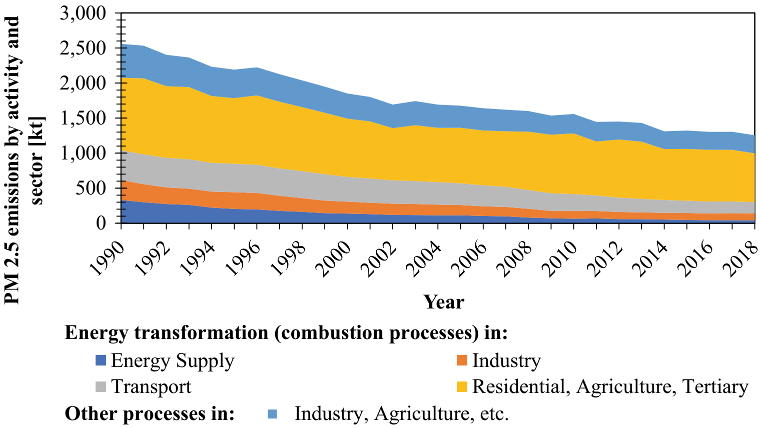


Fig. 2.13 Development of particulate (PM 2.5) emissions in the EU-28 by activities and sector from 1990 to 2018. *Source* Own illustration based on data from EEA (2020)

distinguished, the finest being PM 2.5, which corresponds to particles with diameters below 2.5 µm. Figure 2.13 shows that most of these emissions are again related to energy usage. Yet, the primary sources are the residential sector. Chimneys and other devices for the combustion of solid fuels (mainly wood and coal) contribute the largest share of emissions. Regulations particularly address new devices, and thus, emissions have only gradually decreased due to very long investment cycles in the residential sector, with a decline of about 50% over the period under investigation.

These examples illustrate the broadness of environmental impacts of energy usage. But, they also highlight the role of both technological advances and adequate regulatory settings in curbing down emissions.

As global emissions of CO_2 and other greenhouse gases have still substantially increased over the last decade, global warming remains a major unresolved environmental issue. Mitigating climate change requires deep changes in the existing energy system, and these will be a particular focus of the following chapters.

2.4 Energy Transformation Chain and Energy Balances

As outlined in the previous sections, energy is both a fundamental physical concept and a societal challenge. To deal with that challenge appropriately, it is important to have conceptual and statistical tools to describe the aggregate use of energy at the societal level as well as the use of energy in smaller systems.

2.4.1 Energy Terms and Energy Transformation Chain

To adequately describe the use of energy in our economies and societies, it is important to distinguish different categories of energy usage that may be summarised in the energy transformation chain given in Fig. 2.14.

The first important concept is **primary energy**. It designates energy carriers extracted or captured from the natural environment, e.g. coal, crude oil, natural gas, wind energy and further renewable energies.

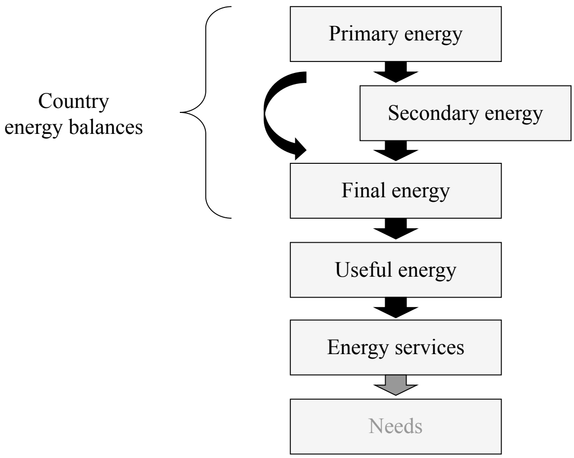


Fig. 2.14 Energy transformation chain

Secondary energy by contrast is a product of human activity, an output of (usually) specifically designed conversion processes. Important secondary energy carriers are electricity, engine fuels (gasoline, diesel, kerosene), fuel oil or district heat. As a consequence of the first and second fundamental law of thermodynamics, the conversion of primary energy carriers to secondary energy will always imply some losses; i.e. the conversion efficiency will be strictly below one. In particular, the conversion of chemical energy as contained in fossil fuels into electrical energy in power plants induces considerable losses (cf. also the row electricity plants in Table 2.6).

Final energy is not a further step down the energy transformation chain. Instead, it describes all the energy usage by entities (firms, households, public institutions) outside the energy sector. As highlighted by the two arrows pointing at final energy in Fig. 2.14, both primary and secondary energy carriers may serve as final energy. E.g. natural gas is a primary energy carrier used directly as final energy, whereas motor fuels are secondary energy carriers that are used as final energy.

Energy conversion does not stop at the level of final energy. Rather inside buildings, industrial or commercial sites, energy is further transformed into **useful energy**. E.g. heating boilers convert natural gas (or other fuels) into heat that is then transported through pipes inside buildings to the rooms to be heated. Also, the share of electricity that is transformed into lighting energy in a light bulb or light emitting diode (LED) lamp is an example of useful energy, which is only a fraction of the corresponding final energy (electricity).

Pushing one step further, the useful energy is in turn used to provide **energy services**. These are typically immaterial products, such as a heated room or a lighted area. The corresponding energy service demand may be approximately quantified using non-energy indicators such as m^2 of heated floor space in buildings.

Ultimately, these energy services serve to fulfil **human needs** such as warmth or eating. These human needs are very diverse, subjective and thus hard to quantify. Therefore, they are usually not explicitly considered in energy economics and energy engineering.

Yet, the preceding steps of useful energy and energy services are essential elements of the energy conversion chain, as energy efficiency is a key building block of sustainable energy systems. Providing the same energy service with less useful energy or the same useful energy with less final energy enables a reduction of resource usage and corresponding emissions (e.g. of greenhouse gases like CO_2). The case of light bulbs versus LED lamps is a perfect illustration of that point. Conventional incandescent light bulbs transform just about 1% of the electric energy into light energy in the visible spectrum – the rest is transformed to heat and effectively lost. By contrast, LED lamps convert 5–15% of the electric energy into visible light. I.e. for the same amount of useful energy, five to fifteen times less electricity is needed. But, one may also go one step further and consider the energy service: if the lamp is used to light an entire room, the same level of luminosity may be achieved if more daylight enters the room or if the walls are painted in white instead of dark colours. The daylight and/or the light reflected by the wall is a

Table 2.6 Aggregated energy balance of the European Union (EU 28) for the year 2019

[PJ]	Coal and similar	Oil and oil products	Natural gas	Nuclear	Hydro	Geotherm./ solar/etc.	Biofuels/ waste	Electricity	Heat	Total
Production	4,549	3,242	3,611	8,800	1,173	2,521	7,136		46	31,078
Imports	3,845	39,511	16,768				1,045	1,418	0	62,588
Exports	−547	−16,547	−2,766				−615	−1,332	−0	−21,807
Stock change	−219	−279	−753				−7			−1,257
Intl. marine and aviation bunkers		−4166	−6				−2			−4,174
Total primary energy supply	**7,628**	**21,762**	**16,853**	**8,800**	**1,173**	**2,521**	**7,557**	**87**	**46**	**66,428**
Electricity plants	−3,125	−291	−2,660	−8,670	−1,173	−2,368	−1,143	9,166	−2	−10,266
CHP and heat plants	−2,112	−282	−2,688	−130	−0	−26	−2,117	2,315	2,339	−2,700
Blast furnaces, cokeries and gas works	−870	−13	−2				−0			−886
Oil refineries and petrochemical plants		−186								−186
Other transformation	−85	142	−39				−45			−28
Energy industry own use	−257	−1,217	−737			−0	−26	−812	−181	−3,230
Losses	−28	−1	−75			−0	−3	−735	−215	−1057
Total final consumption	**1,150**	**19,915**	**10,652**			**127**		**10,022**	**1,986**	**48,075**
Non-energy use	71	3,367	656							4,093
Industry	739	1,084	3,496			1	1,361	3,689	659	11,029
Domestic transport	0	12,734	156				731	231		13,853
Residential	304	1,315	4,314			90	1,820	2,914	904	11,661
Service sector, agriculture and others	69	1,329	2,005			37	319	3,176	422	7,356
Statistical differences and transfers	−34	86	25			−0	−8	13	1	83

Source Own illustration based on Eurostat (2021) substitute (partly) for the electric light in providing the needed brightness in the room. This substitution of commercial energy input by other goods is even more prominent for room heating. The required energy service is a warm room – which may be either obtained with a lot of heating energy and little insulation or instead with little heating energy and improved insulation.

In official energy statistics, the focus is on the first three elements of the above-described energy conversion chain – given that there are no reliable measurements available for the useful energy produced or the energy services delivered. Nevertheless, a full-fledged energy system analysis should also encompass useful energy and energy service estimates to provide valid future-oriented statements. But obviously, the concepts of useful energy and energy service are not always applied as easily as in the case of room heating and lighting. E.g. the useful energy is hardly observable and even not properly defined for a multitude of electronic devices such as TVs or smartphones. And also, the energy service provided by a smartphone or a computer is hard to describe and even more challenging to quantify. Hence, the energy modeller may focus in these cases on the electricity consumption and possible future developments for this energy carrier.

2.4.2 Energy Balances

Energy balances may be determined for systems of different sizes and types. Engineers may be interested in the energy balance of a process or a plant, e.g. an industrial furnace or a power plant. Yet subsequently, we will focus on energy balances for countries or other territorial entities. Although most frequently established for countries,[11] energy balances may also be determined for subnational entities like municipalities or larger geographical units like the EU. The objective of these energy balances is to provide an aggregate documentation of the energy import/export, conversion and use in the region under study.

As with any statistical report, energy balances require a specific set of rules and conventions to allow for reliable and consistent reporting. The first convention is that energy quantities are reported as annual quantities unless otherwise stated. Consequently, energy balances do not enable statements about energy supply and demand at a certain point in time but give an aggregate view for an entire year. This is particularly relevant for electricity where storage limitations are ubiquitous (cf. Sect. 5.2). By themselves, annual energy balances do not indicate whether the regional energy system can match supply and demand in each hour of the year.

The second convention is that energy balances are presented in a table format with energy carriers in the columns and different energy-related activities in the rows. Literally, it may be called a “balance sheet”, yet there is a substantial difference to “balance sheets” used in financial accounting. The latter report stocks of value for a defined reference day, e.g. the value of tangible assets or the amount of

[11] Energy balances for multiple countries based on similar conventions are compiled by the IEA (cf. IEA 2021).

debt. Energy balances by contrast report flows – annual amounts of inputs and outputs of the various activities (cf. Table 2.6).

The overall structure of the energy balance may be summarised in three parts: the primary energy balance, the transformation balance and the final energy balance.

The **primary energy balance** describes the gross domestic energy quantities differentiated by origin (domestic production, import, exports as negative quantities, stock changes) and energy carrier. The line **total primary energy supply** provides the bottom line of the primary energy balance and gives the sum of the energy carriers used domestically.

The **transformation balance** indicates the structure of the energy transformation sector by showing groups of conversion technologies in the lines and (as before) the energy carriers in the columns. Negative entries in the table indicate consumption of the energy carrier in the conversion technology, whereas positive entries designate outputs of the conversion. The column **total** corresponds then to the energy losses for the different technologies. Further lines in the conversion sector balance indicate losses and self-consumption of energy in the transformation sector, e.g. the electricity usage of coal mines and refineries.

The **final energy balance** shows the use of the energy carriers by end-users. The line **total final consumption** gives the bottom line of primary energy supply and conversion sector activities. It also includes non-energy usages of energy carriers, notably in the petrochemical industry. While energy use in industry is generally statistically well-documented and transport energy use may be retraced based on the used fuels, respectively, the main players in the markets, the energy use in the residential and other sectors are less well-known, and data given are rather based on estimates.

The lack of reliable data is also the major reason why no statistics on useful energy or energy services are provided, although consideration of these is primordial when addressing energy efficiency issues.

When interpreting energy balance data, two further conventions are important to keep in mind: firstly, energy balances record energy quantities – not monetary values. E.g. one Peta-joule of crude oil is economically much more valuable than one Peta-joule of hard coal. And under most circumstances, this is even more true for electricity. Secondly, even the energy quantities of energy carriers are not unambiguous. For fossil energy carriers, one may debate whether the energy content should be based on the upper or the lower calorific value (cf. Sect. 2.1.3). But, the question is even more relevant for other, non-fossil energy carriers: what is the energy content of nuclear fuels like uranium or of solar and wind energy? Here, the convention adopted by the IEA, the EU and most national energy balances is to consider the first energy form in the energy conversion process for which multiple uses are practical. For nuclear power plants, this is the heat produced in the reactor, which could be used not only for electricity production but also for supplying (district) heat. For wind turbines, photovoltaic cells and hydropower plants, it is the electricity produced. With a typical conversion efficiency of 33% for electricity generated from nuclear heat, one unit of electricity produced from nuclear energy corresponds to three units of primary energy. Only one unit of primary energy is yet reported for every unit of electricity produced from wind or solar energy. This is to be considered when comparing the contribution of nuclear and renewables to energy supply based on primary energy balances. For the energy balance presented in Table 2.6, this notably means that the contribution of the sum of hydro, geothermal, solar, etc. to electricity supply is higher than the one of nuclear energy, even though primary energy figures indicate the contrary (cf. also additional figures provided by Eurostat (2021).

2.4.3 Energy Flowchart

From the data in the energy balance, various graphics may be derived to facilitate the interpretation. E.g. pie charts may be used to illustrate the shares of various primary energy sources in the total primary energy supply or to give the distribution of energy use by major sectors. A particularly illustrative type of energy diagram is an **energy flowchart** (or Sankey diagram) which shows key elements of the energy balance given in Table 2.6. Figure 2.15 shows this graph with the width of the arrows being proportional to the size of the corresponding energy flows. It notably clarifies that roughly one-third of all primary energy is lost in the energy transformation sector – primarily when producing electricity. Also, the rather even distribution of energy demand over the four main sectors industry, transport, residential and others becomes evident. Moreover, the dependency of Europe on energy imports is visible from the chart, with domestic extraction corresponding only to about 45% of the total primary energy supply. Yet by itself, this number is rather a description of the importance of world trade than an indicator of some problem. Also for other products, Europe is strongly dependent on imports, e.g. in consumer electronics such as smartphones or desktop computers.

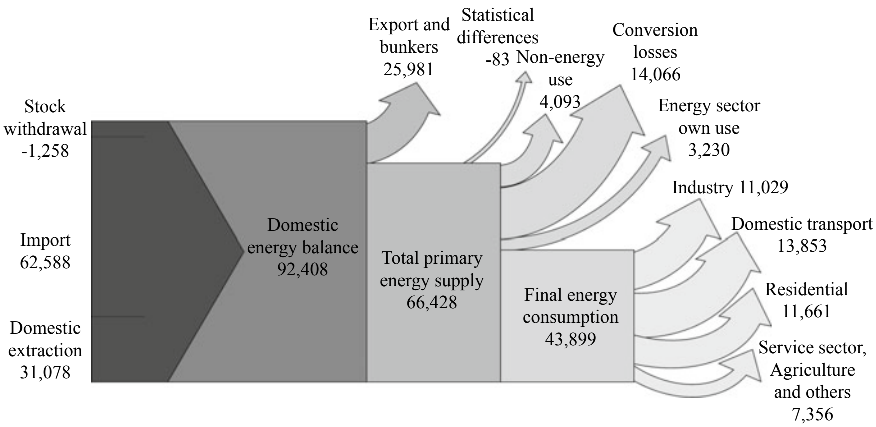


Fig. 2.15 Energy flowchart for EU-28 in PJ in 2019. *Source* Own illustration based on data from Eurostat (2021)

Economists are generally convinced that free trade is beneficial for all trading countries even if they concede that some groups within a country may face welfare losses through increases in trade, e.g. consumers in energy-rich countries who have to face price increases if their suppliers have the option to sell abroad at a better price. We come back to the welfare effects of trade in Sect. 7.2.

Yet, energy import dependency should be monitored, and especially, dependence on one supplier or one group of suppliers could be a subject of concern. Such a unilateral dependency – if there is no backup solution – may lead the supplier or group of suppliers to exert market power – as seen in 1973/74 during the first oil crisis, when OPEC temporally interrupted oil supply to Western countries.

2.5 Particularities of Electricity and the Electricity Sector

As introduced in the previous sections, electrical energy is one particular form of energy. Electricity is of very high importance for societies as it provides easily usable energy for many appliances. However, electricity has some particularities, which differ from other energetic forms and products:

- Electricity is an essential but **intangible good**. I.e. you can neither touch, see nor smell electricity when it is flowing through power lines.
- **Non-storability and equilibrium of supply and demand**: electricity cannot be stored in large amounts. This comes with the challenge of ensuring a permanent balance of electricity supply and demand.
- **Grid-bound**: large quantities of electricity can only be transported in electricity networks, respectively, power grids. Power grids are generally seen as natural monopolies (cf. Sect. 6.1), and in consequence, power grids are usually treated as regulated business.
- The permanent need to balance electricity supply and demand makes **security of supply** a vital issue in electricity systems. Security of supply has to be guaranteed (at a very high level) due to the high economic importance of electricity for industry and consumers. Aspects of security of supply will be addressed mainly in Chaps. 4 and 6.
- Electricity generation has a massive **impact on the environment**. This is not only true for conventional power generation, where several pollutants with global (e.g. CO_2) and regional (e.g. NO_x, SO_x, heavy metals, etc.) impacts are emitted (cf. Sect. 2.3), but also for renewable technologies (e.g. land usage, farming of crops, noise pollutions of wind power, metal depletion, etc.) and electricity transmission. As these environmental impacts are a co-product of electricity production and costs for these co-products are not considered – in economics, these costs are described as external costs (see Sect. 6.2) – there is a strong tendency for market intervention from the government side.
- As the quality of the good is regulated, e.g. voltage and frequency are predefined (at 230 V, respectively, 50 Hz in Europe), electricity is a **homogenous good** and

difficult to differentiate. Origin and type of generation are not directly traceable as is the case with a shipment of coal or oil. However, there are some examples of product differentiation, such as green or regional electricity. Moreover, self-generation is gaining importance due to technological progress (decentralised generation technologies, such as photovoltaic) and allows self-consumption. This results in a new relation to the product through self-generation, including new marketing possibilities.

- Consequently, electricity is often seen as a **low-interest product** with low involvement and quite a low interest in changing the supplier (low willingness to change). Additionally, there is little or **no real-time information** of consumption at the consumer level.
- **Substitutability** of electricity is very low or even zero. There are no real alternatives of energy sources for most applications in households (e.g. laptop, dishwasher, vacuum cleaner) and industry (e.g. cooling, machines).
- **Non-elasticity of demand in the short term**: in the short term, many consumers also do not see real-time wholesale market prices – even among those who have information on their consumption. As a result, no incentives exist to react to market prices in the short term, resulting in an inelastic demand. Additionally, (industrial) customers seeing real-time prices have only a very limited elasticity due to an extremely limited substitutability of electricity. Hence, electricity demand is very inelastic in the short term. In the long term, there is an incentive to invest in more efficient technologies or to reduce the need for electricity services. However, the substitutability of electricity is also very limited due to the lack of alternatives (e.g. how to operate electric appliances, such as laptop, washing machines). In consequence, price elasticity of electricity is also relatively low in the long term.
- Investments in energy technologies are characterised by **very long equipment life**, often between 20 and 40 years and sometimes even longer. This goes along with **high shares of fixed costs**, which is especially true for renewable technologies, where variable costs are close to zero (e.g. operational costs of solar and wind power are nearly zero). This has a substantial impact on investment decisions in these markets. Empirically, such industries are often characterised by **boom and bust cycles**. Boom and bust cycles describe phases of economic expansion and contraction that occur repeatedly. The boom and bust cycle is a crucial characteristic of ‘today's energy markets (not only in the electricity industry but also in the oil industry). During the boom, investments occur at general high prices, jobs are plentiful, and the market brings high returns. In the subsequent bust, prices are pretty low and little investment takes place.
- Low energy and electricity prices are an **essential location factor** mainly for energy-intensive industries. This is why energy-intensive industries, e.g. aluminium production, are located in countries with very cheap electricity costs.

2.6 Further Reading

Cengel, Y.A., Boles, M.A., & Kanoglu, M. (2019). Thermodynamics. An Engineering Approach. 9th edition. New York: McGraw Hill.

This book is an applied introduction to thermodynamics that covers the basic principles along with multiple applications.

BGR. (2021). BGR Energiestudie 2021 – Daten und Entwicklungen der deutschen und globalen Energieversorgung. Hannover: BGR.

The German geological service provides in this study an overview about global energy reserves and resources (in German).

McKenna, R. et al. (2020). On the Socio-technical Potential for Onshore Wind in Europe: a Response to Enevoldsen et al. (2019). Energy Policy, 132, 1092–1100.

The authors provide an up-to-date comparison of various potential estimates for wind energy in Europe and discuss key methodological issues.

IEA. (every year). World Energy Outlook. Paris: IEA.

An authoritative volume of analyses of the current state of the world energy markets and their prospects. Besides updates on the different fuels, technologies and regions, the outlook provides in-depth analyses on varying subjects.

BP. (every year). Statistical Review of World Energy. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/using-the-review.html.

This widely used report provides data, graphs and analyses on all major energy carriers and countries, focusing on fossil energy carriers. The data include time series and are downloadable.

2.7 Self-check of Knowledge and Understanding and Exercises

Self-check Questions

1. What are the basic principles expressed in the fundamental laws of thermodynamics?
2. Describe which concepts are used to measure the available resources of renewable and non-renewable energy carriers.
3. Indicate the different elements of the energy transformation chain and provide the corresponding examples when an electric vehicle is used that has been charged using a wind turbine.

4. Explain the elements of energy balances for countries and use the information contained in Table 2.5 to establish an energy flowchart for oil and oil products in European OECD countries.

Exercise 2.1: Power and Energy
The European Union has banned the use of conventional light bulbs and is now fostering energy-saving LED lamps.

1. How many LED lamps with nominal power of 4 W may be lit over a whole year with the energy produced by a 5 MW wind power turbine? Please assume that the wind turbine is operated at 3500 full-load hours.
2. How many hours must a gas-fired power plant operate at its rated power of 450 MW to produce the annual electricity demand of 2340 GWh of an aluminium smelter?

Exercise 2.2: Energy Conversion and Efficiency
Consider a passenger car with an annual kilometrage of 16,000 km. Useful energy consumption per 100 km is 15 kWh.

1. Compute the final energy consumption if this is an electric vehicle with an overall efficiency of the motor and the powertrain of 75% and overall efficiency of the battery charging and discharging of 80%.
2. Compute the final energy consumption for a conventional car with an internal combustion engine (motor) that consumes 7 l petrol per 100 km. Assume for the sake of simplicity that 1 l fuel corresponds to 10 kWh. What is the combined efficiency of the motor and the powertrain?
3. Compute the primary energy consumption for the three cases:

   (a) An electric vehicle with electricity coming from conventional power plants with an average efficiency of 45% and grid losses of 5%.
   (b) An electric vehicle with electricity coming from wind and solar plants with a conversion efficiency of primary energy to electricity of 100% (statistical convention in energy balances, not physical conversion efficiency). The grid losses are again 5%.
   (c) A petrol car with a conversion efficiency along the total fuel chain of 90%.

4. What are the CO_2 emissions associated with the three cases considered previously? Use the following assumptions:

   (i) No consideration of indirect emissions, e.g. associated with equipment manufacturing
   (ii) Average emission intensity in the conventional (fossil) power plant park is 0.3 kg CO_2/kWh_{fuel}
   (iii) Average emissions for petrol are 0.26 kg CO_2/kWh_{fuel}.

References

BGR. (2017). *BGR Energiestudie 2017 – Daten und Entwicklungen der deutschen und globalen Energieversorgung*. BGR. Available at: https://www.bgr.bund.de/DE/Themen/Energie/Downloads/energiestudie_2017.html [Accessed May 30, 2021].

BP. (2021). *Statistical review of world energy 2021*. Available at: https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/using-the-review.html [Accessed February 15, 2021].

Cengel, Y. A., Boles, M. A., & Kanoglu, M. (2019). *Thermodynamics. An engineering approach* (9th ed.). McGraw Hill.

Contag, M., Li, G., Pawlowski, A., Domke, F., Levchenko, K., Holz, T., & Savage, S. (2017). How they did it: An analysis of emission defeat devices in modern automobiles. In *2017 IEEE Symposium on Security and Privacy (SP)* (pp. 231–250).

EEA. (2020). *National emissions reported to the Convention on Long-range Transboundary Air Pollution (LRTAP Convention)*. Available at: https://www.eea.europa.eu/data-and-maps/data/national-emissions-reported-to-the-convention-on-long-range-transboundary-air-pollution-lrtap-convention-14 [Accessed February 6, 2021].

Eurostat. (2021). *Energy balances in the MS Excel file format*. Available at: https://ec.europa.eu/eurostat/web/energy/data/database [Accessed February 6, 2021].

Gils, H. C., Scholz, Y., Pregger, T., & Luca de Tena, D. (2017). Integrated modelling of variable renewable energy-based power supply in Europe. *Energy, 123*, 173–188.

Hoogwijk, M., & Graus, W. (2008). *Global potential of renewable energy sources: A literature assessment*. ECOFYS.

Hotelling, H. (1931). The economics of exhaustible resources. *Journal of Political Economy, 39*, 137–175.

IEA. (2020). *Key world energy statistics 2020*. IEA. Available at: https://iea.blob.core.windows.net/assets/1b7781df-5c93-492a-acd6-01fc90388b0f/Key_World_Energy_Statistics_2020.pdf [Accessed February 15, 2021].

IEA-PPSP. (2021). *Snapshot of global PV markets*. IEA. Available at: https://iea-pvps.org/wp-content/uploads/2021/04/IEA_PVPS_Snapshot_2021-V3.pdf [Accessed June 11, 2022]

IEA. (2021). *World energy outlook*. IEA.

McAllister, S., Chen, J.-Y., & Fernandez-Pollo, A. C. (2011). *Fundamentals of combustion processes*. Springer.

McGlade, C., & Ekins, P. (2015). The geographical distribution of fossil fuels unused when limiting global warming to 2 °C. *Nature, 517*, 187–190.

McKenna, R., et al. (2020). On the socio-technical potential for onshore wind in Europe: A response to Enevoldsen et al. (2019). *Energy Policy, 132*, 1092–1100.

NREL. (2019). *Renewable energy technical potential*. Available at: https://www.nrel.gov/gis/re-potential.html [Accessed May 14, 2022].

Resch, G., Held, A., Faber, T., Panzer, C., Toro, F., & Haas, R. (2008). Potentials and prospects for renewable energies at global scale. *Energy Policy, 36*, 4048–4056.

Stetter, D. (2014). *Enhancement of the REMix energy system model: Global renewable energy potentials, optimized power plant siting and scenario validation*. University of Stuttgart ITW and Dissertation: U Stuttgart.

Teske, S., Nagrath, K., Morris, T., & Dooley, K. (2019). Renewable energy resource assessment. In S. Teske (Ed.), *Achieving the Paris climate agreement goals* (pp. 161–173). Springer.

The Engineering Toolbox. (2021). *Fuels—Higher and lower calorific values*. Available at: https://www.engineeringtoolbox.com/fuels-higher-calorific-values-d_169.html [Accessed February 15, 2021].

UN. (2015a). *Paris agreement*. United Nations. Available at: https://unfccc.int/sites/default/files/english_paris_agreement.pdf [Accessed February 15, 2021].

UN. (2015b). *Transforming our world: The 2030 agenda for sustainable development*. United Nations. Available at: https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf [Accessed February 15, 2021].

Voß, A. (2003). *Einführung in die Energiewirtschaft*. Unpublished Manuscript.

Zweifel, P., Praktinjo, A., & Erdmann, G. (2017). *Energy economics. Theory and applications*. Springer.

3 Energy Demand

Covering the demand for energy services is the driving force of the whole energy sector, and this indicates the importance of demand in the energy value chain. However, policy instruments in place (see Chap. 6) and investment activities frequently focus on the supply side, whereas the demand side is often treated rather poorly.

To deliver the desired energy services (e.g. an illuminated room), useful energy (e.g. light) has to be provided, which is done by transforming final energy carriers (e.g. electricity). Although energy services may be considered the real driver (see Chap. 2), this chapter will mainly focus on the demand for final energy carriers, particularly for electricity and heat, as energy services and useful energy are difficult to quantify. Nevertheless, it is worth mentioning again that the first step in making the energy transformation chain more sustainable is to analyse whether the requested energy services can be reduced (without any loss of comfort) or provided more efficiently (see Sect. 2.4.1). Key questions to be answered in this chapter are as follows:

- What are the key drivers and characteristics of electricity and heat demand?
- What do electricity load profiles look like?
- By which instruments can the demand side be influenced?
- Which methods are used to forecast the future electricity demand?
- How can electricity tariffs be differentiated?

These questions are discussed starting with the final energy carrier electricity in Sect. 3.1; subsequently, the heat demand is analysed in Sect. 3.2, as it is partly produced jointly with electricity in so-called combined heat and power (CHP) plants (see Chap. 4). Both subchapters will have a focus on households as the most homogenous sector.

Key Learning Objectives
After having gone through this chapter, you will be able to

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_3

- Describe the level, the structure and the key components of electricity and heat demand.
- Understand the basic ideas of different methods to forecast energy demand.
- Define the key concepts of demand-side management (DSM) and different electricity tariffs.

3.1 Electricity Demand

3.1.1 Basics

An important parameter to describe the **electricity demand** of a region/country is the net electricity consumption, composed of the electricity demand of the different sectors (transport, households, tertiary and industry) in this region/country. To calculate the gross electricity consumption, the losses in electricity transportation and distribution as well as the electricity needed to operate the power plants (including pump storage power plants) have to be added. The gross electricity consumption can also be computed by adding electricity imports to and subtracting electricity exports from the gross electricity production in the region/country.

Worldwide electricity consumption has been growing for many years. Meanwhile, the net electricity consumption[1] has gone beyond the mark of 20 PWh (see Sect. 2.2). According to Eurostat, the net electricity consumption of the EU28 was about 2,700 TWh in the year 2015 (cf. Eurostat 2019). The share of the transport sector is minor, the industry sector is responsible for about 40% and the household and the tertiary sector each for about 30% of the net electricity consumption (see Fig. 3.1).

Net electricity consumption in Germany as an example for an industrialised country was about 527 TWh in 2018, the industry sector having a share of slightly below 50% and households and the tertiary sector a share of about a quarter each and transport a share of about 2% (cf. BDEW 2019). The final energy carrier electricity is used to provide a variety of energy services in the different sectors, whereby mechanical energy and process heat dominate in Germany (cf. Ströbele et al. 2010, p. 218). To satisfy its demand of energy services, a German single-person household needs about 2,000 kWh of electricity per year on average. Of course, there are considerable differences in electricity consumption of different households; the demand typically increases with, e.g. the dwelling space and the appliances used. A higher consumption also occurs if electricity is also used for heating purposes. The additional electricity demand if having more than one

[1] Comparing the net electricity consumption with the gross electricity production (see Sect. 2.2) provides an indication of the losses in the system.

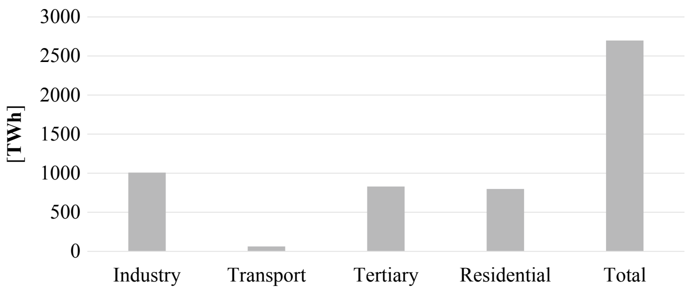


Fig. 3.1 Net electricity consumption in EU28 in the year 2015. *Source* Own illustration based on data from (Eurostat 2019)

member in the household usually is below these 2,000 kWh. In other words, the electricity demand per person is falling with an increasing size of the household. This can be explained, e.g. by the fact that households of double size do not necessarily have double the number of devices and appliances, e.g. to provide the energy services of illuminated rooms or cooling of food.

3.1.2 Applications on the Demand Side

Electricity demand is the sum of the electricity consumption of the demand-side technologies employed in the considered system. Therefore, the electricity demand can be calculated by considering the installed appliances, their specific electricity demand and their time of use. The demand-side technologies can be described by characteristics also used for supply-side technologies (e.g. investment costs, efficiency, etc.; see Sect. 4.3). The main difference is that the demand-side applications transform the final energy carrier electricity into useful energy, which is then used to provide the desired energy service. Examples for such technologies are in the industry sector electrolysis processes and electric arc furnaces, in the residential and tertiary sector fridges, freezers and electric heating systems. Given the relatively homogenous structure of the household sector – especially compared to the industrial sector – the share of the different appliances in the electricity demand of an average household can be determined rather easily (see Fig. 3.2).

With regard to the technologies used, cross-cutting technologies such as pumps, fans and compressors, which are used in many sectors and subsectors, may be distinguished from sector-specific technologies, like electric arc furnaces used in the iron and steel sector. As cross-cutting technologies are applied in totally different sectors, improvements in the efficiencies of these technologies will reduce electricity consumption in many places.

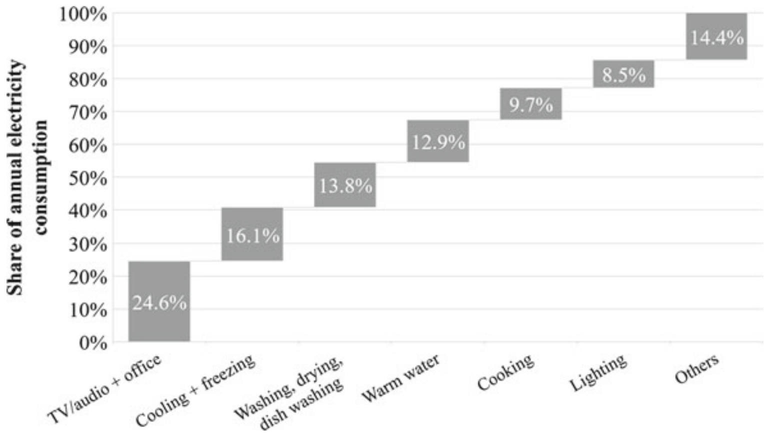


Fig. 3.2 Electricity consumption in German households in the period 2007–2011. *Source* Own illustration based on (Oberascher 2013)

As soon as new technologies using electricity as an input factor emerge, this impacts the electricity demand of the customers, sectors and countries. This might be of huge importance for the future energy system as more and more technologies are expected to use electricity instead of fossil-based energy carriers and sectors like the transport sector will become more and more electrified (so-called sector coupling or energy system integration), as electricity can rather easily be decarbonised. Therefore, promising applications like **electric vehicles** and **heat pumps** can have a considerable impact on future electricity demand, e.g. an electric vehicle will considerably increase the electricity demand of a household. Assuming an average annual mileage of a passenger car of 14,000 km and an average consumption of 15 kWh/km, this results in an additional yearly electricity demand of 2,100 kWh.[2]

3.1.3 Load Profiles

A **load profile**, also called load shape, shows the power consumption variation, measured, e.g. in watt, over time. The applications used in a system determine the shape of the load curve of this system. This load curve is of paramount importance since possibilities for storing electrical energy are limited (see Sect. 5.2). The load profile of a region consists of the superposition of the profiles of the customers (from different industry branches as well as from sectors like transport, households and tertiary) or appliances in this region. Fortunately, the load profiles of different customers have different forms, as these demand electricity at different times; e.g.

[2] From a system perspective, it has to be considered that this new electricity demand will, in most cases, be accompanied by a reduction of energy carriers used up to now in internal combustion engines (ICE) like diesel or gasoline.

households typically demand electricity mainly in the evening in winter time, when people frequently come home from work, whereas demand from bakeries has its peak during early morning hours. So, the electricity demand peak in a region is not equal to the sum of all electricity demand peaks of the customers in this region, as the peaks of the different customers are not simultaneous. The so-called **coincidence factor** is frequently used to consider this effect in the planning process. The coincidence factor is computed by dividing the demand peak of a system/region by the sum of all demand peaks of the electricity demanding entities in this system/region.

Load profiles can be interpreted as a combination of deterministic and stochastic processes. In households, the profiles are the result of switching on and off different electrical appliances. Therefore, load profiles are influenced by factors like occupant characteristics. Although the time of use of some appliances is easily foreseeable, other appliances seem to be switched on and off more or less randomly, introducing a stochastic component to electricity load profiles (cf. e.g. McLoughlin et al. 2010).

For customers with relatively small consumption quantities like households, load profiles have until recently mostly not been measured, but typical load structures have been clustered to obtain so-called standardised **load profiles**. The reason for not measuring the details of the electricity consumption is that due to the limited amount of electricity delivered to these customers, it was for a long time not seen to be economically justified to install the necessary metering equipment – a fact that is just beginning to change due to modern smart meter technologies. The standardised load profiles are typically used by load serving entities to determine the demand structure of a bundle of such customers, e.g. some thousands of households. Of course, the true load profiles of single households (see Fig. 3.3) might be distinct from these standardised load profiles showing much more fluctuations (e.g. due to the use of an electric kettle in the morning). But these fluctuations average out as soon as a larger number of customers is considered. For major customers like industrial plants (typically with a yearly demand of more than 100,000 kWh), the consumed energy is measured in a more elaborated way, e.g. for every 15 min interval.

If the hourly load over the whole year is sorted in descending order, the resulting graph is called a sorted annual **load duration curve** (see Fig. 3.4). Load duration curves have often been used during the planning process of new generation capacities as they indicate, which capacity is needed for which duration. With the help of the load duration curve, the so-called base load, which is the minimum load over the whole year (point A in Fig. 3.4), can easily be identified.

Many factors are influencing the load profile of different customers, like lifestyle, attendance or working times or the stock of installations. Additionally, the diffusion of emerging technologies using electricity as an input factor (e.g. **electric vehicles**) might drastically change the future load profiles of customers and so the total load profile of the system/region. Furthermore, customers tend to produce more and more electricity by themselves, e.g. with the help of rooftop PV (see e.g. Sect. 4.2.3). These so-called **prosumers** (see also Sects. 6.1.4 and 12) reduce the

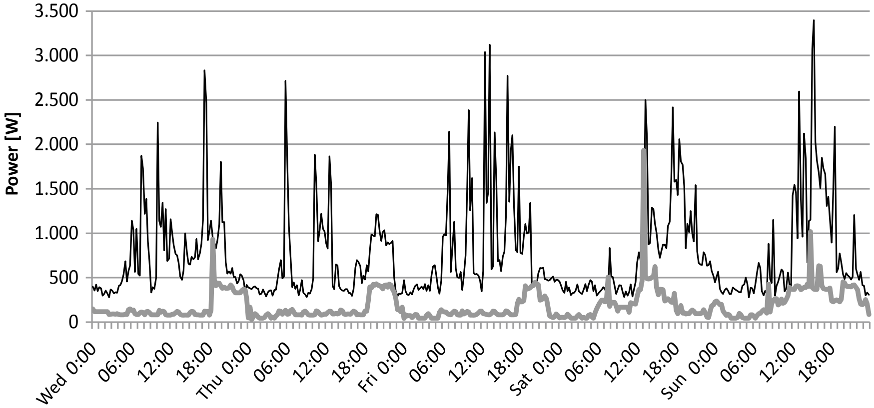


Fig. 3.3 Selected load curves of two households. *Source* Own illustration based on (Kaschub 2017, p. 20)

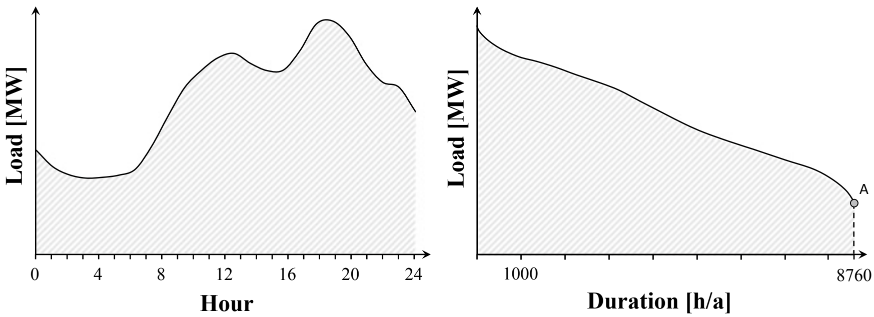


Fig. 3.4 Daily load profile (left) and annual load duration curve (right)

amount of electricity taken out of the power grid and change the electricity demand profile still to be delivered to these customers (so-called net or residual load) from the grid.

3.1.4 Demand-Side Management

Demand-side management[3] comprises activities on the demand side to reduce the load of customers[4] in general (energy efficiency objective) or to reduce or increase their load during specific periods (load shifting objective) (cf. Kostkova et al. 2013). Both ideas have been on the agenda for many years. Increasing the energy efficiency on the demand side helps to diminish emissions and the depletion of

[3] There is no clear differentiation of the technical terms demand-side management (DSM), demand response (DR) and load management (cf. e.g. Albadi and El-Saadany 2008; Kostkova et al. 2013).
[4] Or at least their purchase of electricity from the electric grid.

resources by reducing the electricity produced. Recently, load shifting has received increased attention as it is seen as a flexibility option, which can compensate fluctuations of electricity production by wind and PV.

To react in the desired way, customers need to be incentivised, e.g. by price-based or incentive-based programmes (cf. DoE 2006). Within an incentive-based programme, customers get a payment for actively reducing their demand or for agreeing that a pre-defined entity, e.g. their power company or the system operator, is allowed to remotely control the use of some of their devices during critical hours. Therefore, this form of incentivising load shifting is also called direct load management. In contrast, price-based programmes comprise tariff forms setting different charges in different situations (see Sect. 3.1.6). This form of incentivising load shifting is called indirect load management.

While typically the responsiveness of many customers, e.g. from the household sector, is relatively low, the establishment of such programmes helps increase and use the customers' responsiveness. The **price elasticity** of demand (sometimes – to put it simply – just called demand elasticity) indicates thereby how the (incremental) demand (q) changes in response to an (incremental) price (p) change (see 3.1):

$$\varepsilon = \frac{dq}{dp} \cdot \frac{p}{q}. \tag{3.1}$$

In the electricity sector, demand is said to be relatively price inelastic (absolute value of ε is below one), as a change in prices typically results in a rather small change in the quantity of electricity consumed. Of course, the price elasticities of electricity demand depend on many different aspects, like the customers considered, the point in time, etc. Subject to the considered period, demand elasticity is typically differentiated into a short-term and long-term price elasticity of demand. While the long-term price elasticity of electricity demand is still lower than price elasticities in many other sectors, it is higher than the short-term price elasticity of electricity demand. The reason is that customers can find substitutions for energy-intensive devices, e.g. more efficient devices still using electricity or appliances using other energy carriers, like natural gas, to provide the needed energy service. In the short term, customers have hardly any further possibility of reacting than changing their usage habits.

Not all applications are suitable for **load shifting**. Applications like lighting, TV and PC, which are used when there is a direct demand for the corresponding service, can hardly be used for demand response because there will be little willingness to accept such an intervention into daily life. Much more promising are appliances, the operation of which is projectable, like dishwashers and dryers. Very suitable for demand response are appliances that can decouple the electricity consumption and the service provision, e.g. due to the availability of a storage unit (cf. e.g. Boßmann 2015, pp. 19 and 23). In this context, it is advantageous if the user does not even notice that the regular demand has been changed; an example could be the premature use of a refrigerator's compressor. Hopes are especially placed in potentials for load shifting realised by appliances like electric vehicles and heat pumps, which are becoming more and more important in the context of decarbonisation and associated sector coupling strategies (see Chap. 12), as these appliances lead to a high additional electricity demand that can rather easily be shifted within certain boundaries.

Despite the interest in the topic, there are many barriers to the realisation of demand-side management, which could result in an underinvestment (so-called energy efficiency gap). Many customers do not have the necessary information about energy-saving and load shifting measures. Overcoming this barrier might lead to additional costs, so-called transaction costs. Concerning short-term activities, the main barrier seems to be the resistance to deviate from traditional behaviour patterns. As soon as investments have to be realised, capital limitations and long payback times hinder demand-side activities. Especially in the industrial sector, short payback times are usually required (often in a range between 2 and 3 years). This can be a considerable hindrance for investments in energy-efficient technologies, which often have long lifetimes far exceeding these payback times. Therefore, investment opportunities, that would be rather promising if other investment criteria (e.g. the net present value) were used, will not be realised.

3.1.5 Projecting Electricity Demand

Electricity supply and demand have to be always in equilibrium to keep the frequency at the foreseen level (cf. Chap. 5). Therefore, it is essential to forecast the electricity demand[5] on a short-term basis (e.g. for the next hours), a medium-term basis (e.g. for the next weeks and months), as well as on a long-term basis (e.g. for the next years or even decades). Short-term projections of **electricity demand** are required, e.g. to adjust the operation of the existing generation units (see Sect. 4.4), medium-term projections, e.g. to plan necessary maintenance operations of these installations and long-term projections, e.g. to identify the requirements for additional generation capacity.

In general, electricity demand in the different sectors of the energy system, households, tertiary, industry and transport is influenced by many factors, which have to be projected to forecast electricity consumption. Factors influencing the long-term electricity demand are, e.g. the stock of installations, taking into account the efficiency of the used appliances, the lifestyle of the population, the structure of the whole economy, the industry structure (share of energy-intensive production), the gross value added (GVA), socio-demographic factors, the available income and the price of electricity. In addition to these factors, for short-term projections of electricity demand with a much higher temporal resolution, additional information is necessary, e.g. the start of the broadcast of a football match or the weather conditions in the next hours.

A multitude of qualitative and primarily quantitative methods exists to forecast electricity demand on the short-term, medium-term and long-term basis – many of

[5] As well as, e.g. the fluctuating supply.

these methods are also used to forecast electricity prices (a comprehensive classification of electricity price modelling approaches can, e.g. be found in Weron (2014), an overview of stochastic models used in electricity markets in Möst and Keles (2010). To forecast the development of long-term electricity demand, fundamental drivers of demand and their further development are typically analysed. The so-called top-down approach tries to model electricity demand as a function of macroeconomic variables like the demographic and the economic development, as electricity consumption typically increases with the growth of the gross value added or the population (cf. Zweifel et al. 2017, pp. 89–110 and Chap. 2). In contrast, the so-called bottom-up approach takes into account much more detailed information about the appliances in the different sectors and the intensity of their use (cf. Zweifel et al. 2017, pp. 65–87). Nevertheless, also demand projections based on a bottom-up approach start by estimating the development of macroeconomic variables like the gross domestic product (GDP), the population and wholesale prices. This data is used to determine more disaggregated factors like the sectoral gross value added (GVA), the physical production, the employment and energy prices for the different end-users. In a next step, this information is used to derive the main drivers of the electricity demand in the different sectors, like the value added in the different industrial branches, the numbers of employees in the different subsectors of the tertiary sector, the number of households and their net dwelling areas and the tonne- and passenger-kilometres in the transport sector (cf. e.g. Fraunhofer-ISI et al. 2015). To model electricity demand in such a comprehensive way, a variety of data about the technologies available in the different sectors and their specific electricity consumption have to be taken into account, which already illustrates the challenge of data availability.

More emphasis has to be put on a higher temporal resolution of the forecast for shorter forecasting periods. Here, e.g. econometric time-series models are often applied, which are based on the idea to identify patterns in historical time series and to use these patterns to develop a forecast. Time-series models employ statistical methods, like regression and smoothing techniques, to forecast future electricity demand. Furthermore, approaches from the emerging field of artificial intelligence (AI), like neural networks, can be used to forecast electricity demand. With the help of neural networks, a forecast of the electricity demand can be developed without knowing any details about the relationship between this output and different inputs, like, e.g. weather conditions. Finally, it should be mentioned that different forecasting methods, e.g. a long-term model based on fundamentals of the system and a model trying to represent short-term stochasticity, can also be combined, leading to so-called hybrid methods.

3.1.6 Electricity Tariffs

Electricity tariffs describe the payment structure customers face when they pay for electricity usage. Electricity tariffs might consist of different components, like an annual base rate (€/a), energy rates (€Cent/kWh) and capacity rates (€/kW). Under such a system with different components, the total bill to be paid by the customer cannot be calculated by just multiplying the quantity (kWh) consumed by the specific price (€Cent/kWh), which is why such pricing systems are called **nonlinear pricing systems** (cf. Oren 2012).[6] By paying the tariff, the customer not only settles the costs to deliver the electricity but also the costs caused by the necessary use of the electricity grid (use-of-system charges, see Chap. 6) as well as fees (e.g. to finance the extension of renewable energies or very efficient technologies like cogeneration units) and taxes. Electricity tariffs of different customers vary in their components and their relative importance. Typically, households and other small customers only pay a base rate and essentially a **volumetric electricity price** – a rate per kWh electricity used. A tariff (€Cent/kWh) which is more or less based on volumetric end-user prices typically is much higher than the average electricity wholesale price (€Cent/kWh) as it incorporates, e.g. grid charges, levies and surcharges. However, such tariffs may not give the appropriate incentives to the consumers, especially concerning sector coupling and self-consumption (see Chap. 12). Industrial companies are also charged a rate for their yearly peak demand (in kW) called a non-peak-coincident demand charge. As the peak demand of a company does not necessarily coincide with the peak demand of the whole system, activities to reduce the company-specific peak demand do not necessarily have a positive effect on the entire system. To consider the contribution of the company to the **peak load** in the whole system, so-called **peak-coincident demand charges** can be used. These charges consider the demand of the company during the time of peak demand in the system. Furthermore, different forms of discounts might be given to the customers, e.g. the transportation and distribution system operators can reduce the use-of-system charges for energy-intensive companies if their delivery structure is untypical (high company-specific demand in times of low demand in the whole system).

The energy rates of electricity tariffs (€Cent/kWh) can be constant (simple rate) or vary over time, which seems to fit better to the time-varying costs of electricity production (see, e.g. Boiteux 1960). Energy rates may depend on the amount of delivered electricity; there are tariffs with different thresholds for the delivered electricity with higher energy rates as the demand of the customer increases (tiered rate). Alternatively, energy rates may depend on the time of delivery, e.g. the hour of the year. This form of letting the customers participate in the volatility of wholesale prices is called dynamic pricing (cf. Joskow and Wolfram 2012). Time depending tariffs can be differentiated into

- **Time-of-use (TOU) tariffs**: The prices change according to different pre-defined time periods (e.g. day–night).
- **Critical peak pricing (CPP) tariffs**: Higher prices are applied if pre-defined critical events (high demand) occur.
- **Real-time pricing (RTP) tariffs**: The prices vary rather often, e.g. on an hourly basis – but also here the customers usually know the tariffs in advance.

[6] Another form of price discrimination – so-called Ramsey prices – will be discussed in Sect. 6.1.4.

Tariffs offering dynamic forms of adaptability provide the possibility to incentivise load shifting activities of customers on a short-term basis, a feature that might be beneficial in energy systems with a high share of fluctuating electricity production (see Sect. 3.1.4).

In contrast to time-dependent tariffs, load-dependent tariffs have the feature that the price for electricity (€Cent/kWh) depends on the required capacity. As soon as a threshold concerning the load (kW) is passed, the electricity rate will be changed. Furthermore, the capacity rates (€/kW) can vary (so-called load-variable tariffs): the rates will increase with the customer's load that has to be served. Having the proper measuring and controlling installation in place such tariffs may even foresee that the power consumption of a customer is limited to a contracted guaranteed power level.

3.2 Heat Demand

Heat demand is responsible for a large part of the final energy demand. The importance of the heating sector may be illustrated with the following data: in 2019, more than 50% of the final energy demand has been used for heating purposes in Germany (i.e. more than 1300 TWh), in the household sector even about 90% (cf. e.g. Arbeitsgemeinschaft Energiebilanzen 2020). According to the European heat roadmap, the heat demand in the largest EU member states[7] was almost 4,500 TWh in 2015 (Paardekooper et al. 2018). More than half of this demand is related to space heating purposes, especially in the residential sector (see Fig. 3.5). In industry, most of the heat is needed to provide process heat, e.g. for drying or melting applications. These different forms of heat are insofar different products as they have different temperature levels and therefore different thermodynamic values, as the exergy of heat depends on the temperature level (see Sect. 2.1.2). Process heat is typically needed at much higher temperatures (partly even above 1,000 °C) than space heat or hot water. The temperature level also influences the technology used to provide the requested heat. The heat production process is typically realised locally, rather close to the point of demand,[8] because heat transport is limited by the low-energy density of hot fluids like hot water and steam. Furthermore, unlike electricity, heat can be stored quite efficiently, so heat production does not exactly have to follow heat demand. Typically, the demand for process heat is relatively constant during the year, of course strongly depending on the activity level of the industrial process for which the heat is needed. The demand for **hot water** is also relatively constant over the year but with sporadic peaks during the day (e.g. due to the daily shower in the morning). The absolute level (e.g. in one building/region/country) mainly depends on the number of users. In contrast to process heat and hot water, the demand for **space heating** shows considerable

[7] These 14 EU member states are responsible for about 90% of the EU heat demand.

[8] Nevertheless, there are district heating systems, which transport the heat up to about 50 km to the customers.

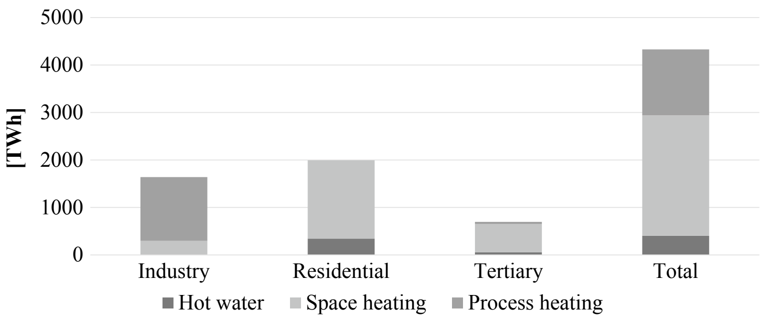


Fig. 3.5 Heat demand in the largest EU member states in 2015. *Source* Own illustration based on data from (Paardekooper et al. 2018)

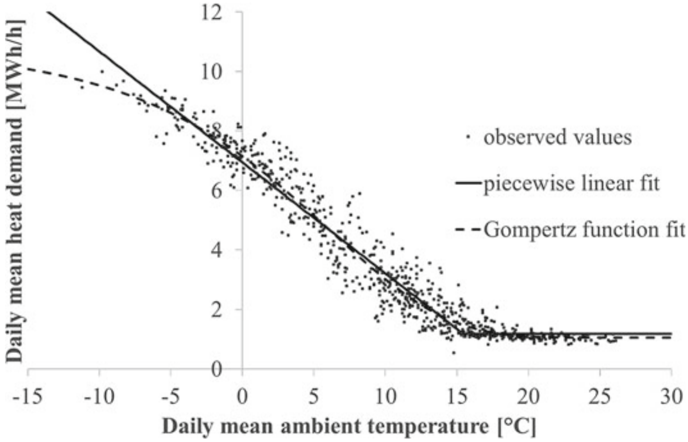


Fig. 3.6 Exemplary energy demand for space heating as a function of outside temperature

seasonal variations as it is strongly dependent on the outside temperature. Furthermore, the demand for space heating depends on the floor area to be heated within the considered system.

According to the first principles of (non-radiative) heat transfer, heating energy demand should increase linearly when the outside temperature drops below a threshold temperature, sometimes labelled “heating limit temperature” (see Fig. 3.6). The dashed, sigmoid line in Fig. 3.6 considers some smoothening and limiting effects in real-world systems, including the heterogeneity of the “heating limit temperature” within any sample of buildings and occupants. On the other hand, the dimensioning of the heating equipment (radiators, heat exchangers, boilers, etc.) limits demand at extremely low temperatures.

Process heat demand arises in multiple sectors, ranging from primary materials fabrication (e.g. steel or plastics) to food processing (e.g. bakeries). There are multiple types of processes that require heat on a relatively high-temperature level, e.g. for drying, melting, chemical conversion or cleaning applications. In the iron and steel, non-ferrous metals, glass, ceramics and cement industries process temperatures of several hundred degrees are needed (cf. e.g. McKenna and Norman 2010). A common characteristic of process heat demand is the independence or limited dependence on the outside temperature. However, some correlation may be observable, e.g. when the ambient air is heated and used for drying purposes.

3.3 Further Reading

Borenstein, S., & Holland, S. (2005). On the efficiency of competitive electricity markets with time-invariant retail prices. RAND Journal of Economics, 36, 469–493.

In this paper, the different effects of real-time pricing are discussed.

Zweifel, P., Praktiknjo, A., & Erdmann, G. (2017). Energy Economics – Theory and Applications. Berlin, Heidelberg: Springer.

The book Energy Economics gives a comprehensive overview of energy economics and focuses in two chapters on bottom-up and top-down analysis of energy demand.

Blesl, M., & Kessler, A. (2017). Energieeffizienz in der Industrie. 2nd edition. Berlin, Heidelberg: Springer Vieweg.

This book gives an extensive introduction into energy-saving measures in the industry in German language.

McKenna, R., & Norman, J. (2010). Spatial modelling of industrial heat loads and recovery potentials in the UK. Energy Policy, 38, 5878–5891.

This paper presents a comprehensive estimation of the heat demand, differentiated according to different temperature levels, and the technical recovery potential for industrial sectors in the UK.

3.4 Self-check of Knowledge and Exercises

Self-check of Knowledge

1. What is the level of the yearly electricity and heat demand in Europe?
2. Which different forms of heat do you know?
3. Which sectors are responsible for which shares of the total electricity/heat demand in the European Union?
4. Which quantitative methods are typically used for forecasting energy demand on a short-term and on a long-term basis?
5. Define the price elasticity of demand!

6. Name different factors influencing short-term and long-term demand for electricity, space heating, hot water and process heating.
7. What is the difference between a load profile and a load duration curve?
8. Which forms of electricity tariffs do you know?

Exercise 3.1: Investments in Efficient Demand-Side Technologies
You need a new washing machine and a new gas-based heating system for your household (interest rate: 4%). To reduce your contribution to climate change, you decide to invest in the first step in one of the two cases in a very efficient appliance. As you have different opportunities, you want to identify this appliance using the performance figure "Euro per tonne of CO_2 avoided" (€/t CO_2). Your local energy company runs a steam power plant fired with hard coal and offers the following tariffs: electricity tariff: 30 Cent/kWh, gas tariff: 7 Cent/kWh. In which of the two more efficient technologies should you invest? The local energy company announces to switch from coal to gas (CCGT). Does this influence your decision?

	Lifetime [years]	Additional investment compared to the standard	Yearly energy demand compared to the standard
Washing machine	10	200 €	45 kWh (standard: 90 kWh)
Heating system	10	650 €	9,500 kWh (standard: 10,000 kWh)

Exercise 3.2: Forecasting Electricity Consumption
In the following table, the net electricity consumption in Germany in the years 2010–2019 is depicted. Forecast the net electricity consumption in Germany in the year 2025, applying linear regression using the least square method. Critically discuss your result.

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
[TWh]	540	535	530	535	525	525	525	530	520	515

Exercise 3.3: Introducing Time-of-Use Tariffs
Calculate the yearly electricity bill of a household with an annual consumption of 4500 kWh according to the following load profile (in % of the total demand) under a fixed electricity tariff of 30 Cent/kWh.

hour	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00
[%]	3%	2%	2%	2%	2%	3%	4%	5%	4%	4%	4%	5%
hour	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	0:00
[%]	4%	4%	4%	4%	5%	6%	7%	7%	6%	5%	5%	3%

The energy supply company wants to introduce a time-of-use tariff with two different zones: peak time from 8 am to 8 pm and off-peak time from 8 pm to 8 am. During the off-peak time, the tariff is supposed to be 20 Cents/kWh. How high must the tariff be during peak time to realise the same revenues assuming

(a) that customers will not change their behaviour,
(b) that customers will shift 10% of their current demand in peak hours into off-peak hours?

References

Albadi, M., & El-Saadany, E. (2008). A summary of demand response in electricity markets. *Electric Power Systems Research, 78*, 1989–1996.

Arbeitsgemeinschaft Energiebilanzen (2020). *Anwendungsbilanzen zur Energiebilanz Deutschland.* Available at: https://ag-energiebilanzen.de/wp-content/uploads/2020/10/ageb_20v_v1.pdf [Accessed May 27, 2022].

BDEW. (2019). *Nettostromverbrauch nach Verbrauchergruppen*. Available at: https://www.bdew.de/service/daten-und-grafiken/nettostromverbrauch-nach-verbrauchergruppen/ [Accessed May 5, 2019].

Boßmann, T. (2015). *The contribution of electricity consumers to peak shaving and the integration of renewable energy sources by means of demand response*. Fraunhofer Verlag and Dissertation: Karlsruhe Institute of Technology (KIT).

Boiteux, M. (1960). Peak-load pricing. *The Journal of Business, 33*, 157–179.

DoE. (2006). *Benefits of demand response in electricity markets and recommendations for achieving them. Technical report*. Washington. U.S. Department of Energy. Available at: https://www.energy.gov/sites/default/files/oeprod/DocumentsandMedia/DOE_Benefits_of_Demand_Response_in_Electricity_Markets_and_Recommendations_for_Achieving_Them_Report_to_Congress.pdf [Accessed May 27, 2022].

Eurostat. (2019). *Supply, transformation and consumption of electricity, Final energy demand by sectors, Table nrg_cb_e*. Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_cb_e&lang=en [Accessed April 19, 2019].

Fraunhofer ISI, IREES, & TEP. (2015). *FORECAST/eLOAD*. Available at: http://www.forecast-model.eu/forecast-wAssets/img/methodology/forecast_tabelle_05.png [Accessed May 5, 2019].

Joskow, P., & Wolfram, C. (2012). Dynamic pricing of electricity. *American Economic Review, 102*, 381–385.

Kaschub, T. (2017). *Batteriespeicher in Haushalten unter Berücksichtigung von Photovoltaik, Elektrofahrzeugen und Nachfragesteuerung Produktion und Energie (23)*. KIT Scientific Publishing and Dissertation: Karlsruhe Institute of Technology (KIT).

Kostkova, K., Omelina, L., Kycina, P., & Jamrich, P. (2013). An introduction to load management. *Electric Power Systems Research*, *95,* 184–191.

McKenna, R., & Norman, J. (2010). Spatial modelling of industrial heat loads and recovery potentials in the UK. *Energy Policy*, *38,* 5878–5891.

McLoughlin, F., Duffy, A., & Conlon, M. (2010). The generation of domestic electricity load profiles through Markov chain modelling. *Euro-Asian Journal of Sustainable Energy Development Policy, 3*.

Möst, D., & Keles, D. (2010). A survey of stochastic modelling approaches for liberalised electricity markets. *European Journal of Operational Research, 207*, 543–556.

Oberascher, C. (2013). *Stromverbrauch und Stromverwendung der privaten Haushalte in Deutschland. Ergebnisse einer Studie im Auftrag von HEA, BDEW und EnergieAgentur.NRW, 2012*. Available at: https://www.yumpu.com/de/document/view/10605406/energieverbrauch-und-energieverwendung-im-haushalt-hea [Accessed May 26, 2022].

Oren, S. (2012). Nonlinear pricing. In Özer, Ö., & Phillips, R. (Eds.), *The Oxford handbook of pricing management*. Oxford University Press. Available at: https://doi.org/10.1093/oxfordhb/9780199543175.013.0022 [Accessed May 27, 2022].

Paardekooper, S., Lund, R. S., Mathiesen, B. V., Chang, M., Petersen, U. R., Grundahl, L., David, A., Dahlbæk, J., Kapetanakis, I. A., Lund, H., Bertelsen, N., Hansen, K., Drysdale, D. W., & Persson, U. (2018). *Heat Roadmap Europe 4: Quantifying the Impact of Low-Carbon Heating and Cooling Roadmaps*. Aalborg Universitetsforlag.

Ströbele, W., Heuterkes, M., & Pfaffenberger, W. (2010). *Energiewirtschaft - Einführung in Theorie und Politik* (2nd ed.). Oldenbourg.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting, 30,* 1030–1081.

Zweifel, P., Praktinjo, A., & Erdmann, G. (2017). *Energy economics—Theory and applications*. Springer.

4 Electricity Generation and Operational Planning

Electricity is a crucial resource for modern societies (see Sect. 2.2). The sustainable provision of electricity is at the same time at the core of the current debates about the energy system of the future. Key questions for designing future electricity systems are thereby:

- Which technologies may be used for electricity generation?
- What are physical and technical characteristics of these technologies?
- What are economic characteristics and potentials of these technologies?
- How may electricity demand be met using a mix of technologies, including notably also variable renewable generations?

These questions will be discussed subsequently, starting with available conventional electricity generation technologies in Sect. 4.1. Then renewable energy technologies are in the focus in Sect. 4.2. Key characteristics for both types of energy technologies are summarised in Sect. 4.3 and the problem of scheduling the different generation technologies to meet demand is discussed in Sect. 4.4, whereas the intermediaries of transport and storage are left for Chap. 5.

Key Learning Objectives
After having gone through this chapter, you will be able to

- Describe the basic principles and the key components of (conventional and renewable) electricity generation technologies.
- Describe the key techno-economic characteristics of power generation and state the magnitude of these characteristics for different generation technologies.
- Understand the scheduling of electricity generation.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_4

- Interpret the role of electricity generation technologies in day-ahead planning.
- Explain the merit-order approach and formulate the corresponding mathematical optimisation problem of plant scheduling.

4.1 Conventional Generation Technologies

One key advantage of electricity is that many technologies may be used for its production. There are technologies available to convert almost any primary energy source into electricity. In the following, first, an overview is given of the major technologies. Electricity generation technologies may be categorised according to different characteristics such as the underlying physical principles or the typical size of installations. We still follow the established distinction by primary energy sources, although electricity generation from biomass (see Sect. 4.2.4) mostly exploits similar technologies as fossil fuels.

4.1.1 Fossil-Fired Technologies

Fossil-based electricity generation systems consist of three main parts: (1) a combustion unit, where fossil energy carriers are burnt to convert the stored chemical energy into heat; (2) a thermal engine which converts the heat into mechanical energy and (3) the electricity generator itself converting the mechanical energy of a rotating shaft into electricity. For the thermal engine, three major types may be distinguished according to the basic thermodynamic processes used: steam turbines, gas turbines and motor engines. Before discussing these in some more detail, first, the combustion process is considered. Subsequently, we also discuss the electricity generator and fuel cells as conversion engines that are not based on combustion. Finally, the scaling up from physical principles to utility-scale units is briefly discussed.

4.1.1.1 Combustion Process

Fossil fuels are composed primarily of carbon and hydrogen (see Table 4.1).

Correspondingly, the main chemical reactions[1] in a combustion process may be described as:

$$C + O_2 \Rightarrow CO_2$$

[1] It should be noted that these are not chemical reactions in a narrow sense. Rather these equations abstract from the chemical equations and just summarize the essential main chemical components.

Table 4.1 Fossil fuels and their most essential components

Fuel	Components (**bold**: main components)
Coal	**Carbone** (**C**) plus varying proportions of water (H_2O), sulphur (S) and other substances (nitrogen (N), hydrogen (H_2), phosphorus (P), trace elements, ...)
Oil	(Various, mostly saturated) **Hydrocarbons** ($\mathbf{C_nH_{2n+2}}$ with n between 6 and 30) plus varying proportions of sulphur (S) and other substances (oxygen(O_2), hydrogen (H_2), metals, ...)
Natural Gas	**Methane** ($\mathbf{CH_4}$) plus varying proportions of ethane (C_2H_6), propane (C_3H_8), etc., as well as nitrogen (N_2), hydrogenic sulphur (HS) and other substances (carbon (CO_2), water (H_2O), ...)

$$4\ H + O_2 \Rightarrow 2\,H_2O$$

Both reactions are exothermal; i.e. energy is released in the form of heat. Given the heterogeneous composition of most fossil fuels and the specifics of the combustion, further combustion reactions are occurring in parallel, notably the sulphur in the fuel is converted to sulphur dioxide SO_2 and sulphur trioxide SO_3, and part of the nitrogen in the combustion air forms nitrous oxides of the general formula NO_x (cf. Sect. 6.2).

4.1.1.2 Steam Cycle and Steam Turbines

The heat produced through the combustion process may be used in steam turbines. A **steam turbine** is part of a closed-cycle thermodynamic process, generally called the steam or water-steam cycle. The key elements of this cycle process are depicted in Fig. 4.1.

Energy enters the process through heat transfer in the boiler (c). In the case of fossil (or biomass) plants, the energy is obtained through the combustion of fuels – they may be solid (coal), liquid (oil) or gaseous (gas). The energy increases temperature and pressure of the fluid and eventually transforms the water into steam. In the so-called p–V diagram (see Fig. 4.2), the pressure–volume diagram, this heating and phase-shifting correspond to the movement from point I to point II. In the turbine, part of the enthalpy contained in the fluid is transferred to the rotating shaft of the turbine (e) – it is transformed into mechanical energy. This leads to a decrease in pressure and temperature with a simultaneous increase in volume (point II–III). The cool, low-pressure steam then enters the condenser (g), where it transfers the remaining energy to the cooling medium (in general water). In the p–V-diagram, this corresponds to the horizontal, i.e. isothermal and isobaric,[2] move from III to IV. The final step to complete the cycle is the pressure increasing pumping (h and b) of the water from the condenser to the boiler.

[2] An isothermal process is a type of thermodynamic process in which the temperature of the system remains constant. Analogical, an isobaric process, is a type of thermodynamic process in which the pressure of the system remains constant.

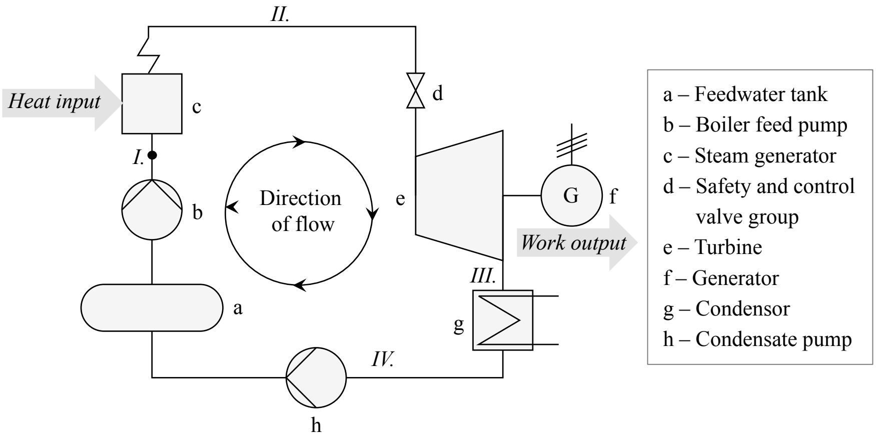


Fig. 4.1 Schematic representation of the steam cycle process of electricity generation

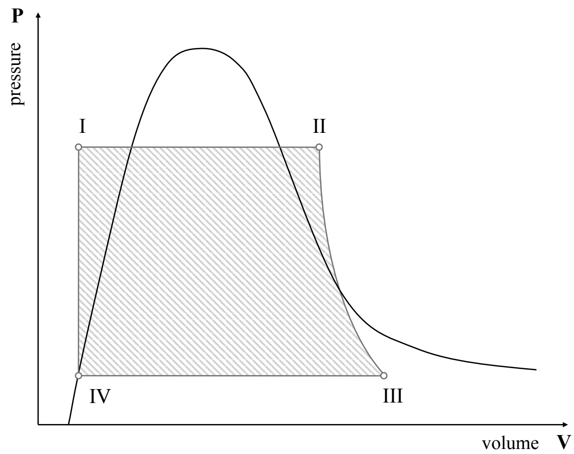


Fig. 4.2 p–V-diagram of the Rankine cycle (steam cycle process)

This description focuses on the key elements in terms of energy transformation. The thermodynamic, ideal reference process, which may be used to determine the achievable process parameters and efficiencies, is called the Rankine cycle or Clausius–Rankine cycle. Further process steps (partly indicated in Fig. 4.1) are needed for the technical operation or to improve overall efficiency.

The efficiency of the process cannot exceed the Carnot efficiency discussed in Sect. 2.1.2, which is the theoretical maximal efficiency of the process depending on the temperature difference. In practice, the Rankine cycle will be more efficient than a process using gas as a working fluid, notably because the mechanical energy needed for compression (IV–I) is small, given that water has a much higher density than steam. Instead of water also other working fluids like ammonia or organic substances may be used. The latter technical option is pursued in the so-called organic Rankine cycle (ORC), transforming lower-temperature heat (100–200 °C) into mechanical or electrical energy.

In large-scale fossil power plants, the Rankine cycle is operated with fresh steam temperatures of up to 600 °C and pressures of 280 bar. Separate steam turbines for high, medium and low pressure successively expand the steam. The combination of several turbines allows to increase the overall efficiency of the process as different temperature and pressure levels can transform heat energy into mechanical energy. The condenser is operated at about 25 °C and a pressure of 0.03 bar, i.e. below air pressure. Such a configuration allows achieving a net efficiency of up to 46% at modern steam power plants. Typical plant sizes range from 10 to 1000 MW_{el} (MW electrical output), with efficiency increasing with larger plant sizes due to lower losses.

To further improve the conversion efficiency and thus lower resource consumption and emissions, the following measures may be taken:

- Use of new materials and alloys to enable an increase of steam temperature and pressure: in fact, the steam parameters are not limited by the combustion process but by the materials' ability to sustain high temperature and pressure over long time periods. Further increases in temperature and pressure levels allow further net efficiency improvements.
- Intermediate reheating of steam and preheating of feeding water which allows to increase the efficiency by lower condenser temperatures respectively a higher average temperature of the energy transferred in the boiler.
- Heat use in combined heat and power (CHP) systems: this does not increase the efficiency of electricity generation but the overall efficiency (see Sect. 4.1.3).

  Besides improvements in energy efficiency, increases in operational flexibility are another primary goal of plant manufacturers' current research and development (R&D) activities. Traditionally, the flexibility of operation is limited by several factors:
- Thermal stress resulting from rapidly changing operation conditions. In operation, ramping rates are therefore limited, e.g. to 2% of the rated capacity per minute.
- Minimum operation times and minimum downtimes are also frequently defined to limit thermal stress.
- Start-up times as a consequence of defined start-up processes and the necessity to heat-up the equipment. For coal plants, these may be 4 h or even more and depend on the status of the plant (starting from cold or heated status).
- Stable operation requirements for grinding mills, pumps and combustion processes. This leads to a minimum stable generation level which for many plants used to be around 50% of the nameplate capacity.

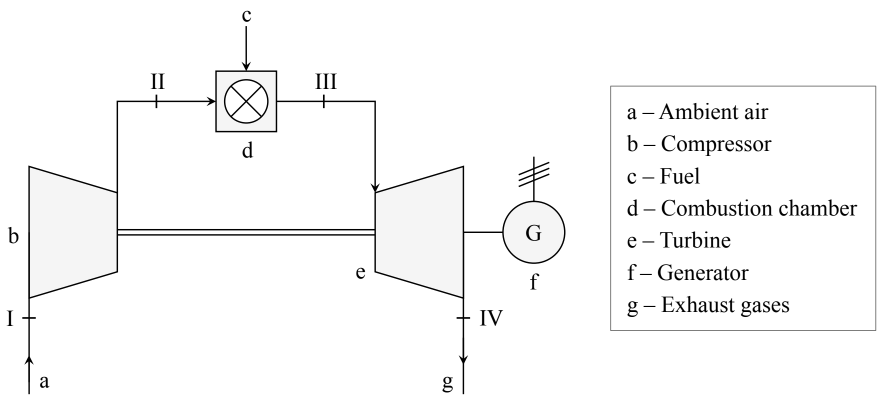


Fig. 4.3 Schematic representation of the gas turbine process for electricity generation

4.1.1.3 Joule Cycle and the Gas Turbine

A **gas turbine** is in its basic process details somewhat similar to a jet engine for aircrafts. The main components of the thermodynamic process are depicted in Fig. 4.3.

The process is "open" (in contrast to the above shown Rankine cycle) since ambient air enters the compressor and exhaust gases leave the process after the turbine through a chimney with possibly some intermediate cooling facilities. Therefore, this type of engine is also frequently called open-cycle gas turbine or **OCGT** for short. The term "gas turbine" refers to the fact that the working fluid is gaseous; it does not necessarily imply that the fuel is natural gas. Rather commercial gas turbines may also run on fuel oil or in the future on biogas (biomethane) or hydrogen. Even coal-fired plants may include a gas turbine: in that case the plant design consists of a gasification stage preliminary to the gas turbine. This concept, known as **IGCC** (**integrated gasification combined cycle**) has been realised in a limited number of industrial large-scale demonstration projects (e.g. in Puertollano, Spain and Priolo, Italy). This plant design is intended to make use of two processes, the conventional steam process (Rankine cycle) and the gas turbine process, to increase overall efficiency.

The thermodynamic reference process for the gas turbine is called the Brayton or the Joule cycle. The compression corresponds in the p–V diagram to an increase in pressure with a simultaneous decrease in volume (I–II), ideally at constant entropy (isentropic). The combustion is isobaric, i.e. occurs at constant pressure, with an increase in temperature and associated increase in gas volume (II–III). The conversion of this thermal energy into mechanical energy occurs in the gas turbine itself. It leads to a decrease in pressure and temperature but an increase in the volume of the combustion gas (III–IV). For the thermodynamic representation, the cycle is closed by the isobaric step from IV–I. The idealised cycle is depicted in Fig. 4.4.

Gas turbines are currently built with an output range of 5–300 MW. Microgas turbines with rated power starting at 5 kW have been intensively researched over the last decades, yet have hardly been a commercial success so far. Modern

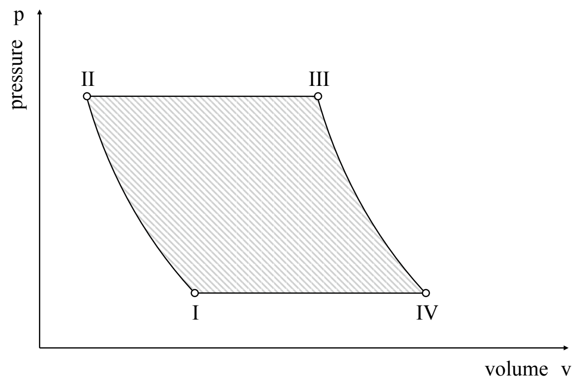


Fig. 4.4 p–V-diagram of the Brayton or Joule cycle (gas turbine process)

large-scale gas turbines achieve an efficiency of 38% (relative to the lower heating value, see definition Chap. 2). The efficiency is limited by the maximum inlet temperature of the gas turbine, which is currently around 1500 °C. For small-scale turbines, efficiency is substantially lower due to higher losses, notably related to a higher share of reverse flows in the compressor.

The maximum turbine inlet temperature effectively limits the attainable pressure ratio and thus the efficiency of the almost isentropic and isobaric processes of compression and combustion. As a consequence, the outlet temperature of the gas turbine is then also in the order of 500 °C. The hot outlet temperature implies a possibility for very substantial conversion efficiency increases: the use of the exhaust gases for heating the steam for a steam cycle process. This is the basic principle of the combined cycle gas turbine (CCGT) units described in the following section. Other possibilities for efficiency increases arise from higher inlet temperatures, possible with new materials such as ceramics.

Operation of gas turbines is much more flexible than operation of power plants with steam cycles. Pure open-cycle gas turbines may be started within less than a quarter-hour and may also ramp their entire operation range within the same time interval. However, a limitation in many designs is the low part-load efficiency, which makes operation at less than the rated output rather unattractive. Furthermore, flexibility improvements similar to those indicated for steam turbines in Sect. 4.1.1.2 are currently strived for.

4.1.1.4 Combined Cycle Gas Turbine

As indicated by the name, **gas combined cycle units** (short: CCGT) consist of a combination of one or several gas and steam turbines. As shown in Fig. 4.5, the fuel is burnt in the gas turbine process and the exhaust gases of the gas turbine are then used to heat the steam in a steam cycle. Whereas the efficiency of a gas turbine may reach up to 38%, the overall efficiency for **CCGT** plants may exceed 60%, based on the lower heating value.

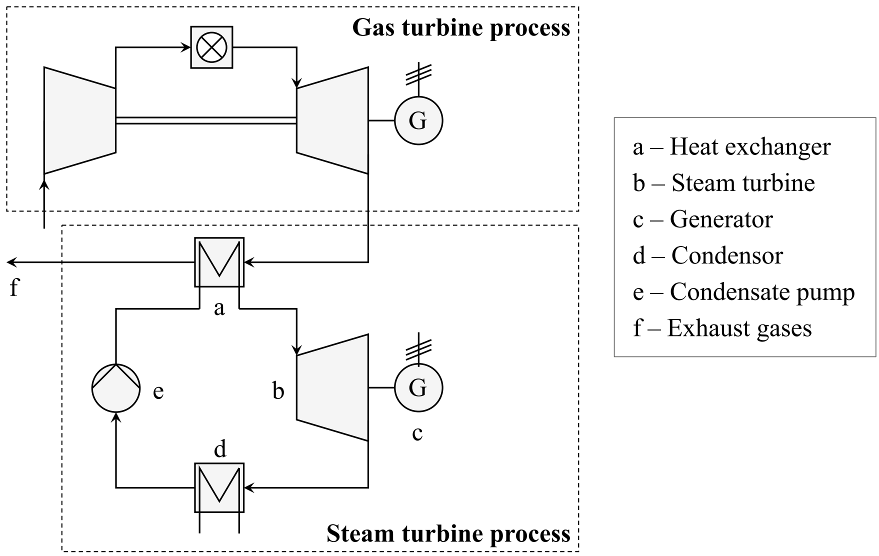


Fig. 4.5 Schematic representation of the combined cycle process for electricity generation

Commercial plants exist with nameplate capacities ranging from 20 to 800 MW and various turbine configurations. In terms of cost efficiency, the single shaft concept seems advantageous: gas turbine, steam turbine and generator are mounted on one single shaft, thus avoiding a separate generator unit for each turbine. Yet this implies that the gas and the steam turbine have to be continuously operated as a single unit. Moreover, a 2-1 configuration is preferable in terms of sizing: two gas turbines provide the heat input for one single steam turbine. Then gas turbines and steam turbines may be scaled to their economically efficient size, providing efficiency and cost advantages.

The operational flexibility of combined cycle units is strongly dependent on the actual plant design. Many of the CCGT units built in the 1990s and early 2000s in Europe were built as baseload units to run on cheap gas or in a carbon-constrained context. The plant layout was then optimised to maximise efficiency at full load. Newer designs and current development efforts aim to provide operational flexibility and high efficiency over a broader range of operation points, along the same lines as indicated for pure steam and pure gas turbines. Also, more frequent start-ups are a design feature important for new developments.

4.1.1.5 Motor Engine

Whereas gas and steam turbines dominate the large-scale electricity generation segment, **motor engines** are the preferred option for generation capacities in the range of 5 kW–2 MW. The newest generation even goes to a unit size of up to 10 MW with a nameplate electrical efficiency of 45%. The engines are similar to car motors, although they are rather designed for continuous operation than for frequent starts, accelerations or other regime shifts. For illustration: a typical car use of 1 h per day corresponds to less than 400 h of operation per year. By contrast, standard stationary motor engines are operating 4000–8000 h per years. Nevertheless, the basic processes are identical; i.e. Otto and Diesel cycles (cf. Granet and Bluestein 2014) are in use. The motor engines may be fuelled with liquid fuels like fuel oil or biogenic oils (e.g. palm oil or rapeseed oil), but in urban areas in Europe, they are more frequently run on natural gas, biogas or landfill gas.

Compared to gas and steam turbines, motor engines offer better possibilities for down-scaling while maintaining reasonable electrical efficiency. Well-established manufacturing concepts also make them cost-effective in terms of costs per capacity unit – at least at engine sizes above 100 kW. Operational flexibility is at least as high as for open-cycle gas turbines, although maintenance expenditures tend to be higher. Engines could also be used as CHP plants (see Sect. 4.1.3) when recovering the exhaust heat.

As with other combustion plants, research and development aim at increasing energy efficiency through higher combustion temperatures and pressures. For applications in residential housing, new designs for small-scale motor engines are looked for. One alternative that has received considerable attention is the so-called Stirling motor. It also has pistons and cylinders; yet combustion is done outside the cylinders. This provides, in principle, advantages in terms of fuel flexibility and, at the same time, allows small engine sizes. Yet still, it remains to be seen whether they may be produced at affordable costs when moving to larger scale production. So the main development focus is on cost reduction in parallel with further efficiency improvement.

4.1.1.6 Electricity Generators

Conversion of mechanical energy to electrical energy is done using the electrical induction principle discovered in the nineteenth century by Faraday and others. A **generator** consists of a rotating part, the rotor, and a stationary part, the stator. In general, two types of electromagnetic generators can be distinguished: dynamos induce pulsing direct current using a commutator, while alternators generate alternating current. With the rotation of the rotor, a time-varying magnetic field induces electrical currents in a conductor, and this effect is amplified if the conductor is wound up to form a coil.

Generators in large-scale power plants are in general so-called synchronous generators; i.e. their rotation speed is proportional to the frequency of the alternating current (AC) in the grid (see Sect. 5.1). The alternating current is induced in the coils of the outer, fixed part of the generator, the so-called stator (see Fig. 4.6). It results from the rotating magnetic field in the inner part, the rotor. For small-scale units, permanent magnets may be used for the rotor; yet in large-scale power plants, the magnetic field is generated by electromagnets, the so-called field coils. Synchronous generators primarily used in power plants can deliver pure active power or supply reactive power, which is required to compensate for inductive and capacitive loads. Thus, they can serve as active phase shifters in electrical power supply networks.

Apart from PV and fuel cells, generators are ubiquitous for electricity generation for most energy carriers, including fossil, nuclear and most renewables: wind, biomass, solar thermal, geothermal, hydro, tidal, wave, etc.

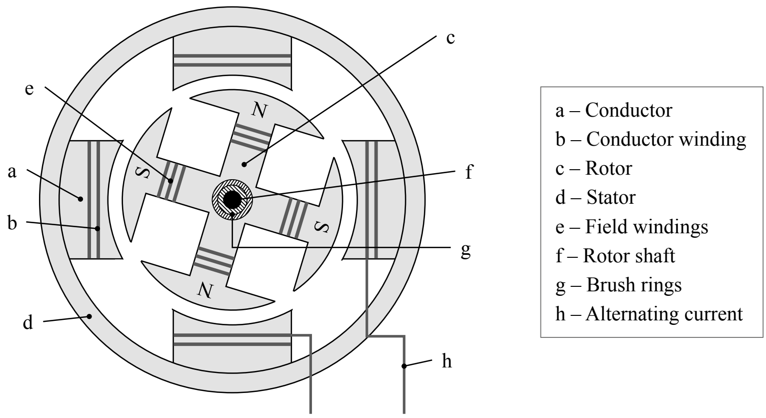


Fig. 4.6 Cross section of an electricity generator

4.1.1.7 Fuel Cells

As opposed to all the technologies discussed so far, fuel cells do not rely on combustion. Instead, they directly convert the chemical energy stored in the fuel to electricity in reversing the electrolysis process. The main components of a fuel cell are the two electrodes – cathode and anode, and an electrolyte (Fig. 4.7). The electrolyte selectively transports ions from one electrode to the other, e.g. protons from the anode, which is the negative pole of the fuel cell, to the cathode. There they recombine with oxygen, previously reduced by the electrons transported through the electrical system. Hence the basic reaction equations for a hydrogen fuel cell are:

$$\text{Anode}: \ H_2 \Rightarrow 2\,H^+ + 2\,e^-$$

$$\text{Cathode}: \ O_2 + 4\ e^- \Rightarrow 2\ O^{2-}$$

$$2H^+ + O^{2-} \Rightarrow H_2O$$

The electrolyte, e.g. a membrane, ensures that the fuel and the oxidant (oxygen or ambient air) do not recombine directly.

By avoiding the intermediate conversion of chemical energy into thermal energy, fuel cells have a higher efficiency than conventional power plants as they circumvent the limitations of the Carnot efficiency for the conversion of heat into power. Hence, in principle, very high electrical efficiencies may be obtained. Furthermore, the concept is scalable since a single cell typically delivers a voltage of 0.6–0.7 V. Thus, they have to be combined to form larger stacks. Yet, various practical issues make this appealing concept rather complicated for stationary and mobile applications.

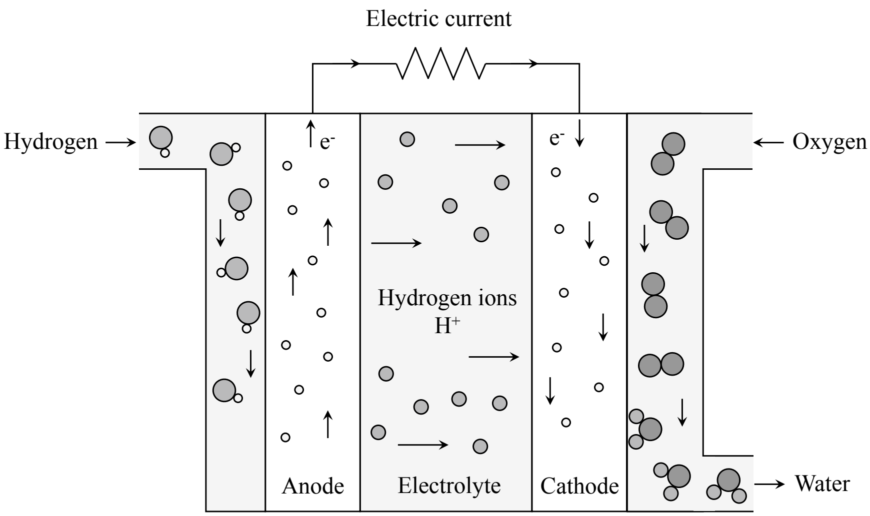


Fig. 4.7 Schematic representation of a fuel cell for electricity generation

Research, development and early commercialisation currently focus on four different types of fuel cells:

PEMFC – polymer fuel cells: these fuel cells are also called PEMFC—proton exchange membrane fuel cells, since they use a polymer membrane as electrolyte. They operate at low temperatures (40–80 °C) and are very flexible in operation. Yet, they require pure hydrogen as fuel. Unless this fuel is immediately available, the hydrogen conversion reduces the overall electrical efficiency so that at a plant level approximately 40–50% are achieved.

PAFC – phosphoric acid fuel cells: these fuel cells were among the first to be commercialised on a small scale. They use phosphor acid as an electrolyte which leads together with the operating temperature of 150°–200 °C to considerable corrosion problems. Commercialisation of these cells has been stopped in the early 2000s, as plant efficiency has been relatively low with 38%.

SOFC – solid oxide fuel cells: these are high-temperature fuel cells based on a ceramic as electrolyte. Here oxygen O^{2-}ions are transported through the electrolyte. The operating temperature is in the range of 800–1000 °C, which implies limited operational flexibility. Yet, on the other hand, this fuel cell type does not require pure hydrogen as fuel but may also run on methane, i.e. natural gas, which is then internally reformed. The electrical efficiency at the plant level is in the range of 55–60%.[3]

[3] The fuel cell efficiency maybe around 50%. However, because of a high temperature, a gas turbine could be connected to a SOFC, thus increasing the plant efficiency beyond 70%.

MCFC – molten carbonate fuel cells: also the MCFCs are high-temperature fuel cells based on carbonate ions CO_3^{2-}. Like SOFCs, the start-up is slow due to the necessary preheating (operating temperature around 700 °C). Plant-level efficiencies are in the range of 50%.

After high growth expectations around the turn of the millennium, prospects for fuel cells have been rather bleak for about a decade. Yet, in recent years, the commercialisation of fuel cells has expanded considerably, with first applications in vehicles and home-size CHP systems.

4.1.1.8 Large-Scale Fossil-Fired Power Plants

Power generation in practice requires industrial processes, which are considerably more elaborate than the basic physical processes sketched in the previous sections. Figure 4.8 summarises the key material flows in a large-scale **fossil-fired power plant**. The combustion process and the steam cycle are in the centre of the graph; yet here multiple steam turbines are included. Typically, large-scale power plants have two to three different steam turbines—from high pressure to low pressure. This design allows intermediate reheating, which generally improves efficiency. Moreover, some steam may also be diverted from the steam cycle between the high and medium pressure turbines to produce high-enthalpy steam and heat. This is notably important for combined heat and power generation (see Sect. 4.1.3). Besides these core elements, the cooling apparatus (in the bottom part of the figure) and the post-treatment of the combustion gases (upper part) have to be considered. When a cooling tower is present, this is typically the most prominent building on the plant site. Furthermore, the post-treatment of exhaust gases may occupy more than one-third of the entire area. The post-treatment in modern coal-fired power plants includes dust cleaning, desulfurisation and denitrification (see Sect. 6.2.2.3). The desulfurisation avoids emissions of SO_2 (and some SO_3) and leads to a solid by-product, plaster gypsum, which is partly used in the building sector. The remaining quantities and the fly dust from dedusting have to be deposited in landfills. By contrast, denitrification, mostly done via a selective catalytic reduction (SCR), has molecular nitrogen and water as final products. Especially when assessing the environmental impact of power generation, it is crucial to consider the multiple auxiliary units and processes making up a real-world power plant. This becomes evident when key material flows (as shown in Fig. 4.8) with their mass balances are considered.

4.1.2 Nuclear Energy

Like most electricity generation based on fossil fuels (see Sect. 4.1.1), nuclear power plants also consist of the three parts: (1) heat source (analogue to the combustion unit of a fossil-fuel based plant), (2) thermal engine and (3) electricity generator. But in contrast to steam power plants fired by fossil fuels like coal or gas, the needed heat in nuclear power plants comes from splitting heavy atoms (so-called fission). Nuclear power reactors in commercial operation are exclusively

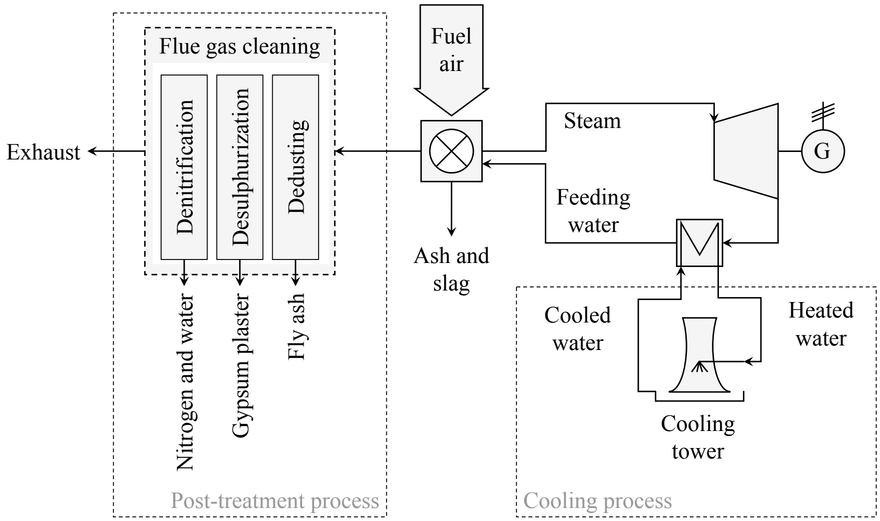


Fig. 4.8 Key material flows in a large-scale fossil power plant

nuclear fission reactors, while nuclear fusion reactors are still under research and development. In nuclear fission reactors, primarily the uranium isotope uranium-235 (^{235}U) is used as fuel input, as it can relatively easily be fissioned.

The **nuclear fuel cycle** (cf. e.g. Murray and Holbert 2020, p. 440) can be differentiated into the front end and the back end. The front end consists of mining and milling, conversion, enrichment and fabrication (cf. Larson 2019). The first step is mining and milling as uranium occurs on earth (e.g. in Canada and Australia) in the form of uranium ore, which first has to be extracted. Natural uranium is mainly composed of ^{238}U (about 99.3%[4]), ^{235}U only has a share of about 0.7%. The extracted uranium ore is then treated chemically, e.g. with acids. This treatment results in a mixture of different uranium oxides (so-called yellowcake), mainly triuranium octoxide, U_3O_8. The yellowcake is then transported to a conversion factory to produce UF_6 (uranium hexafluoride). UF_6 is then enriched in an enrichment plant to increase the concentration of ^{235}U to about 4%, which is necessary for the use in a light-water reactor. The final step in this supply side of the nuclear fuel cycle is the manufacturing of uranium dioxide (UO_2) filled in fuel rods used in nuclear power plants (cf. Zweifel et al. 2017, p. 251).

The principle of a nuclear fission reactor is to bombard ^{235}U with neutrons to split it into lighter atoms. By splitting such nuclei, the so-called mass defect can be used for energy generation (see Fig. 4.9). Due to the lower mass of the produced nuclei, energy will be delivered according to the formula[5] $E = mc^2$. The energy released is used to produce steam, and with this steam turbines are operated.

[4] As the nucleon number already shows, ^{238}U has three neutrons more in the atomic nucleus than ^{235}U.

[5] Note that c here stands for the speed of light, whereas m is the mass and E is the energy.

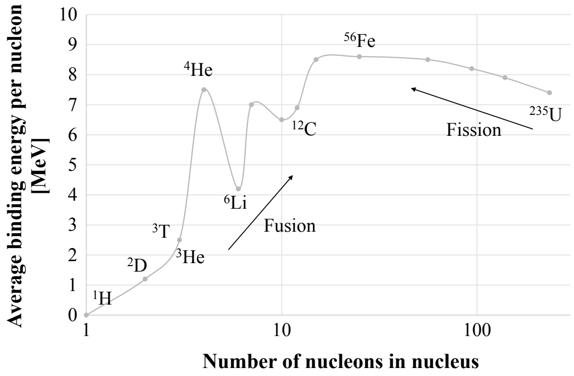


Fig. 4.9 Binding energy per nucleon. *Source* Own illustration based on Pistner (2012, p. 28)

Besides releasing kinetic energy (and radiation), free neutrons are unleashed by the fission process. These neutrons will then partly be absorbed by other atoms, resulting in new fission processes – the so-called nuclear chain reaction has started. The control of this chain reaction is realised with the help of so-called control rods by absorbing neutrons. The nuclear fuel is typically used for some years (about 3–5 years) in the reactor, and then the concentration of ^{235}U has become too low (below 1%). Hence, during the yearly fuel replacement, one-third to one-fifth of the used uranium is exchanged every year, typically during times with relatively low electricity demand (e-g. in summer months).

The spent nuclear fuel is highly radioactive and still producing heat; here, the back end of the nuclear fuel cycle starts. Due to its radioactivity, the spent nuclear fuel has to be stored safely for a long period. Nowadays, the disused fuel is first stored in (storage) pools close to the reactor and then (after some years) in dry casks [e.g. in casks for storage and transport of radioactive material (CASTOR)] in interim storage facilities directly located at the nuclear power stations. There is the possibility of realising a challenging reprocessing process to reuse parts of the nuclear waste (e.g. plutonium and ^{235}U) and produce mixed oxide (MOX) fuels. Without reprocessing, the spent nuclear fuel directly has to be transported to a final disposal facility, capable of storing this nuclear waste for thousands of years, e.g. somewhere in the underground. In this context, it has to be mentioned that up to now, no final storage facility for this kind of waste exists worldwide.

Besides splitting heavy atoms, the mass defect can also be used for energy generation by fusing light atomic nuclei like hydrogen (so-called **fusion**), which is the process with the help of which the sun generates energy for billions of years (see Sect. 4.2.3). To realise such a fusion process, the electrostatic force, which pushes the atomic nuclei apart, must be overcome. Therefore, extremely high temperatures and extremely high pressure are needed. Additionally, the so-called tunnel effect helps to increase the probability of the fusion of nuclei. There is still a lot of research to pursue to realise nuclear fusion on earth. Significant challenges are to reproduce conditions like inside the sun, e.g. to realise a stable plasma operation with a temperature of millions of degrees Celsius. Some prominent research institutions have succeeded in very short reactions at a small scale. Currently, an International Thermonuclear Experimental Reactor (ITER)[6] is under construction near to Cadarache, France, to demonstrate the scientific and technological feasibility of fusion energy. Since research, demonstration and implementation of nuclear fusion have been delayed several times, the promise that fusion is "always 50 years away" is valid for several decades.

4.1.2.1 Power Plant Technologies

Nuclear energy is used for electricity production commercially since the 1960s. Over the last sixty years, light-water reactors,[7] using enriched uranium as input, became the dominating technology to generate electricity by fission. Light-water reactors can be further subdivided into pressurized water reactors (PWR) and boiling water reactors (BWR).

The **pressurized water reactor** has a central characteristic: there are different circuits to prevent water from flowing through the reactor to reach the turbine (see Fig. 4.10). For cooling and as a moderator, which is needed to slow down the neutrons, ordinary water is used. In the core of the reactor, the water is heated up to temperatures of about 320 °C under a pressure of about 150 bar (Murray and Holbert 2020, p. 327). This heated water is used to produce steam with the help of steam generators. Due to the fact that there is no steam in the upper part of the reactor,[8] the control rods can be mounted at the top in PWRs, which comes along with the advantage that the rods may enter the core of the reactor by gravity if there is an interruption of power supply (cf. e.g. Murray and Holbert 2020, p. 331).

The second most common technology used for electricity production based on the fission process is the boiling water reactor. This kind of nuclear power plant has the characteristic that the water, which is again moderator and cooling medium, is also used as the steam source for the turbine, resulting in radioactivity reaching the turbine (see Fig. 4.11). Compared to pressurized water reactors, the temperature (about 290 °C) and pressure (about 70 bar) (Murray and Holbert 2020, p. 327) are lower; thus, the construction is easier and no steam generators are needed. Despite these differences the efficiency of both types – BWRs and PWRs – is rather similar and at about 33% (Lamarsh and Baratta 2001, pp. 140 and 147).

With a share of about 70% of the worldwide capacity of nuclear power plants, (light water) pressurized water reactors are the dominating plants (see Table 4.2). The typical capacity of such a nuclear power unit is approximately between 800 and 1200 MW, the capacity of one plant might even be higher as nuclear power plants

[6] https://www.iter.org/, accessed 13th May 2022.

[7] Light-water reactors use normal water, while heavy-water reactors use heavy water. Heavy water is water that contains essentially deuterium (^{2}H or D, also called heavy hydrogen). In contrast most of the hydrogen in normal water consists of the hydrogen-1 isotope (^{1}H or H, also called protium).

[8] The effect of inserting the rods in a steam area is lower compared to inserting them in a water area (Lamarsh and Baratta 2001, p. 147).

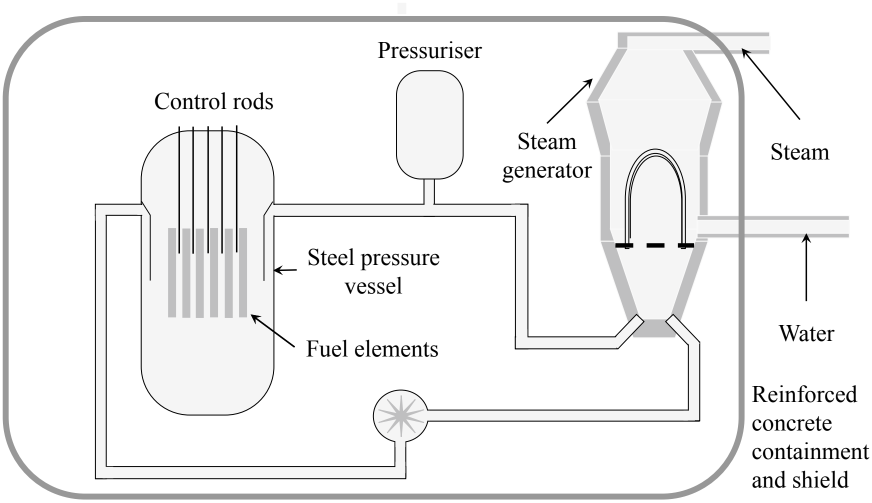


Fig. 4.10 Schematic representation of a typical pressurized water reactor. *Source* Own illustration based on World Nuclear Association (2018)

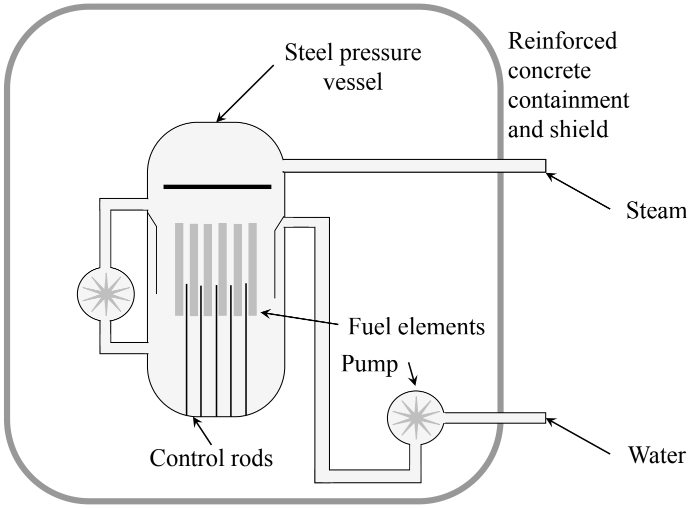


Fig. 4.11 Schematic representation of a typical boiling water reactor. *Source* Own illustration based on World Nuclear Association (2018)

often consist of two or even more units. Besides the dominating technologies (PWRs and BWRs) fast breeder reactors, where simultaneously to electricity production plutonium fuel is breeded, graphite-moderated reactors and pressurized heavy-water-moderated and cooled reactors, where natural uranium is used as fuel and heavy water as a moderator, can be found worldwide. According to Schneider et al. (2020, p. 21), nuclear energy had a share of about 10% of the world's gross

Table 4.2 Operational reactors by type

Reactor type	Reactor type descriptive name	Number of reactors	Total net electrical capacity (GW)	Relative share in capacity (%)
BWR	Boiling light-water cooled and moderated reactor	63	64	16.2
FBR	Fast breeder reactor	3	1.4	0.4
GCR	Gas-cooled, graphite-moderated reactor	14	7.7	2.0
LWGR	Light-water-cooled, graphite-moderated reactor	12	8.4	2.1
PHWR	Pressurized heavy-water moderated and cooled reactor	49	24.5	6.2
PWR	Pressurized light-water moderated and cooled reactor	303	288	73.1
Total		**444**	**394**	

Source IAEA PRIS (2021)

electricity production and of about 4% of the world's primary energy consumption in 2019.

Actual developments to improve nuclear power plants are focusing on improvements with regard to standardisation, efficiency and safety (so-called third-generation reactors) as well as on small modular reactors; long-term developments try to develop closed fuel cycles and to minimise nuclear waste (Generation IV) (cf. e.g. Murray and Holbert 2020, pp. 339–345).

4.1.2.2 Environmental Effects and Risks of Nuclear Technologies

From an environmental point of view, the main advantage of nuclear power plants is that there are no local air pollutions or CO_2 emissions as long as the plant is operated under normal operating conditions – disregarding the emissions arising during the construction of the plant. On the other hand, the operation of nuclear power plants leads to diverse forms of (radioactive) waste. Radioactive waste emits different types of radiation, representing a threat for human beings and the whole environment (for more details, cf. e.g. Krieger 2019).

Environmental effects can already be seen during the mining process, as in the mining areas, a lot of excavation material is produced. The uranium concentrations in the deposits worldwide tend to be relatively small. But also the other steps of the fuel cycle have significant environmental effects; e.g. the enrichment process leads to a lot of depleted uranium (so-called tails). The quantity of depleted uranium, for which also long-term storage is needed, is many times higher than the quantity of the enriched uranium – in two material balances for a nuclear reactor (cf. Murray and Holbert 2020, p. 443), the amount of depleted UF_6 is about 7–9 times higher than the quantity of enriched UF_6. Spent nuclear fuel is often referred to as the major source of environmental effects. According to the German Association of Energy and Water Industries, a German nuclear power plant produces an amount of

20–25 t of spent nuclear fuels per year, which corresponds to 0.0021–0.0027 g per kWh of electricity produced (cf. BDEW 2019, p. 37). But radioactive waste not only results from the used nuclear fuel but also from contaminated parts of the nuclear power plant – e.g. from components located within the primary circuits of the pressurized water reactors.

Radioactive waste can be classified into different types. The classification might vary from country to country, but often a differentiation, based on the radioactivity, into low-level waste (LLW), intermediate-level waste (ILW) and high-level waste (HLW) can be found. The waste submitted for final disposal (long-term storage) is differentiated into waste with negligible heat generation (LLW and partly ILW) and into heat-generating waste (partly ILW and HLW) (cf. Working Panel Waste Management 2012, p. 8).

To have sufficient financial resources to pay the nuclear liabilities caused by the deconstruction of nuclear power plants, the storage of components and the used fuel, energy companies have to set up accruals. Due to the many uncertainties coming along with this task (e.g. at what point in time will which amount of money be needed?) and due to the huge gap between the point in time, when setting up accruals, and that of the real cash flows, the calculation of accruals for nuclear liabilities is a challenging task.

Besides environmental effects that already take place under normal operating conditions, the use of nuclear energy is connected with different risks (Zweifel et al. 2017, pp. 256–258). First of all, there is the risk that a severe accident during the nuclear power plant operation leads to an uncontrolled release of radiation into the atmosphere. Nuclear accidents are measured with the help of the so-called International Nuclear and Radiological Event Scale (INES), which was developed by the International Atomic Energy Agency (IAEA), differentiating seven levels of severity. Up to now, two accidents at nuclear power stations have been classified at level 7 as "Major Accidents": the Chernobyl disaster in Ukraine in 1986 and the Fukushima Daiichi nuclear disaster in Japan in 2011 (Giraldo et al. 2012, p. 53). Besides, there is the risk that nuclear fuels are abused, e.g. by terrorists, or that nuclear power stations become the target of terrorist attacks. Furthermore, different risks come along with the final disposal of radioactive waste: there is a high uncertainty and related risk of how to ultimately dispose radioactive waste for thousands of years, resulting in the fact that so far, no repository for high-level waste is in operation worldwide. Nowadays, deep geological repositories seem to be the preferred solution for this problem. Most probably, a long-term disposal facility of this form will come into operation in the next years in Olkiluoto, Finland.

To calculate the expected monetary damage (e.g. per kilowatt-hour electricity produced), so-called probabilistic safety analyses of nuclear power plants are performed (cf. Zweifel et al. 2017, pp. 258–267). Thereby, for different scenarios, the probabilities of damage events are multiplied by the monetary damage. It has to be mentioned in this context that despite all efforts to improve the security of the plant operation, e.g. the provision of redundant safety systems (cf. e.g. Giraldo et al. 2012, pp. 58–61), the probability of occurrence is indeed very low, but will never become zero – risks are estimated between 1:33,000 and 1:10,000,000.

Furthermore, such calculations have to face the problem that for many damages (e.g. loss of life, environmental damage, etc.) an economic value is difficult to estimate.

4.1.3 Combined Heat and Power Generation (CHP)

The combined production of heat and electricity is frequently labelled by the abbreviation CHP or the term cogeneration. It is, in principle, a desirable means to enhance the overall efficiency of the energy conversion process. The Carnot efficiency limits the efficiency of electricity generation in thermal power plants (see Sect. 2.1). The usage of the waste heat is, on the other hand, only limited by the usefulness and usability of the heating energy which are dependent on the temperature level of the heat and the heat sinks in the vicinity of the power plant. Whereas heat demand, in general, is discussed in Sect. 3.2, the focus is here on the main technical characteristics of CHP generation systems, but also relevant heat demand patterns for CHP are addressed at the end of this subsection.

In principle, cogeneration may be applied to all thermal[9] power generation technologies and is even an option for fuel cells, given that their operating temperature is also considerably above the ambient temperature. Yet up-to-now, large-scale cogeneration has mostly been implemented with steam cycle-based power plants including CCGTs. By contrast, CHP solutions at intermediate and small scale are frequently based on motor engines or industry-scale gas turbines (size 5–100 MW). For the usage of the different options, it is furthermore essential to consider the application context of CHP solutions – i.e. the heat demand.

4.1.3.1 Large-Scale Cogeneration

Large-scale electricity generation is usually done based on the steam cycle (see Sect. 4.1.1). The simultaneous production of heat then implies that part of the energy contained in the steam is diverted for heat production. The heat collected in the condenser of conventional power plants does not have the temperature level necessary for almost any heat application. Moreover, its transport would require vast amounts of transportation fluid.

Two basic construction principles are possible for steam turbines with simultaneous heat and power production: **backpressure turbines** and **extraction-condensing turbines**.

Backpressure turbines: these turbines are similar to conventional steam turbines, only the pressure level at the rear end of the turbine is higher and correspondingly the temperature. Consequently, less heat energy is transformed into electricity and the remaining heat may be transferred through a heat exchanger to the district heating or process heating system (see Fig. 4.12, left part). The electricity produced is (at given steam parameters) almost proportional to the steam mass flow, as is the

[9] Be aware that CHP is not limited to conventional energy carriers as also renewable energy carriers, for example biomass, are converted to thermal energy.

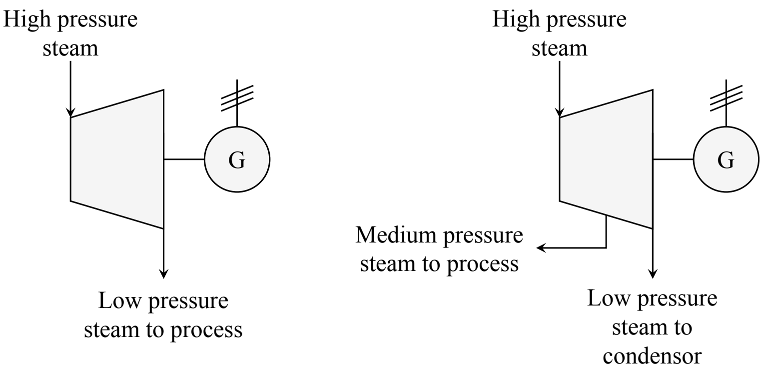


Fig. 4.12 Steam turbines for combined heat and power: backpressure turbine (left) and extraction-condensing turbine (right)

useful heat generated. Hence, in this case, there is an (almost) constant **ratio of heat-to-power**.[10]

Extraction-condensing turbines: in these turbines, part of the steam may be flexibly extracted, e.g. through openings in the casing at higher pressure levels (see Fig. 4.12, right part) or through extraction of part of the steam between the high pressure and the medium pressure turbine. The remainder of the steam is expanded as in conventional condensing turbines to below atmospheric pressure and corresponding low temperatures. Yet, the extracted steam may provide heat at useful temperatures via a heat exchanger to a district heating or process heating system. The advantage of this technology is its flexibility in the heat-to-power ratio.

The so-called **PQ chart** is used in thermal power plant engineering to characterise joint heat and power production possibilities and limits. P stands here for the electric power output and Q for the heat output of a plant. Given the fixed heat-to-power ratio, the permissible operation range for a backpressure turbine is depicted by a line segment in the PQ chart (see Fig. 4.13, left part). For an extraction-condensing turbine, the operation range has the form of a polygon (see Fig. 4.13, right part), the possible operation modes are limited by the maximum steam flow (line 1), the maximum generator output (line 2) and the maximum heat transfer capacity of the heat exchanger (line 3). Moreover, the lower limit of power generation – when all steam is extracted from the turbine – is given by line 4, the so-called backpressure line. Line 5 finally represents the minimum stable operation limit of the power plant (minimum steam flow).

4.1.3.2 Medium- and Small-Scale Cogeneration

In the case of medium-scale cogeneration (range of 2–100 MW), backpressure and extraction-condensing steam turbines may also be employed, notably in the context of **CCGT** units. But the smaller the demand, the more advantageous (notably in cost terms) is it to use only an open-cycle gas turbine together with a heat recovery

[10] Note that sometimes also the inverse ratio, the **power-to-heat ratio** is computed.

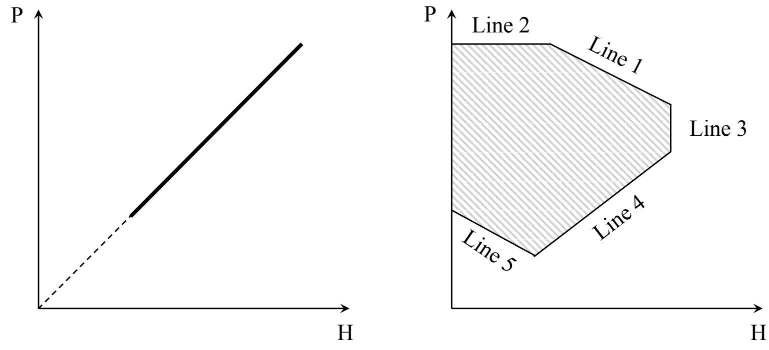


Fig. 4.13 PQ charts for combined heat and power: backpressure turbine (left) and extraction-condensing turbine (right)

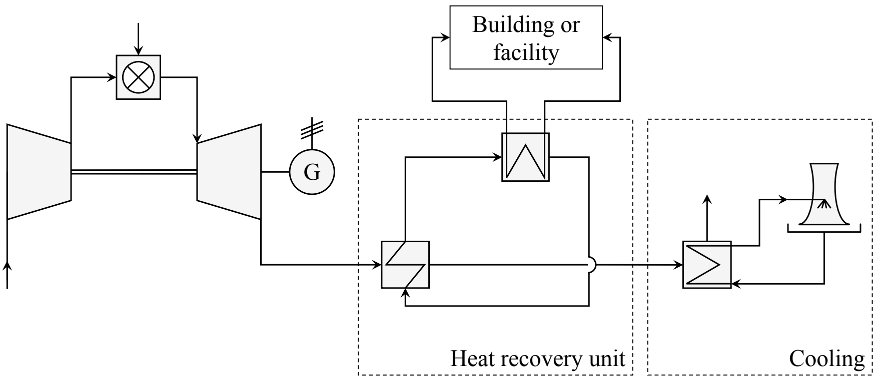


Fig. 4.14 Gas turbine with heat recovery boiler for combined heat and power

boiler, which allows capturing the heat contained in the flue gases of the turbine (see Fig. 4.14). Given the high exhaust temperatures of the gas turbine, such a solution also enables the provision of process heat at temperature levels up to several hundred degrees centigrade.

The flexibility of the combined heat and power production depends on the possibility to operate the gas turbine without the **heat recovery boiler**, i.e. whether sufficient cooling is available when the heat recovery boiler is bypassed. Without that flexibility, the heat-to-power ratio is fixed as for backpressure turbines (see Fig. 4.15 left); otherwise, flexible ratios may be possible. As opposed to steam turbines, increased heat use will then not induce any loss in power output.

4.1.3.3 CHP Based on Motor Engines and Further Technologies

For a long time, motor engine-based CHP has been an alternative for power ranges below the one of gas turbines. But in recent years, motor CHP engines have been scaled up and the largest reach 10 MW electric output (see Sect. 4.1.1.5). But they

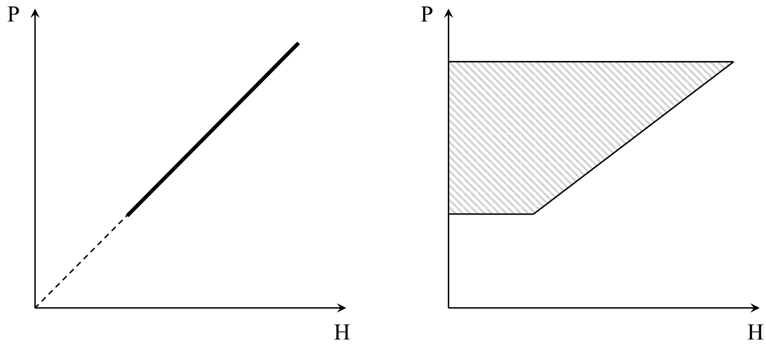


Fig. 4.15 PQ charts for combined heat and power with open-cycle gas turbine: without auxiliary cooling (left) and with auxiliary cooling (right)

may also be built in the size of (small) vehicle engines with an output of 5 kW_{el}. As with gas turbines, waste heat of the engine is used for heating purposes. The heat in the cooling water and the flue gases is typically used for heat sinks with temperatures up to 70–90 °C.

Besides the technologies above, also others may be used for CHP. In the last decade, fuel cells and Stirling engines have notably been developed and commercialised in small series. They have mainly been designed to provide combined electricity and heat at low nameplate capacities with a prime application field being small residential buildings, notably single-family dwellings. Such CHP units are technologically much more demanding than a conventional heating system, such as a condensing boiler, and correspondingly they are also much more expensive. At the same time, their electric capacity is very limited, notably in the case of Stirling engines, so that they are sometimes labelled "electricity-generating boilers". The economic viability of such solutions strongly depends on the difference between wholesale and retail electricity prices.

There are various technologies and application areas for CHP, which vary according to the size and temperature of the heat demand, the economic sector and the generation technology. While several technologies are applicable over a broader range of sizes and applications, some such as Stirling engines are more suited to smaller-scale applications. An overview of the specifications of commonly employed generation technologies for small- and medium-sized systems is given in Table 4.3, from which it is clear that these technologies differ in their suitability for part-load operation as well as their stage of technological development. Some technologies, such as fuel cells, do not yet have high market penetration and suffer from high costs.

Table 4.3 Overview of typical generation technologies for medium- and small-scale cogeneration

	Gas/petrol combustion engine	Diesel engine	Stirling engine	Fuel cell	Gas turbine
Electrical power [kW]	1–5000	5–20,000	1–40	1–250	30–250,000
Overall efficiency [%]	up to 90	up to 90	up to 85	up to 90	up to 85
Electrical efficiency [%]	25–42	28–44	10–30	30–47	25–30
Part-load behaviour	Good	Good	Less good	Very good	Less good
State of the art	Proven	Proven	Small series	Pilot systems	Proven
Usual fuel	Gas, petrol	Diesel	Gas, wood	Gas	Gas, diesel

Source Pehnt and Schneider (2010, p. 124)

4.1.3.4 Heat Demand and Combined Heat and Power Plants

Different technologies can be used to satisfy the heat demand (see Chap. 3). Whereas central (e.g. biomass-fired district heating stations) and decentral (e.g. gas-fired calorific value boilers in households) heating (heat-only) installations are not part of this book, the combined production of heat and electricity (CHP or cogeneration) in large-scale, intermediate-scale and small-scale installations are. Yet it is primordial to match the cogeneration technology to the local heat demand. Therefore, some aspects are subsequently discussed which must be considered when planning the dimensioning and operation of a CHP installation.

CHP can be used to provide all three different kinds of heat: space heat, hot water and process heat. One of the most important contextual factors for applying CHP solutions is the distribution of heat demand in space and time. Weather conditions characterise the provision of space heat. Heat demand in systems with space heating has a clear seasonal pattern, although there are substantial stochastic variations like early snow in October. The sorted annual **duration curve** of heat demand provides a graphical means to describe and analyse the heat demand over time. Figure 4.16 gives an illustration for an exemplary municipal district heating grid.

The peak heat demand there exceeds the minimum demand by a factor of ten or more. The **capacity factor** of the heating installations (energy production during a period, e.g. one year, divided by the maximum possible energy production during that period, cf. Sect. 2.1.1) is frequently below 0.4, corresponding to only 3000–3500 **full-load hours** per year (capacity factor multiplied by the hours of the year). In single buildings or households, the heat provider has to cover peak demand, while in larger networks, peak load boilers are also used in addition to the CHP units.

When providing process heat by CHP the continuity of heat demand and the temperature level of the demand are decisive: some industrial processes, especially in the primary conversion sector, require constant heat supply, which might provide

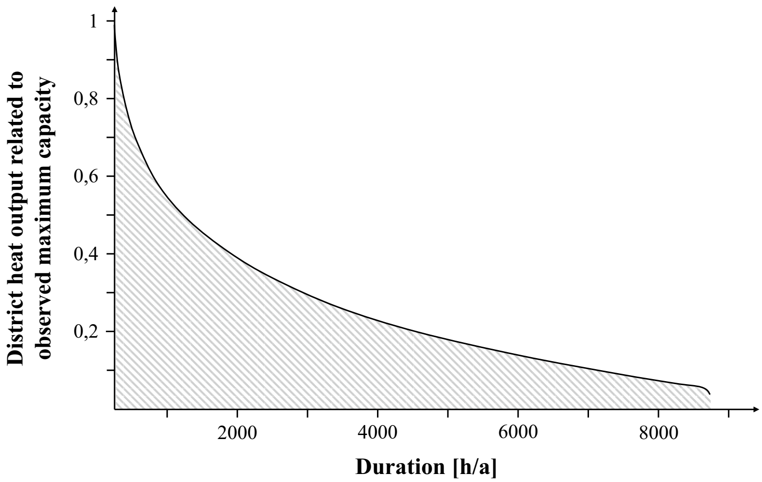


Fig. 4.16 Sorted annual duration curve of heat demand in an exemplary municipal district heating grid

opportunities for high capacity factors. Other processes are organised as batch processes. Here intermediate heat storage is necessary if CHP is to reach high capacity factors. Even more important is the temperature level for the heat provision from CHP. Process temperatures of 500 °C and more are hardly suitable for CHP applications. At best, the exhaust gases of open-cycle gas turbines may provide heat at temperatures of about 500 °C. Advantageous are process heat requirements at temperatures below 120 °C, since hot water may then be used as heat transfer medium. At temperature levels between 120 °C and 200 °C, CHP may offer potential, but then steam will usually be used for heat transfer.

Furthermore, the possibilities to market the produced electricity must be analysed to determine the optimal investment and production plan for a CHP installation. Here, the opportunity to feed in electricity into the electrical grid often provides a kind of backup solution if the customers' electrical and thermal demand profiles do not match adequately.

4.2 Renewable Generation Technologies

Electricity can be produced from various renewable sources, including wind, solar, hydro, tidal, geothermal and biomass. Biomass contributes most to the total renewable energy supply (see also Table 4.4), which is due to its high share in heat supply. Hydropower has a long tradition and is currently still the second-largest renewable power source in Europe. However, there has been no growth in hydropower electricity generation in the last years (as most potentials are already in

Table 4.4 Share of renewables in gross inland (primary) energy consumption, 2016 in %

	Renewable energies – total	Bioenergy[a]	Hydroenergy	Wind energy	Solar energy	Geothermal energy
EU-28	13.2	8.6	1.8	1.6	0.8	0.4
Belgium	6.8	5.4	0.1	0.8	0.5	0.0
Bulgaria	10.7	7.2	1.9	0.7	0.8	0.2
Czech Republic	10.3	9.3	0.4	0.1	0.5	0.0
Denmark	28.7	21.7	0.0	6.3	0.7	0.0
Germany	12.3	8.2	0.6	2.1	1.2	0.1
Estonia	15.5	14.7	0.0	0.8	0.0	0.0
Ireland	7.5	3.4	0.4	3.6	0.1	0.0
Greece	10.7	4.7	2.0	1.8	2.2	0.0
Spain	14.3	5.6	2.6	3.4	2.6	0.0
France	9.9	6.6	2.1	0.7	0.3	0.1
Croatia	23.3	15.2	6.9	1.0	0.1	0.1
Italy	16.8	8.5	2.4	1.0	1.4	3.6
Cyprus	5.9	1.7	0.0	0.8	3.3	0.1
Latvia	37.2	32.0	5.0	0.3	0.0	0.0
Lithuania	20.8	18.7	0.6	1.4	0.1	0.0
Luxembourg	5.3	4.6	0.2	0.2	0.3	0.0
Hungary	11.7	10.8	0.1	0.2	0.1	0.5
Malta	3.4	1.3	0.0	0.0	2.1	0.0
Netherlands	4.7	3.5	0.0	0.9	0.2	0.1
Austria	29.6	17.3	10.1	1.3	0.8	0.1
Poland	8.8	7.4	0.2	1.1	0.1	0.0
Portugal	24.1	12.3	5.8	4.6	0.7	0.7
Romania	19.1	12.0	4.8	1.7	0.5	0.1
Slovenia	16.5	9.7	5.7	0.0	0.5	0.7
Slovakia	9.6	6.9	2.3	0.0	0.3	0.1
Finland	30.7	26.0	3.9	0.8	0.0	0.0
Sweden	37.1	23.6	10.8	2.7	0.0	0.0
UK	8.1	5.7	0.2	1.7	0.5	0.0

[a] The category bioenergy includes wood and solid biofuels, liquid biofuels, biogas and renewable wastes
Source ec.europa.eu/eurostat.[11]

use). In contrast, solar energy, wind power as well as biomass have seen high growth rates. In the following sections, an overview of the different renewable energy sources focusing on electricity generation will be given.

[11] http://ec.europa.eu/eurostat/statistics-explained/images/d/dc/Renewable_energy_statistics-2018-v1.xlsx, accessed 13th May 2022.

4.2.1 Hydropower

The share of **hydropower** in power generation varies considerably in different European countries between 0% (e.g. Netherlands) and more than 90% (e.g. Norway) (see Table 4.4). These differences are mainly due to geographic conditions. The most extensive hydroresources in Europe and consequently the largest hydropower potential are in Scandinavia and the Alpine countries. No or only a slight increase in hydropower generation in the Alps can be expected as most of the economic potential in the Alpine region has already been developed.

In general, hydropower utilisation is characterised by

- very long technical and economical machine life (e.g. up to 80 years), especially when appropriate maintenance is carried out,
- the provision of electrical energy with terse lead times, making them interesting for peak load and reserve energy (see Sect. 10.3),
- the opportunity to provide black start[12] (see Sect. 10.4) as this type of power plant does not rely on the external electric power transmission network to recover from a total or partial shutdown,
- a very high efficiency of energy conversion (e.g. more than 90%), as well as
- quite a broad availability of hydropower worldwide (especially in nations with rapidly growing energy needs: China, India, South America).

Only water flows on the ground can be used for electricity generation, while the potential energy of all water raining down from the sky cannot be used yet (which would result in a much higher potential than only the share of water flows on the ground). Consequently, geographic conditions in combination with precipitations form water flows, which enable the exploitation of hydropower. Different geographic conditions and water flows result in various types of hydropower, which will be discussed in the following.

4.2.1.1 Types of Hydropower and Power Calculation

Hydropower plants convert water's potential and kinetic energy into electrical energy via mechanical energy, with potential energy making up the bulk of electricity generation. Hydropower plants can be differentiated according to various criteria, e.g. operation, head of water, type of turbine, capacity, annual output, etc. (see Table 4.5). In principle, three main types of hydropower can be distinguished:

Run-of-river power plants do not have a storage facility and process the inflows continuously. Run-of-river power plants, therefore, have only limited possibilities for changing the production volume. Storing water in the inflow is not or hardly achievable. If the incoming water volume exceeds the plant's capacity (e.g. in the case of a flood), the excess water remains unused and is bypassed in an overflow. Variations in production over a day are small, but output can vary

[12] A black start is the process of restoring an electric power station without relying on electricity from the power transmission network to recover from a total or partial shutdown.

Table 4.5 Characteristics of hydropower

<table>
<tr><th>Characteristic</th><th colspan="6">Attribute [a]</th></tr>
<tr><td>Operation</td><td colspan="2">Run-of-river power plant</td><td colspan="2">Reservoir power plant</td><td colspan="2">Pump storage power plant</td></tr>
<tr><td>Head of water</td><td colspan="2">Low pressure (1–20 m)</td><td colspan="2">Medium pressure (20–100 m)</td><td colspan="2">High pressure (>100 m)</td></tr>
<tr><td>Capacity</td><td colspan="2">Small power plants (<1 MW)</td><td colspan="2">Medium power plants (1–100 MW)</td><td colspan="2">Large power plants (>100 MW)</td></tr>
<tr><td>Type of head (depending on used turbine)</td><td colspan="3">Pressure head</td><td colspan="3">Velocity head</td></tr>
<tr><td>Type of turbine</td><td colspan="3">Reaction turbine</td><td colspan="3">Impulse turbine</td></tr>
<tr><td>Main turbine type</td><td colspan="2">Kaplan turbine</td><td colspan="2">Francis turbine</td><td colspan="2">Pelton wheel</td></tr>
</table>

[a] A specific hydroplant may be characterised by attributes from different columns

considerably between seasons. Hence, run-of-river power plants are mostly used to cover base load. A characteristic feature of run-of-river power plants is a large flow amount of water at relatively low heads, typically achieved in large rivers.

Reservoir power plants: the incoming water is stored as long as free storage volume is available. The stored water is used to provide electrical energy. In principle, differences can be made according to the storage volume in relation to the capacity. Depending on the duration required for emptying the reservoir at full load (using the full capacity of the plant), reservoir power plants can be classified in day storage plants (e.g. up to 6 h at full load), week storage power plants (between 6 and 25 h at full load), seasonal storage power plants (up to 500 h at full load), annual storage power plants (over 500 h at full load) or multi-year storage power plants. In addition to purely natural hydro inflows, the water can also be supplied to the reservoirs by so-called feeder pumps from side valleys. The basic rules for the utilisation of the stored water are subsequently discussed in Sect. 4.4.1.2.

Pump storage power plants (PSP) use only water to provide electrical energy which was previously pumped from a lower location into the upper storage basin.

Pumps and turbines are typically connected to the same lower and upper storage basin. The electrical energy required for pumping is usually consumed during off-peak hours (or rather at cheap electricity wholesale prices) and stored in the form of potential energy of water. In peak load periods (expensive electricity wholesale prices), the potential energy of the water is converted back into electrical energy. However, it should be noted that losses of approximately 10–35% (depending on the geographic situation, the efficiency of the plant, etc.) are associated with the conversion of electrical energy in pump storage plants from off-peak to peak load periods. The efficiency of storing electricity is called cycle- or round-trip efficiency and is accordingly between 65% and to a maximum of 90%.

Beyond this strict classification, combinations of run-of-river and reservoir power plants as well as reservoir and pump storage power plants may occur. If water in a river can be stored by (significantly) increasing the upper water level, this is a mixture between a run-of-river and a reservoir power plant. Furthermore, natural inflows can result in a combination of a pure pump storage and a reservoir power plant, sometimes also referred to as pump storage power plants with natural inflow. Hydropower plants, especially reservoir and pump storage plants, are generally characterised by very fast start-up times in the minute range and very high ramping rates compared to thermal power plants, which result in high flexibility, e.g. relevant for the provision of balancing energy as non-spinning reserve.

In areas with large river systems or lakes, the outflow of one plant may be the inflow of another. The lower basin of one hydroelectric power station becomes the upper basin of the next. Such linked formations are called hydropower cascades. Especially in the Alps, several hydropower cascades are made up by various reservoirs, pump storage and reservoir power plants. The interlinkages of hydro flows strongly impact power plant operation, necessitating the consideration of interlinkages in power plant dispatch planning.

Besides these main types of operation, hydropower plants can be distinguished according to the head of water from low heights and low pressure up to high heights and high pressure. Furthermore, the size of hydropower plants is strongly influenced by the geographic location, mainly by the amount of available water. Power plants with a capacity smaller than 1 MW are commonly labelled as small hydropower plants; sometimes, power plants with less than 100 kW are further distinguished as microhydropower, while capacities above 100 MW are classified as large power plants (see also Table 4.5).

The hydroturbine as core component of the hydropower plant converts water's potential and kinetic energy into mechanical energy, which is then converted with a generator in electrical energy. The power of a water turbine is derived from the physical equation of potential energy ($m \cdot g \cdot h$), by taking the time derivative. The electrical power output is calculated from the product of the water volume flow through the turbine ($\dot{V}$ in m^3/s) with the density of water (ρ in kg/m^3), the gravitational acceleration (g with 9.81 m/s^2), the head of water (h in m) and the total efficiency of the plant η_{total}, resulting in the following formula:

$$P_{el} = \dot{V} \cdot \rho \cdot g \cdot h \cdot \eta_{\text{total}}. \tag{4.1}$$

From this formula, it is obvious that a large head of water can compensate for low water flows and vice versa. A relatively small amount of water in a small mountain stream but with a high head of water of several hundred metres may generate more electricity than a large amount of water in a river that uses a shallow head of water in a weir. These different framework conditions necessitate different turbine types, which are explained in the next section.

4.2.1.2 Turbine Types

Several technical turbine concepts have been developed for different site conditions, enabling to exploit optimally the specific characteristics concerning water inflow and water height. Turbines can be differentiated according to various aspects: the application (part load or full load), the wheel shape (radial, diagonal, axial), the construction (vertical or horizontal shaft position) and the mode of action, which is probably most commonly considered distinguishing feature. Accordingly, there are **impulse turbines** and **reaction turbines**. Impulse turbines change the direction of flow of a high-velocity water flow. The corresponding impulse drives the turbine, and the water leaves the turbine with diminished **kinetic energy**. The pressure of the water does not change in the turbine blades; rather before and after the turbine, the same (standard) pressure applies. Consequently, no pressure casement around the rotor is required. The water jet's momentum is transferred to the turbine. In effect, "impulse" energy does work on the turbine. **Pelton turbines** and (historical) water wheels are impulse turbines. Pelton turbines are most efficient when the flow is low and the inlet pressure is high. These characteristics occur at small mountain streams with high water heads and are thus typical for power plants in mountain regions. Sometimes several Pelton wheels are mounted on one shaft. A giant Pelton wheel with 423 MW is in the power plant Bieudron within the reservoir plant complex Grande Dixence in Switzerland, with a head of 1883 m (world record), a total capacity of three turbines with each 423 MW and a maximum water flow of 75 m^3/s.[13]

In the Pelton turbine (see also Fig. 4.17), the water flow comes as a high-speed water jet from one or more nozzles tangentially to the wheel's blades. Each of the up to 40 blades is divided by a sharp edge, the so-called centre blade, into two approximately hemispherical half-blades, called buckets. In the middle of the cutting edge, the water jet from the nozzles hits tangentially. The buckets (or blades) redirect the water in the opposite direction so that the kinetic energy can be released to the impeller according to the principle of actio and reactio. The split into two water streams balances the side-load forces on the wheel. This enables an efficient transfer of the impulse from the water jet to the turbine wheel (cf. Giesecke et al. (2014)).

Besides Pelton turbines also cross-flow turbines, sometimes also called Ossberger turbines, are impulse water turbines as pressure remains constant at the

[13] http://www.grande-dixence.ch, accessed 13th May 2022.

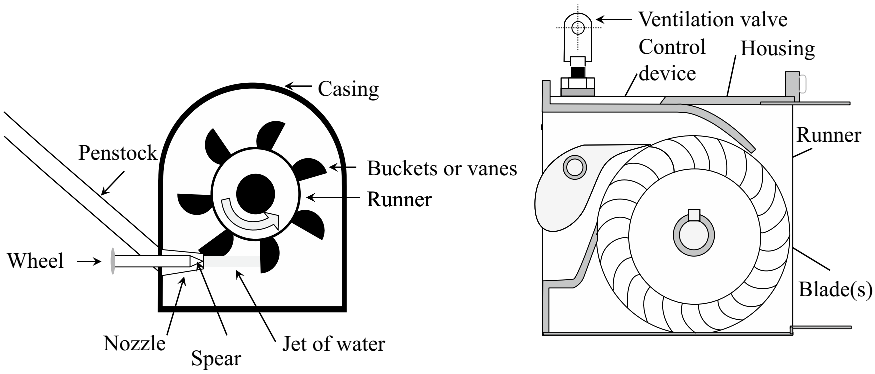


Fig. 4.17 Schematic illustration of a Pelton turbine (left) and an Ossberger turbine (right). *Source* Own illustration based on Giesecke et al. (2014, p. 532)

runner. Ossberger turbines (see also Fig. 4.17) consist of a cylindrical water wheel or runner with numerous blades. These are arranged radially and tangentially around a horizontal shaft. The water flows through the blade channels first from the outside to the inside and then back to the outside. Ossberger turbines typically have a nameplate capacity of less than 2000 kW. They are often used in mini- and microhydropower units with heads less than 200 m since they have a low price, an excellent behaviour under part load and are easy to operate, although the turbine's efficiency is somewhat lower than that of the other turbine types (cf. Giesecke et al. 2014, Chap. 14).

In contrast to impulse turbines, reaction turbines develop torque by reacting to the water pressure or mass. As the water passes through the turbine rotor blades, its pressure drops. Correspondingly, a pressure casement is needed to contain and direct the water. This serves also to maintain the suction imparted by the draft tube. **Francis turbines, Kaplan turbines** as well as Jonval turbines (as well as most steam turbines) make use of this concept.[14]

The Kaplan turbine (see Fig. 4.18 right side) is a propeller-type water turbine, combining features of radial and axial turbines. The waterflow is directed inwards and changes pressure as it moves through the turbine and gives up its energy. Thereby both the potential energy of the water head and the kinetic energy of the flowing water are converted in rotating energy. Kaplan turbine efficiencies are typically over 90% and they are broadly used worldwide for electricity generation, especially at sites with a low head of water combined with high water flows. Several variations of Kaplan turbines exist, such as, e.g. simpler propeller turbines, which have – in contrast to Kaplan turbines – non-adjustable propeller vanes.

[14] In contrast to water turbines, where only one turbine stage is needed as water is nearly incompressible, multiple turbine stages are usually used to harness the expanding gas efficiently for compressible working fluids, such as steam and gas.

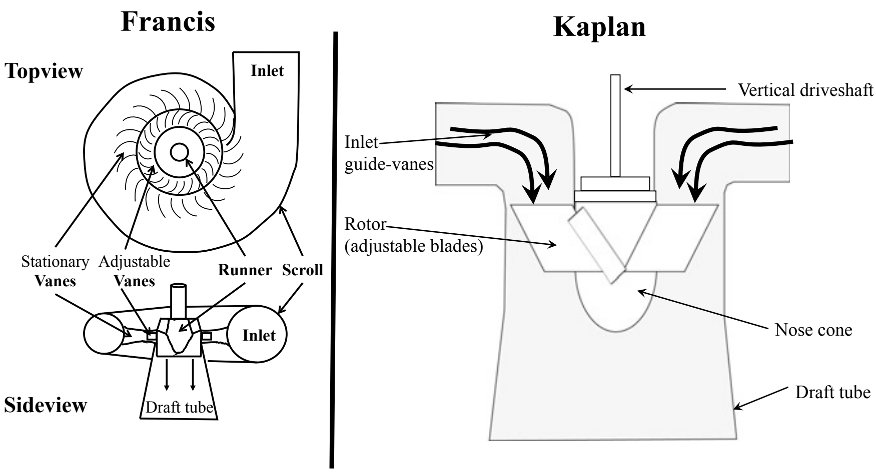


Fig. 4.18 Schematic illustration of a Francis turbine (left) and a Kaplan turbine (right). *Source* Own illustration based on Giesecke et al. (2014, p. 532)

Francis turbines (see Fig. 4.18, left side) are also reaction turbines and are the most common water turbines today. They operate at a water head from 40 to 600 m and have an electrical power output ranging from just a few kilowatts up to 800 MW. Francis turbines are also used for pumping water in pump storage plants. In the Francis turbine, the incoming water is directed by a volute housing, the so-called spiral, in an additional twist (thus sometimes also referred to as Francis spiral turbine). Then, the water is steered by a non-rotating, fixed stator blade ring with adjustable blades on the counter-curved blades of the impeller. The vanes at the inlet act as an actuator. With the setting of their angle, the speed and thus the turbine's power may be kept constant during load changes at changing water levels. The Francis turbine is a reaction turbine. At the impeller inlet, the pressure is higher than at the impeller outlet. Modern Francis turbines achieve efficiencies of over 90% (cf. Giesecke et al. 2014, Chap. 14).

As depicted in Fig. 4.19, the different turbine types are particularly suited for different water conditions. While Pelton turbines are designed for high water heads in combination with low flows, Kaplan turbines are used for low water heads combined with high flows. Francis turbines make use of medium water heads and flows.

As each hydropower plant is designed for individual site conditions (especially for larger plants), electricity generation costs strongly depend on the specific situation. The main cost drivers are the available water (flow), the head of water and the investment for installing a weir or reservoir. Thus, electricity costs vary broadly between 15 and 80 €/MWh (cf. Möst 2006). As the list of the top ten largest hydropower plants shows (see Table 4.6), the capacity of single power plants and their electricity production can be gigantic. The largest hydropower plant worldwide is the Three Gorges Dam in China at the river Yangtze with a capacity of 22.5 GW and an electricity production of almost 100 TWh. The 10th largest power plant Krasnoyarsk, located on the river Yenisei in Russia, still has a capacity of 6 GW and produces 15 TWh of electricity per year.

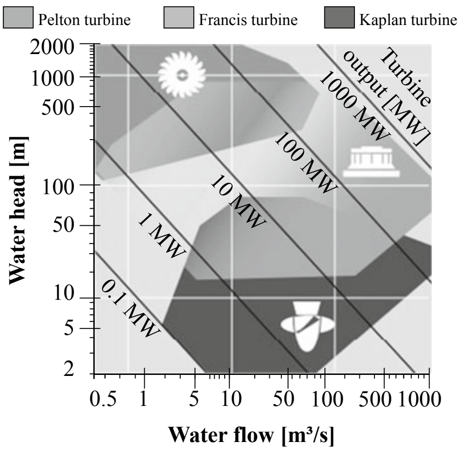


Fig. 4.19 Typical utilisation of turbine types dependent on head of water and hydro inflow. *Source* Own illustration based on Giesecke et al. (2014, p. 534)

Table 4.6 Top ten list of largest hydropower plants worldwide

Name	River and country	Years of completion (first and latest units)	Installed capacity (MW)	Yearly production (TWh)
Three Gorges Dam	Yangtze, China	2008/2012	22,500	98.8
Itaipu Dam	Paraná, Brazil/Paraguay	1984/1991, 2003	14,000	103.1
Xiluodu	Jinsha, China	2014	13,860	55.2
Guri	Caroní, Venezuela	1978, 1986	10,235	53.4
Tucuruí	Tocantins, Brazil	1984, 2007	8370	41.4
Grand Coulee	Columbia, USA	1942, 1950–1991; 1983 and 1984	6809	20.0
Xiangjiaba	Jinsha, China	2014	6448	30.0
Longtan Dam	Hongshui, China	2007/2009	6426	18.7
Sayano–Shushenskaya	Yenisei, Russia	1985/1989, 2010/2014[15]	6400	26.5
Krasnoyarsk	Yenisei, Russia	1967/1972	6000	15.0

Source https://www.power-technology.com/

4.2.1.3 Multipurpose Use of Hydropower

Many hydropower plants serve different infrastructural purposes and do not only generate electrical energy. If hydropower projects also provide other services, these are called multipurpose plants. Depending on the original requirements, the use of hydropower may be the primary objective (or secondary), while the other services are additional (or primary). Likewise, the available potential of the water can be used as a side benefit of another necessary measure. Multiple services represent an external benefit of hydropower and are often not directly compensated. However, not all multipurpose effects have to be external benefits; instead, they may be harmful. An example is the potential production of large quantities of methane from reservoirs due to decomposing organic matter under the water.

Positive multipurpose services of hydropower use are:

- flood control through the creation of artificial dams and thus the delayed release of strong water inflows after heavy or long-lasting precipitation events;
- irrigation by meeting high demands for agriculture when rainfall is low – reservoirs store runoff during times of high rainfall and low demand;
- provision of drinking and service water through dams;
- ensuring navigation through the regulation of running waters using cascaded storage reservoirs;
- regulation of groundwater levels by the attenuation of water-level fluctuations;
- delivery of minimum water to the lower reaches by the water retention during higher inflows, which is relevant to maintain the ecosystem functions – to mitigate negative impacts, adequate flows downstream of a barrier have to be maintained;
- the promotion of biodiversity by artificially damaging rivers and creating new water areas and riparian zones that allow the development of new biodiversity;
- creation of recreational space, suitable for boating, swimming and fishing, which are only secondary uses.

The fulfilment of these multipurpose services and the use of hydropower always raise economic, ecological and social issues in every hydropower project, which should be considered in a comprehensive assessment for each project.

4.2.2 Wind Power

Wind energy is a product of the sun, as the solar radiation leads to different temperatures of the (ground and the) air in different regions. This in turn induces areas of different air pressures. The differences in air pressure lead to wind formation; typically, the wind transports air to places with lower pressure (e.g. from the sea to the mainland during daytime). Near the earth surface, the wind speed is lower than at higher altitudes. This results from the fact that the wind is slowed down by surface friction. In this context, surface roughness plays an important role.

The wind speed v_i at an altitude h_i can be estimated, e.g. by the **Hellmann** approach (cf. Kaltschmitt et al. 2007, p. 55) from a given wind speed v_j at an altitude h_j and using approximate values for the exponent k:

$$v_i = v_j \cdot \left(\frac{h_i}{h_j}\right)^k. \tag{4.2}$$

Furthermore, wind speed is not only strongly varying with the altitude above ground but also over time. Distribution functions of measured wind speeds at a specific site and their mathematical approximations, e.g. in the form of Weibull distribution functions, are typically used to describe the frequency of different wind speeds – from hardly any wind (calm) to heavy storms with velocities of more than 25 m/s.

Wind power plants exploiting the aerodynamic principle transform parts of the **wind's kinetic energy** into mechanical energy and finally into electricity (cf. Kaltschmitt et al. 2007, pp. 295–348). In general, the kinetic energy of any body having the mass m and the velocity v can be calculated with the following formula:

$$E_{\mathrm{kin}} = \frac{1}{2} \cdot m \cdot v^2. \tag{4.3}$$

Considering the mass with a given mass density ρ, that flows during a time period through an area A, (the so-called mass flow $\dot{m}$)

$$\dot{m} = A \cdot v \cdot \rho, \tag{4.4}$$

the power of the wind can be calculated using the following formula (for area A, see Fig. 4.20):

$$P = \frac{1}{2} \cdot A \cdot \rho \cdot v^3. \tag{4.5}$$

The wind speed is of highest importance for the power to be extracted from the wind: the available power is a cubic function of the wind velocity.

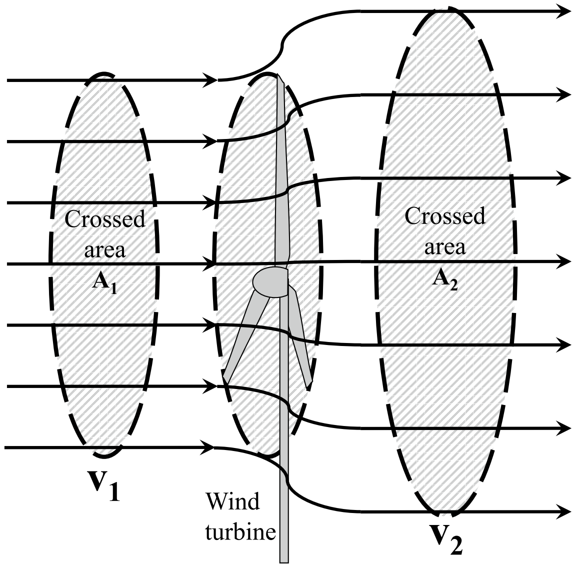


Fig. 4.20 Schematic illustration of wind speed and fluid flow before and after a wind turbine. *Source* Own illustration based on Kaltschmitt et al. (2007, p. 297)

Of course, not the wind power plants' whole kinetic energy can be used, as this would imply that the wind has a velocity of zero behind the wind power plant, leading to congestion hindering the further use of wind energy. The power to be utilised by the wind power plant is the difference between the power in front of the wind power plant (index$_1$) and behind it (lower velocity and larger area A_2, index $_2$) (see Fig. 4.20 and Kaltschmitt et al. 2007, p. 297):

$$P_{\text{WPP}} = \frac{1}{2} \cdot A_1 \cdot \rho \cdot {v_1}^3 - \frac{1}{2} \cdot A_2 \cdot \rho \cdot {v_2}^3 = \frac{1}{2} \cdot \dot{m} \cdot \left(v_1^2 - v_2^2\right), \tag{4.6}$$

which can be reformulated using the average of the wind velocities before and behind the wind power plant as the velocity v of the wind in the wind power plant and the area of the rotor surface (A_{WPP}):

$$P_{\text{WPP}} = \frac{1}{4} \cdot A_{\text{WPP}} \cdot \rho \cdot {v_1}^3 \cdot \left(1 + \frac{v_2}{v_1}\right) \cdot \left(1 - \left(\frac{v_2}{v_1}\right)^2\right). \tag{4.7}$$

Dividing the power to be utilised by the wind turbine by the total power of the wind leads to the so-called power coefficient (maximal efficiency comparable to the Carnot efficiency in the case of thermal power plants, which takes its maximum at 16/27 = 0,593 (see Fig. 4.21), when a ratio of wind velocities v_2/v_1 of one-third is realised. This shows that at best about 60% of the wind power can be used, independently of the type of the wind power plant (so-called **Betz's law**).

Wind power plants use two different principles to harness wind energy: the lift principle or the drag principle or both (cf. Kaltschmitt et al. 2007, pp. 301–308). Different pressure levels below the blade (higher pressure on the underside of the blade) and above the blade are needed to use the **lift force**, which is similar to the use of the lift force in the case of aeroplanes. This is realised by the form of the blade

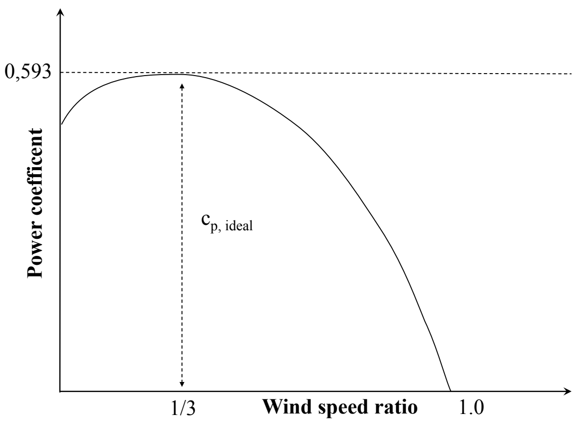


Fig. 4.21 Power coefficient curve for wind turbines exploiting the aerodynamic effect. *Source* Own illustration based on Kaltschmitt et al. (2007, p. 300)

(shape and concavity of the airfoil) and the angle of attack of the wind, which leads to longer distances to be covered by the air and a smaller current cross section at the upper side. The smaller current cross section results in a higher velocity at the top side and thereby lower pressure (Bernoulli's principle). The **drag force** is created by striking the air on a surface moved by this force. As the power coefficient achievable by the lift principle is considerably higher than the one realisable by the drag principle, modern wind power plants are designed to use the lift force primarily.

In general, wind power plants can be differentiated according to their axis into vertical axis wind turbines (VAWT, e.g. Savonius and Darrieus wind turbines) and horizontal axis wind turbines (HAWT) with one or more blades. HAWT have some advantages in comparison with VAWT, such as higher towers making use of higher wind speeds and by using cambered airfoils, which are more efficient due to a higher lift-to-drag ratio. Consequently, HAWT have a better cost-effectiveness and have thus obtained a much greater market penetration in today's wind power market. Wind power plants with a horizontal axis are provided with one to three blades, in most cases with three blades.

Wind power plants are typically equipped with a control mechanism to avoid overloads at high wind velocities and operate the turbine under optimal conditions. Different control mechanisms exist. A relatively simple control mechanism is the so-called stall control. Here the blades are fixed to the rotor hub without a possibility to twist the blades (therefore, this concept is also called passive pitch control). The power control is realised by the profile of the blade, which is constructed in such a way that in the case of strong winds the angle of attack, which is the angle between the chord line of the blade and the wind vector (see Fig. 4.22), is drastically increasing. This results in the so-called stall, meaning that the airflow separates from the contour of the blade. The control concept has been further developed to the so-called active stall control: here, the blades can be twisted slightly to initiate the stall. If the blades can be twisted more extensively, the mechanism is called pitch control. Pitch control enables an optimal angle of attack by automatically rotating each of the blades to maximise efficiency. In extremely high wind speed situations, the pitch control causes the blades to be totally turned out of the wind.

Wind power plants with a horizontal axis typically consist of the following components (cf. Kaltschmitt et al. (2007, pp. 308–322) and Fig. 4.23):

- Foundation: the design of the foundation of wind power plants depends mainly on the particular soil conditions. It is more expensive in the case of offshore wind power plants due to the more challenging framework conditions (e.g. water depth) compared to onshore sites, where typically a gravity foundation based on concrete and steel can be used. However, this may change if swimming foundations are used for offshore sites in the future.
- Tower: monopole towers of modern wind power plants are made of steel or concrete and reach heights of up to 200 m to be able to use higher wind speeds.
- Rotor: the rotor consists of a rotor hub and up to three rotor blades, typically made of fibreglass. Wind power plants can be equipped with rotors having up to 200 m rotor diameter.

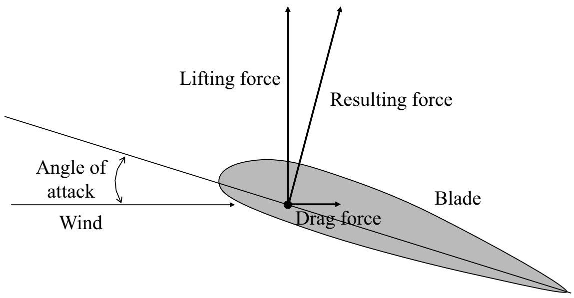


Fig. 4.22 Lift and drag forces at the blade. *Source* Own illustration based on Anderson (2011, p. 20)

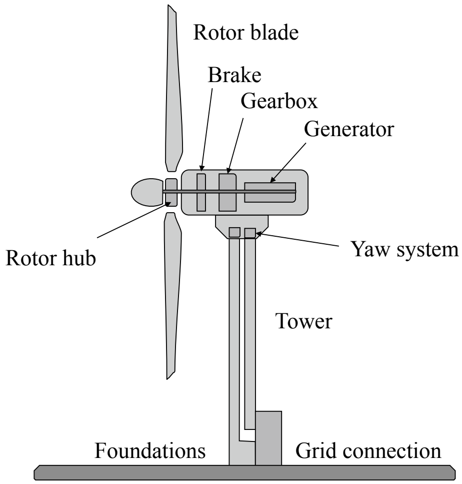


Fig. 4.23 Schematic representation of a modern wind power plant. *Source* Own illustration based on Kaltschmitt et al. (2007, p. 310) and Hau (2016, p. 73)

- Nacelle and its equipment: the nacelle of wind power plants might be equipped with a brake, a gearbox (not necessarily in the case of gearless turbines) to increase the rather low speed of the rotor revolution to higher speeds, and a generator to convert this mechanical energy into electricity, which is then transported through the tower to the grid.
- Yaw mechanism: the yaw system is needed to turn the whole nacelle into the wind and readjust its orientation according to the wind conditions.

Offshore wind turbines typically consist of larger towers and rotors than onshore wind turbines, which results in a greater electric capacity. In both cases the transportation, the foundation and the installation of the different components can be very demanding.

Wind power plants start converting the wind's kinetic energy into electricity as soon as the wind speed exceeds a minimal threshold (so-called cut-in speed, see

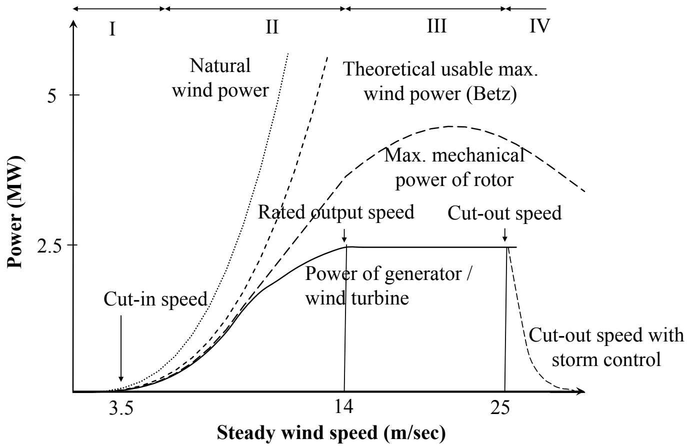


Fig. 4.24 Typical power curve of a wind power plant. *Source* Own illustration based on Wood et al. (2013, p. 23) and Kaltschmitt et al. (2007, p. 326)

phase I in Fig. 4.24), because a certain friction has to be overcome. Then the output of electrical power increases with the cube of the wind velocity until the rated output speed is reached; there, the wind power plant has reached its maximal capacity. To operate the wind turbine with increasing wind velocities at a constant output level requires an adequate control strategy (see above: stall or pitch control). If the wind speed outstrips the so-called cut-out speed, which is typically at about 25 m/s, the wind power plant is turned out of the wind and – with the help of the braking system – the blades are brought to a halt to avoid damages at the wind turbine (see phase IV in Fig. 4.24). As a sharp drop of power production at high wind speeds may be challenging for electricity grids, modern wind power plants have a so-called storm control, which slowly decreases the power with increasing wind speeds to avoid a sharp drop of power.

A wind power plant is rather seldom installed as a stand-alone unit. Typically many wind power plants are installed nearby, creating a so-called wind park. To reduce shadowing effects concerning the wind speeds available for the different wind power plants, certain design principles, e.g. the distance between the wind turbines of the park, have to be considered. Finally, it should be mentioned that also this way of using renewable energies leads to some environmental effects. In fact, there are no direct emissions like those of pollutants or greenhouse gases, but there are, e.g. noise emissions (audible and infrasonic sound), shadow impacts caused by the rotor blades and visual impacts (changes in the natural scenery) as well as the reduction of wind speed (which may result in local impact and which could be a non-negligible factor in the case of large-scale energy extraction).

4.2.3 Solar Energy

After a short introduction to the basics and characteristics of solar energy, subsequently, technologies will be in focus, which generate electricity from solar energy.

Solar energy is the energy of solar radiation from the sun. The sun has a mass of about 1.99×10^{30} kg (about 330,000 times the mass of the earth) and probably consists of about 91% hydrogen, 8.9% helium and 0.1% other elements. Temperatures of about 8 to 15 million Kelvin [K], as well as high pressure of more than 200 billion bar, prevail in the core of the sun. Under these conditions, nuclear fusion of hydrogen into helium takes place (see also Sect. 4.1.2). It amounts to about 4.3 million tonnes per second and thus corresponds to an average output of about 3.8×10^{26} J. This results in an average intensity of solar radiation at the boundary of the earth's atmosphere of about 1367 kW/m^2, which is also called solar constant. The solar constant is not a proper physical constant; it is just an average of a time-varying value.

Global radiation is the direct and diffuse radiation incident on a horizontal surface. Irradiance describes the power measured in W/m^2 and irradiation the energy measured in Wh/m^2. The global radiation at different locations and times depends on the angle of incidence (due to the earth's spherical shape), altitude and weather conditions. The mean annual value of the global irradiation is approximately 2200 kWh/m^2 at the equator and approximately 800 kWh/m^2 at the poles. The annual potential of solar energy, which is often derived as the energy from the sun at the total surface of the earth, is estimated to be between 1575 and 49,837 exajoules (EJ), which is several times larger than the total world energy consumption, being in the magnitude of approximately 570 exajoules (in 2015). Although this physical potential can only be used to a small share given technical and economic restrictions, it clearly shows that this energy source is abundant.

Solar energy, respectively, solar radiation, can be used technically in the form of electricity, heat or chemical energy. A range of technologies harnesses solar radiation, such as solar heating, photovoltaics, solar thermal energy, artificial photosynthesis, solar architecture (for light and heat usage), molten salt power plants, etc. Either passive or active solar usage can be distinguished:

- Active solar techniques transform the solar energy into other energy carriers by technical components (e.g. electricity or heat flows) and include photovoltaic systems, concentrated solar power and solar heating.
- Passive solar techniques include orienting a building to the sun and designing spaces and materials, making use of solar energy in the form of lighting and heating.
- Active solar techniques are required to transform solar energy into electricity, such as photovoltaic systems and concentrated solar thermal power plants, explained in the next sections.

4.2.3.1 Solar Thermal Power Plants

Solar thermal power plants make use of solar energy by converting it to heat. Thereby, concentrating collectors bundle the direct solar radiation and are thus called concentrated solar power plants (CSP). These power plants generate heat – solar thermal energy – by using mirrors or lenses to focus sunlight arriving on a large surface area onto a much smaller surface called the (energy) receiver. This concentration of sunlight enables a higher energy density at the receiver. Electricity is

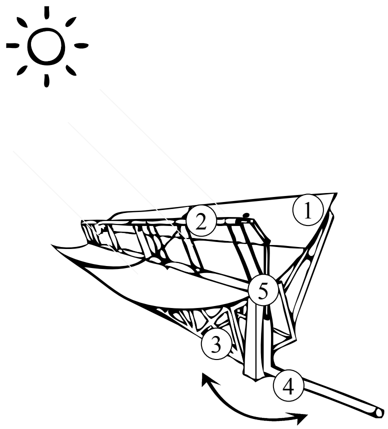


Fig. 4.25 Parabolic trough collector with sun tracking (1 reflector/mirror, 2 absorber tube, 3 frame, 4 solar field piping, 5 sun-tracking motor). *Source* Own illustration based on Kaltschmitt et al. (2013, p. 196)

generated using this concentrated solar thermal energy, respectively, heat, usually by a thermal engine (e.g. steam turbine as in a conventional power plant). Solar thermal power plants can either derive their primary energy exclusively from solar energy (solar only) and are then often equipped with solar thermal storage. Or they are equipped with fossil auxiliary firing (hybrid) to bridge (short) interruptions of solar radiation, but also a combination with conventional fossil power plants is possible. The advantage of systems with solar thermal storage is that besides a longer period of operation and thus a higher capacity factor, their use is dispatchable (at least in a short time horizon). Therefore, electricity generation is more independent from solar radiation, which is an advantage compared to weather-dependent technologies.

Different types of concentrating collectors exist for concentrating the solar radiation, such as parabolic trough collectors, Fresnel collectors or solar towers. Parabolic trough collectors and Fresnel collectors focus the solar radiation on a line, while solar radiation is concentrated on a point in solar towers. Parabolic trough collectors concentrate the solar radiation on an absorber tube (which serves as an energy receiver). Sun-tracking systems may be necessary and allow to adjust the mirror to the angle of incidence of the sun, as depicted in Fig. 4.25. Instead of using a curved reflector, Fresnel technology uses flat glass mirrors (instead of parabolic ones), making production and operation more straightforward and cheaper. In solar tower power plants, the solar light is concentrated at one point in the tower, resulting in extremely high temperatures at the receiver, which is a challenge for the material of the heat converter but results in higher temperatures for the steam process.

At the best sites (locations with very high solar radiation, such as in North Africa), the thermal efficiency of the solar field is on average about 50% and the total solar-electric efficiency of parabolic through power plants is up to 18%.[15] Concentrating solar power had a total installed global capacity of approximately 5 GW_{el} in 2016.

4.2.3.2 Photovoltaics

As opposed to solar thermal power plants, photovoltaic (PV) cells or systems convert light into electricity using semiconducting materials with the help of the photoelectric effect. The photoelectric effect describes a physical phenomenon of the interaction of photons with matter: an electron is dissolved from a bond – e.g. in an atom or the valence band – by absorbing a photon or, in other words, when light shines on a material. Electrons are only dislodged when photons reach or exceed a threshold energy (proportional to the frequency of the electromagnetic radiation). Below this material-dependent threshold, no electrons are ejected from a suitable material regardless of the light intensity or the length of exposure to the light. To explain that light can eject electrons even if its intensity is low, the concept of photons as a collection of discrete wave packets is necessary, which goes back to a proposal by Einstein. The electrons are lifted using the energy of the photons from the valence band into the energetically higher conduction band. The energy of each photon must correspond to at least the (photonic) bandgap of the irradiated semiconductor. The size of the bandgap depends on the used material. This explains the maximum wavelength for each material up to which the photoconductivity occurs (e.g. gallium arsenide: 0.85 µm, germanium: 1.8 µm, silicon: 1.1 µm). With the help of light, more precisely, the photons, voltage and electric current are created in the related materials (see Fig. 4.26).

For solar cells, the photovoltaic effect is used in combination with two differently doped materials, the so-called p- and n-layer. In the p-layer, trivalent-doped materials (doped with elements with three valence electrons) can take up an additional external electron and thus leave a hole in the valence band of the silicon atoms. This makes the electrons mobile in the valence band. In consequence, the "p" (positive) side contains an excess of holes. When photons hit the solar cell, they generate

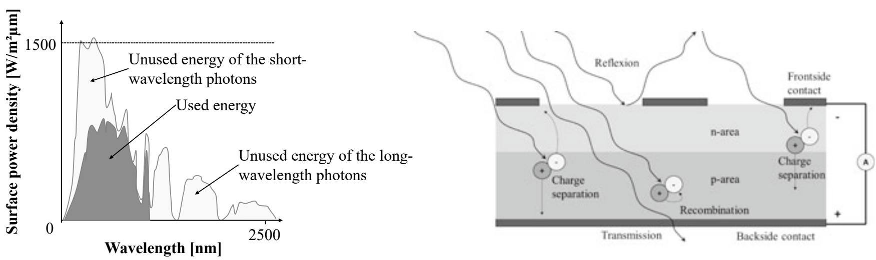


Fig. 4.26 Used energy by PV cells (left) and processes in PV cells (right). *Source* Own illustration based on Unger et al. (2020, p. 45) (left) and Quaschning (2004) (right)

[15] Efficiency is determined by the ratio of produced electricity and irradiated solar energy on the surface of the plant.

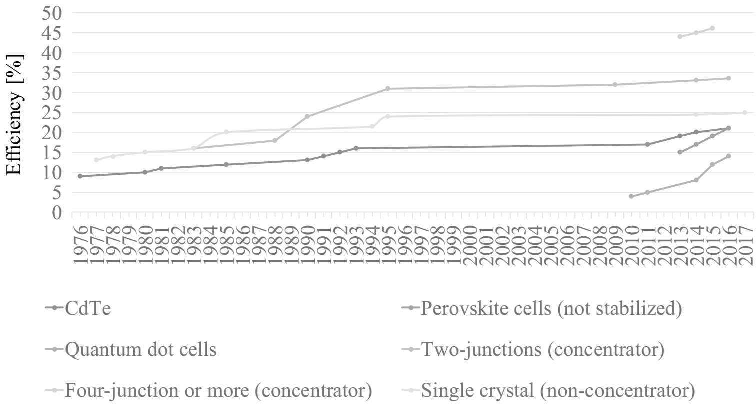


Fig. 4.27 Laboratory efficiencies of different types of solar cells. *Source* Own illustration based on NREL (2021)

so-called electron–hole pairs in the p-layer (positive). Electrons pass through the boundary layer in the n-region (negative), where they can move without problems. In contrast, the "n" (negative) side contains an excess of electrons in the outer shells as in the n-layer, 5-valent dopant (doped with elements with 5 valence electrons) has an outer electron. Due to the differently doped materials in the two layers, electrical current can only pass through the junction in one direction. The doping of the two layers is achieved by ion implantation or diffusion of dopants (e.g. n-doping with phosphorus and p-doping with boron). By separating the individual charges, a plus and a minus pole arise, and by the connection of the two poles with the interposition of a consumer, an electric current can flow (see Fig. 4.26).

In consequence, these solar cells can generate electrical power when light hits the cell. Several solar cells are combined into a solar panel. These solar panels may be ground-mounted, rooftop-mounted or wall-mounted. The solar panel can be mounted as a fixed installation or use a solar tracker to follow the sun across the sky (providing a higher solar yield but with a more expensive construction). The following advantages accompany solar PV: once installed, its operation generates no pollution and no greenhouse gas emissions. Solar PV is easy to scale regarding the provision of electrical energy but is strongly dependent on solar radiation.

Different types and materials of solar cells are used and are still in research. Solar cell (lab) efficiencies[16] vary from 6% for amorphous silicon-based solar cells to 44.0% with multiple-junction concentrated photovoltaics (see Fig. 4.27). Solar cell energy conversion efficiencies for commercially available photovoltaics are currently around 14–22%.

In 2020, approximately 140 GW of photovoltaic systems were installed, reaching a cumulative capacity of almost 800 GW worldwide (cf. IEA 2021). The growth rate of PV installations has been steadily above 30% in the last years: between 1998 and 2015, photovoltaic power grew on average by 38%. It is

[16] Solar cell efficiency is defined as quotient of solar power and solar irradiance.

expected (by several institutions, among others the International Energy Agency) that the installed capacity will increase to between 3000 and 10,000 GW by 2030. While different types of solar cells exist (see Fig. 4.27), the global market share of crystalline silicon cells was about 90% (in 2014). Silicon cells are expected to remain the dominant photovoltaic technology.

Photovoltaics has long been the most expensive form of electricity generation through renewable energy. However, substantial cost reductions of the system components have been realised, making solar power attractive for electricity generation (see Sect. 10.7.4). From 2011 to 2017, the cost of photovoltaic power generation decreased by almost 75%. With each doubling of the total installed capacity, the costs of photovoltaic modules have been decreasing by approximately 20%, often referred to as a learning rate (cf. Junginger et al. 2019). At locations with high solar irradiance, power generation costs are currently in the magnitude of 2–3 €ct/kWh or even less. In many locations, PV has reached grid parity, which is usually defined as PV electricity generation costs at or below retail electricity prices (cf. Sect. 10.7.4 for a detailed discussion). PV electricity generation costs are in the range of thermal power generation costs for coal or gas-fired plants or even significantly below (see the discussion of levelized cost of electricity, LCOE in Sect. 4.3.2). However, the weather-dependent feed-in requires complementary electricity storage. Batteries are becoming more and more attractive, also from an economic perspective, as they can store surplus energy of the day for demand at night. This energy management function strongly depends on the local regulations and the resulting incentive structure (cf. Sects. 6.1.4 and 10.7): if surplus energy at noon (high solar irradiance) cannot be injected into the grid (or only at low prices) and electricity purchase at high costs can be substituted in the night, batteries may be an economically attractive option for the PV owner. However, the gap between high summer solar irradiation (and a summer surplus) and a high winter demand necessitates shifting energy from summer to winter. This shifting of energy is still and will remain very costly, as infrastructure has to be installed for this one cycle per year and would necessitate enormous battery and PV capacities.

4.2.4 Bioenergy

Many different organic materials, including woody and agricultural residues, energy crops (like maize, miscanthus or jatropha) and various forms of waste, are subsumed under **biomass**, respectively **bioenergy** if these materials are used for energy provision. Biomass can be assigned to renewable sources of energy, if biomass is realised sustainably. The manifold forms of biomass can be classified in different ways, e.g. according to their aggregate state (e.g. solid, liquid and gaseous biomass), according to their origin (e.g. biomass from forests, agriculture, fishery and waste) or according to the type of feedstock (e.g. non-lignocellulosic biomass, lignocellulosic biomass). Fuels produced out of biomass, so-called biofuels, are typically differentiated according to their starting substances, e.g. into first-generation biofuels (produced from food biomass, e.g. sunflower oil), second-generation biofuels (produced from non-food biomass, e.g. straw) and third-generation biofuels (produced from algae).

Biomass-based energy carriers have their origin in the energy from the sun, as plants transform solar energy into chemical energy in the form of glucose via photosynthesis (cf. Kaltschmitt et al. 2007, pp. 81–88). Besides the sunlight, only water and carbon dioxide (from the air) are needed for this reaction, which is realised in the cells of plants – in the so-called chloroplasts containing chlorophyll – according to the following simplified overall reaction equation:

$$6\ H_2O + 6\ CO_2 \rightarrow C_6H_{12}O_6 + 6\ O_2.$$

The use of biomass for energy production is not only characterised by manifold forms of biomass, that can be employed, but also by different conversion pathways, that can be pursued (see Fig. 4.28 and Kaltschmitt et al. 2007, pp. 511–516). In contrast to the use of wind and PV for electricity production, the use of biomass has the advantage that energy production can be adjusted to the current energy demand. This holds even true for the biochemical conversion via anaerobic digestion used to produce so-called biogas. Indeed, the reactions of the biochemical conversion processes require a continuous production process, but there is the possibility to fulfil this requirement, without producing electricity, by storing the produced biogas. Furthermore, different products can be obtained from biomass that can meet various demands by displacing existing supply to different extents.

However, the energetic use of biomass is also connected with several challenges. Different biomass pathways potentially compete for the same resources with each other and other sectors like food and materials, which is crucial. Especially the use of first-generation biofuels produced from food biomass has led to discussions

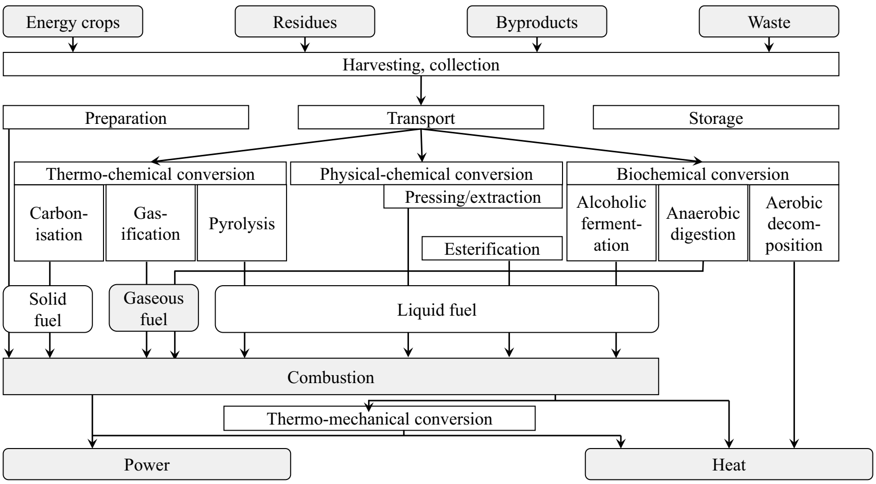


Fig. 4.28 Biomass conversion pathways. *Source* Own illustration based on Kaltschmitt et al. (2007, p. 512)

about whether the cultivation of biomass to produce energy should be realised, as long as we have problems feeding the world. In addition, depending on the form of the biomass used, the logistic steps necessary to provide the needed input can be rather complex and accordingly costly. Another disadvantage is that this form of renewable energy production leads to direct emissions, e.g. emissions of greenhouse gases (e.g. CO_2) and pollutants (e.g. heavy metals). It is argued that the absorption and the subsequent release result in a closed-loop of CO_2 as the used plants have absorbed the corresponding CO_2, which is needed for the photosynthesis relatively shortly before. Therefore, biomass can be seen as a CO_2-free energy carrier (so-called carbon neutrality). There is also a lot of discussion about creating negative emissions with the help of biomass combined with carbon capture and storage (CCS, see Sect. 6.2.2.3), especially in long-term scenarios of the Intergovernmental Panel on Climate Change (IPCC).

A widely used and relatively easy way to energetically use biomass, like wood (products) or straw, is to burn it to produce heat. The heat might then be used for heating purposes, e.g. via a district heating system, cooking, e.g. in developing countries, or producing electricity via the well-known steam cycle (see Sect. 4.1.1.2). This transformation pathway to produce electricity is based on the same components as other steam power plants. It is possible to switch partly – so-called **co-firing** – or totally from burning coal to burning solid biomass in existing steam power plants, e.g. hard coal power plants, when they are designed for switching these fuels. This strategy has the advantage that the power plant operator can – at least partly – use components of the existing power plant.

In addition to the combustion of biomass, which is provided in solid form by nature, the following pathways to provide solid, gaseous or liquid biomass can be differentiated:

- Thermo-chemical conversion (carbonisation, gasification, pyrolysis),
- biochemical conversion (alcoholic fermentation, anaerobic digestion, aerobic decomposition) and
- physical–chemical conversion.

Concerning thermo-chemical conversion, the gasification of the raw materials to produce a synthetic gas (syngas) is essential, not only because this gas can be used as the primary material for producing biofuels. Furthermore, anaerobic digestion, which belongs to the biochemical conversion possibilities, is found in many energy systems worldwide.

The example of biogas for electricity production shall be used for illustrating the biomass conversion pathway. The following four steps characterise this conversion pathway: (1) feedstock management, which comprises, e.g. harvesting, transport and pre-treatment of the biomass, (2) biogas production in biogas plants, (3) biogas treatment and use and (4) digestate treatment and use (cf. Balussou 2018, pp. 15–25).

The central element of the conversion via anaerobic digestion is the biogas plant (see Fig. 4.29), where the production of the biogas is realised in heated, oxygen-free tanks, the so-called fermenters. There are different ways of charging

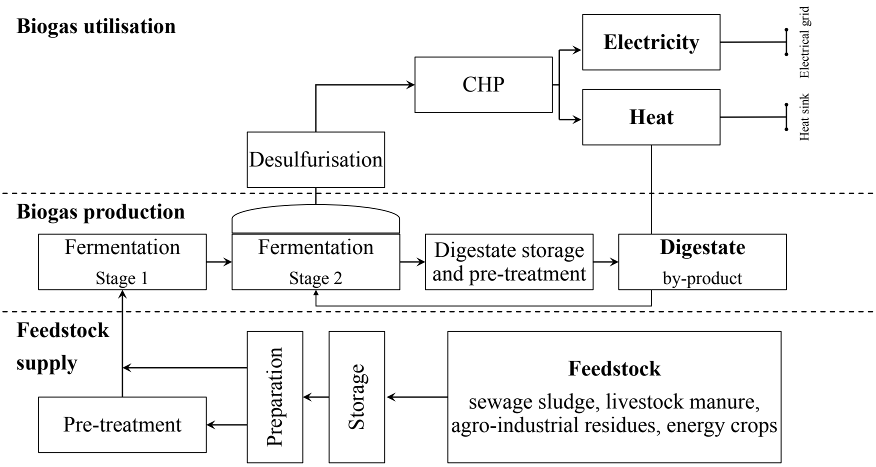


Fig. 4.29 Components of a biogas plant. *Source* Own illustration based on Bidart (2013, p. 81)

and discharging the fermenter (e.g. continuous mode or batch mode). The fermenters are equipped with stirring devices. As soon as the input materials, like energy crops and manure (wet fermentation), have been brought into the fermenter, bacteria begin to decompose the organic biomass. Two different products are obtained: biogas, which is a mixture of different gases, mainly methane and carbon dioxide, and a residue called digestate, which can later be used as a fertiliser. The produced biogas is often burned in small cogeneration units typically consisting of a motor engine installed at the site, e.g. the biogas-producing farm. The biogas is in general desulphurised and dried to reduce pollutant emissions and increase efficiency. The usage of produced heat increases the total efficiency of the small cogeneration unit, e.g. to serve local heat demand for houses and mews. In the absence of heat demand, biogas injection to a natural gas grid may be favourable. If the produced gas is injected into the natural gas grid, an even more sophisticated treatment process is necessary to comply with the quality requirements of the grid operators; one crucial part of this treatment is the removal of CO_2 from the biogas.

4.2.5 Other Renewable Energy Technologies

There are several other renewable energy technologies, such as geothermal energy, wave energy, tidal energy, seaflow energy, osmosis energy, ocean thermal energy conversion, etc. As geothermal energy and wave energy are also promising for electricity production, a short introduction will be given to these two technologies, while the other ones are not further detailed here.

4.2.5.1 Geothermal Energy

Geothermal energy corresponds to the energy stored in the form of heat below the earth's surface. Like solar and wind energy, hydropower and biomass, it is one of the regenerative forms of energy and has the unique advantage that it is continuously available. In general near-surface geothermal energy is differentiated from deep geothermal energy. The differentiation of both types is a consequence of legal aspects (e.g. mining law), but also refers to the used technologies. The focus is subsequently on technologies for power generation, which need higher temperature levels at the geothermal site. These higher temperature levels are available at either attractive geothermal areas[17] or by deep drilling (deep geothermal energy).

Geothermal energy in the earth's crust originates to approximately one-third from the original formation of the planet and to approximately two-thirds from radioactive decay of materials within the earth. It is estimated that 99% of the earth is warmer than 1000 °C. Temperatures at the core–mantle boundary may reach over 4000 °C. This results in a heat flux density of about 0.07 W/m^2 on average at the earth's surface (cf. Kaltschmitt et al. 2007, p. 94). This corresponds approximately to a total thermal power of 33 TW. Although geothermal energy is not infinitely available, it is considered renewable as the time dimension of earth cooling is "forever" compared to human life. As long as the geothermal resource is not locally overexploited, i.e. the inflowing heat flow from the earth is sufficient to serve the heat extraction, heat can be permanently extracted.

Geothermal energy in its simplest form is water from hot springs and has been used since Paleolithic times. In Roman times, geothermal energy was already used for space heating. Since the second half of the twentieth century, geothermal energy has also been used for electricity generation. Worldwide, a capacity of approximately 12 GW of geothermal power for electricity generation is currently installed (cf. Matek 2013). Historically, thermal energy usage has been limited to areas near tectonic plate boundaries as these result in higher temperature close to the surface (including hot springs, etc.). Also, geothermal electric plants have been first built in these regions, since the high-temperature geothermal resources enable higher conversion efficiencies (see Carnot efficiency, Sect. 2.1.3).

Global power generation from geothermal energy is mostly based on **high-enthalpy deposits** that provide heat at high temperatures. These are geological heat anomalies often associated with active magmatism. There are several hundred degrees hot fluids (water/steam) to be found at a depth of a few hundred metres. Their occurrence correlates strongly with active or formerly active volcanic regions, e.g. around the Pacific Ocean or in rift zones and hot spots. Depending on the pressure and temperature conditions, high-enthalpy deposits may be more vapour or water-dominated. Vapour-dominated sites offer temperatures from 240 to 300 °C that produce superheated steam. Liquid-dominated reservoirs are more common with temperatures higher than 200 °C, and they are also close to geological heat anomalies. The hot fluid or steam can be used to provide industrial steam, to feed local and district heating networks or to produce electricity. If electricity is produced, the closed circuit in the circulation system is under pressure so that boiling of the injected water is prevented and the steam is only generated at

[17] Geothermal attractive areas already provide the required heat temperatures near surface and thus allow for a cheaper power generation than in geothermal less attractive areas. Examples of countries with geothermal attractive conditions are New Zealand, Iceland and Italy.

the turbine (flash evaporation). At some sites, the steam was released into the air after use, which could lead to a significant smell of sulphur (e.g. as in Larderello, Italy), but nowadays, it is in general reinjected. As temperature levels of geothermal fluids are quite low (compared to temperatures in conventional steam power plants), the conversion efficiency of geothermal electric plants is also relatively low. The laws of thermodynamics (see Sect. 2.1) limit the efficiency of turbines or other heat engines in extracting useful energy, and thermal efficiency is in general between 10 and 23%. Essential is the selection of a proper working fluid, which is in general not water (as is the case for conventional steam power plants). The essential thermodynamic characteristic of a working medium is its saturation curve with the critical temperature and the critical pressure. Above the critical temperature, the medium can no longer be condensed, no matter how high the pressure is. These variables are essential parameters for selecting a working medium and the best possible adaptation to the temperature profile of the (low temperature) heat source. The potential environmental hazard resulting from the working fluid is, besides its thermodynamic characteristics, an important selection criterion. Examples of used working mediums in geothermal power plants are isopentane and the refrigerant R245fa (1,1,1,3,3 pentafluoropropane).

In non-volcanic areas, the temperatures in the ground can be very different. In general, however, deep drillings for power generation are neccessary as temperatures above 80 °C are required. For commercial use, the fluid temperatures should be above 100 °C. These areas are typically called **low-enthalpy deposits**. In the field of deep geothermal energy (more than 400 m), different types of heat extraction are generally distinguished (see also Fig. 4.30), which of the eligible processes is used depends on the respective geological conditions, the required amount of energy and the required temperature level of heat utilisation:

- **Petrothermal systems** are often referred to as HDR systems (hot dry rock) or enhanced geothermal systems. In case the rock or underground is less permeable, it is artificially cracked. This expands existing rock fissures and so enables the water to freely flow in and out (hydraulic stimulation or fracking). The technique has been adapted from oil and gas extraction techniques. As a result, flow paths are broken or existing ones widened, which increases the permeability of the rock. This procedure is necessary because otherwise, the heat transfer surface and the permeability would be too low. Subsequently, this system of natural and artificial cracks forms an underground geothermal heat exchanger. Consequently, two drillings are necessary, one bringing the carrier medium to the natural and artificial cracks, and through the second, the production well, the carrier medium is conveyed back to the surface.
- **Hydrothermal systems**: if an aquifer provides sufficient energy, thermal waters can be extracted (at one point), used and reinjected at another point (avoiding that the reinjected colder water flows back to the extraction point, resulting in a so-called heat short circuit). Hydrothermal energy can be used to generate heat or electricity, depending on the temperature.

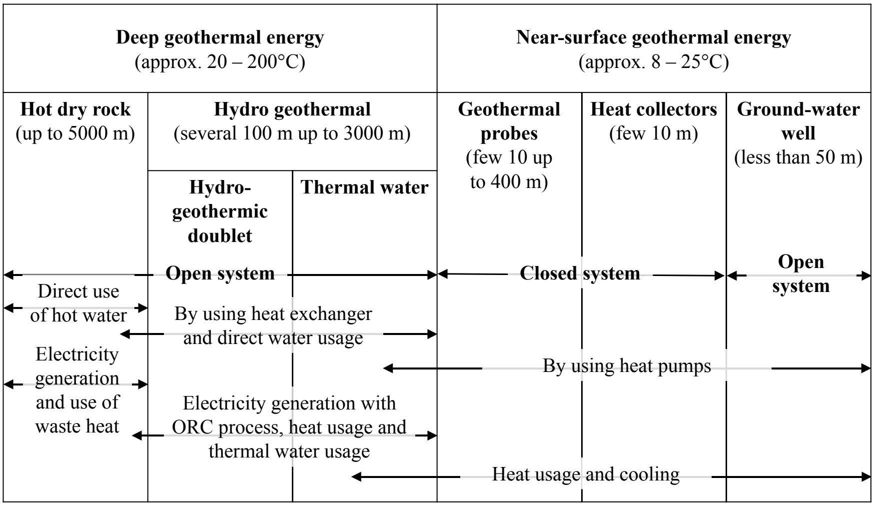


Fig. 4.30 Types of geothermal energy. *Source* Own illustration based on Bundesverband Geothermie (2021)

- A **deep geothermal probe** is a closed system for geothermal energy generation, in which – compared to "open" systems – comparatively little energy is extracted. Closed refers to the fact that the fluid used for heat transport is circulated in a closed system, while in an open system (e.g. hydrothermal system) the fluid is taken from and reinjected to the environment. The probes consist of a single borehole, sometimes drilled over 1000 m deep, in which a fluid is circulated, that is usually trapped in a coaxial tube. The cold heat transfer fluid flows down, is heated at depth in the probe and then rises again in the thinner inside riser. In such geothermal probes, there is no contact with the groundwater. Thus disadvantages of open systems are eliminated, and such systems are hence possible at any location with sufficient geothermal energy. Their withdrawal rate depends not only on technical parameters but also on the mountain temperatures and the conductivity of the rock. However, it will only be a few hundred kW and thus much smaller than a comparable open system. This is because the heat transfer surface is significantly smaller since it only corresponds to the lateral surface of the borehole. Geothermal probes are in general only used for heating and cooling and not for the production of electricity.
- Furthermore, **near-surface geothermal energy** is extracted with heat collectors at varying deeps or with groundwater wells. These are used for heating and cooling purposes, notably using heat pumps, and not for electricity generation.

Geothermal energy is used for electricity production only at a few sites, although it is available everywhere. Yet both the heat flow with only ≈ 0.07 W/m^2 and the temperature increase with the depth of approximately 3 K/100 m make it challenging to access it economically, especially when conventional fuel prices are low.

With the recognition of CO_2 emissions as a problem, a more intensive geological exploration and technical development of geothermal energy started. However, drilling is expensive and several challenges arise, such as: can the rock be made permeable, is sufficient heat available, do seismic events (sometimes interpreted as positive signal for making the rock permeable) result in lacking acceptance and a stop of the projects[18] or are the water amounts in hydrothermal systems sufficient? In contrast, near-surface use of geothermal energy for heating (and cooling) of buildings using heat pumps is usually competitive and thus an interesting option for heating (and cooling).

4.2.5.2 Wave Energy

Waves are caused by the effect of wind forces on the water surface. Their structure depends mainly on the wind speed, the wind duration, the wavelength and the water depth. Wave energy converters attempt to use the energy of these waves for electricity generation or pumping water. Wave power is not to be confused with tidal power. Tidal power makes use of the energy of the falling and rising tide due to the gravitational pull of the sun and the moon. Furthermore waves and tides have to be distinguished from ocean currents which are caused by other forces, e.g. different water temperatures.

Up to now, wave power generation is not widely employed, notwithstanding that several wave power devices have been examined for power generation. Wave energy conversion units can be categorised by the methods used to harness the wave energy, by the location (e.g. shoreline, nearshore and offshore), and by how the power is drawn. Four main technological approaches are distinguished in the following enumeration (see also Fig. 4.31):

- **Point absorber buoys** generate electricity by the rise and fall of swells, e.g. by hydraulic pumps.
- **Surface attenuators** use multiple floating segments connected to one another. These units are oriented perpendicular to incoming waves. Again hydraulic pumps are used to generate electricity taking advantage of flexing motions caused by the swells. A well-known example of this form of using wave energy is the Pelamis Wave Energy Converter which was the first offshore wave conversion unit feeding electricity to the grid from 2004 onwards.
- **Oscillating water columns** consist of air chambers. In these chambers the swells press the air into a turbine with the help of which electricity is produced.
- **Overtopping devices** use the waves to fill an upper reservoir. As the water level is above the surrounding ocean, this potential energy can be transformed into electricity using low-head turbines (see Sect. 4.2.1.2).

[18] As it has been the case at a plant in Basel: the hot dry rock enhanced geothermal project induced seismicity in Basel and led to a suspension of the project. After the induced seismicity a seismic-hazard evaluation was conducted, resulting in the cancellation of the project in December 2009 (cf. Glanz 2009).

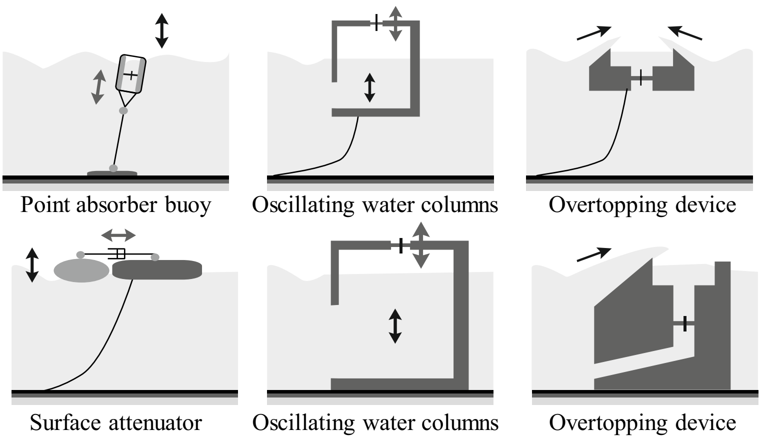


Fig. 4.31 Different technical implementations of wave power plants. *Source* Own illustration based on Graw (1995)

The worldwide resource of coastal wave energy has been calculated to be more than 2 TW (cf. Gunn and Stock-Williams 2012). However, only a few research and demonstration plants in the magnitude of several MW are installed worldwide, so the impact on electricity production has been very limited up to now.

4.3 Key Characteristics of Electricity Generation Technologies

In the following, key characteristics of electricity generation technologies are discussed. Thereby technical, environmental and economic elements are relevant. Also, the availability of the corresponding resources, i.e. primary energy carriers, is a prerequisite. Therefore, this section provides a first, concise overview of these aspects. However, they all deserve closer scrutiny when operating and designing electricity systems since they considerably vary over time and between locations.

4.3.1 Technical and Environmental Characteristics

From an engineering standpoint, the typical plant size and the conversion efficiency are certainly key characteristics of power plants. For future electricity systems, two further aspects are yet relevant: the controllability of the electricity output ("dispatchability") and the flexibility to adjust output to changing circumstances.

In ecological terms, multiple effluents and impacts on eco-systems may be considered. Yet the focal point subsequently is on air-borne emissions and mostly on CO_2 emissions given the primordial challenge of global warming. Other emissions, like e.g. SO_x, NO_x, heavy metals, particular matter, noise, etc., are usually regulated by emission limit values, which are reached with corresponding technical equipment.

Before providing overview tables including relevant information on all these aspects, it is worth reflecting briefly what these indicators measure and what has to be considered when interpreting them.

Typical plant size: these numbers refer to new state-of-the art generation units and are meant to illustrate differences in size between various generation technologies.

Conversion efficiency: this corresponds in general to the ratio of electricity (and heat) output to fuel or other primary energy input. The instantaneous efficiency – typically indicated at reference conditions such as output at nameplate capacity level – has to be distinguished from the annual efficiency, which considers all sorts of real-world losses, e.g. through start-up processes or part-load operation. Subsequently, we indicate the instantaneous efficiency at nameplate capacity since these numbers are better documented than annual efficiencies. The numbers refer to new state-of-the-art generation units.

Dispatchability: this binary indicator highlights when the electricity generation is not dependent on some highly variable input source such as wind or solar energy. Also, other generation units only deliver output if the logistics for the input energy carrier have been secured and if the power plant is available. Yet, these factors are primarily under human control as opposed to weather-dependent energy flows.

Operation lead time: this is one indicator of flexibility – indicating how long it takes to start a power plant from standstill – and hence the ability to adapt to changing demand and supply conditions in the system. Other flexibility indicators would be minimum operation and minimum downtimes, limiting the on-/off-switches of power plants, and the minimum stable operation limit, which informs about the possibilities to modulate output during operation. For the operation lead time, warm start-ups and cold start-ups are differentiated depending on the duration the power plant has been standing idle. Especially in the case of **CCGT** and coal-fired power plants cold start-up times (shown in Table 4.7) are considerably longer than hot start-up times (cf. Agora Energiewende 2017, pp. 44 and 48).

Ramping constraint: this indicator describes the short-term flexibility of a unit in operation, i.e. how fast it can change its output while running (Fig. 4.32).

Technical plant availability: The unavailability of power plants may be due to unplanned outages that typically arise due to some component failure. Or power plants may undergo planned maintenance or retrofit. For variable renewable sources like wind, solar and hydro, resource availability may also restrict power plant output. Yet, the indications given in Table 4.7 only refer to the (yearly average) technical availability, disregarding the possible unavailability of the natural resource.

Capacity and utilisation factor: in contrast to the plant availability, the **capacity factor** describes ex-post, the utilisation of a power plant. The **utilisation factor** (mostly synonymous to full-load hours) is just calculated by dividing the produced

Table 4.7 Key technical and emission indicators for electricity generation technologies

Key indicator	Dispatchability	Net capacity—typical plant K	Net electrical efficiency[a] η	Operation lead time t^{lead}	Ramping constraint $grad^{max}$	Technical plant availability	Operational CO_2 emission factor ε^{fuel} (fuel-specific)
Unit	[.]	[MW_{el}]	[–] (%)	[h;min]	[1/min] (%)	[–] (%)	[t CO_2/MWh_{th} and *MWh_{el}]
Natural gas—CCGT	*y*	480	60	1 h	5–12	~ 85	0.20 *0.33
Natural gas—OCGT	*y*	240	38	5 min	8–12	~ 90	0.20 *0.52
Coal	*y*	770	45	2 h	3–8	~ 85	0.34 *0.75
Nuclear	*y*	1300	36	120 h	3–10	~ 85	0.00
Solar PV—residential	*n*	0.005	100*	N.a	100[b]	~ 92	0.00
Solar PV—commercial	*n*	0.22	100*	N.a	100[b]	~ 95	0.00
Solar PV—large	*n*	2.5	100%*	N.a	100[b]	~ 99	0.00
Onshore wind	*n*	2.5	100*	N.a	100[b]	~ 98	0.00
Offshore wind	*n*	6	100*	N.a	100[b]	~ 95	0.00
Hydro—small	*n*	2	100*	< 1 min	100	~ 95	0.00
Hydro—large	*n*	50	100*	< 5 min	100	~ 95	0.00

[a] applying conventions from energy balances, see Sect. 2.3
[b] within the limits of the time variable natural energy supply
* electrical efficiency of renewable energy plants are set per definition at 100%
Source IEA and NEA (2015), VGB (2013), Dena (2014), own research

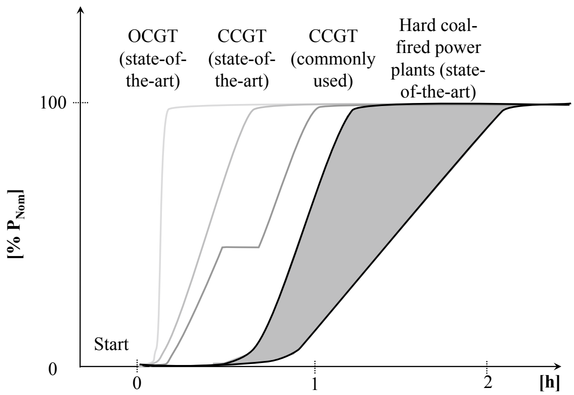


Fig. 4.32 Ramp rates and hot start-up times of selected power plants. *Source* Own illustration based on Agora Energiewende (2017, p. 47)

energy of a year by a plant's capacity. Accordingly, the utilisation factor is between 0 and 8760 h. While the utilisation factor is often used in Europe to describe production of power technologies, in USA typically the capacity factor is specified, which is the utilisation factor divided by 8760 h. In consequence, the capacity factor is between 0 and 1.

Emission factors: these may be given for various pollutants and varying operating conditions. **Emissions** from power generation systems include CO_2, NO_x, SO_x, heavy metals, particular matter, noise, etc. In the following, a focus is set on the CO_2 emission intensity since climate change remains the most challenging global environmental threat and CO_2 emissions from thermal power plants account here for roughly 40% of all emissions (see Sect. 2.3.2). However, other emissions are addressed in Chaps. 2 and 6. We provide emissions per energy unit of input fuel as these figures are only dependent on the fuel used—so they are also applicable for other power plants using the same fuel.

The data compiled in Table 4.8 reveal that conventional electricity generation is typically based on much larger units than renewable generation. Several wind turbines or solar panels may be grouped into utility-scale generation parks with nameplate capacities in the tens or hundreds of MW. Especially offshore wind farms reach connection capacities of more than 1000 MW (e.g. Hornsea, UK). Yet also more small-scale applications and hence more distributed systems exist.

Among the conventional units, the gas-fired **CCGT** plants reach the highest efficiency – and a fortiori the lowest CO_2 emissions per unit of electricity produced. At the same time, **open-cycle gas turbines** (**OCGT**) are the most flexible thermal generation units.

Dispatchability is an advantage of conventional generation units, although large-scale hydroreservoirs and biomass plants are also fully dispatchable (see Sect. 4.2.1 (hydropower) and 4.2.4 (biomass)). In terms of ramping, renewable-based generation is advantageous compared to thermal power plants, as it may usually downregulate (and upregulate, if the natural resource is available) within less than a minute.

Table 4.8 Key economic indicators for electricity generation technologies[19]

Key indicator	Investments c^{inv}	O&M expenditures $c^{O\&M}$	Fuel expenditures p^{fuel}	Technical lifetime T^{life}	Levelized cost of electricity c^{av} [a]
Unit	[€/kW_{el}]	[€/MWh_{el}]	[€/MWh_{th}]	[a]	[€/MWh]
Natural gas—CCGT	810	5	15–75[b]	35	50–150
Natural gas—OCGT	560	12	15–75[b]	25	70–230
Coal	1800	7	5–12	50	60–77
Nuclear	3900	11	1.5–4	50	70–77
Solar PV—residential	1400	24	0	25	108
Solar PV—commercial	1000	30	0	25	90
Solar PV—large	700	18	0	25	60
Onshore wind	1400	19	0	25	65
Offshore wind	3200	36	0	25	104
Hydro—small	4200	30	0	80	87
Hydro—large	2000	8	0	80	35

[a] Based on: interest rate in real terms i = 6%; CO_2 price p^{CO2} = 20 €/t; full-load hours H = 4500 for all conventional plants and according to Table 2.5 for renewables.
[b] In 2022, sudden reductions in Russian gas supplies (related to the war in Ukraine) have sent shock waves through the European gas markets. These have resulted in natural gas prices far above 100 €/MWh, which are much higher than the long-term range mentioned here.
Source IEA & NEA (2015), ISE (2018), own research and computations

In terms of technical availability, photovoltaic plants may reach almost 100%, although ageing may reduce the actual electricity generation, and wind attains approximately 95–98%. Yet, in terms of system **reliability** (cf. Sect. 5.1.4.1), the availability of the corresponding natural resource has also to be considered. For most thermal power stations, including coal, (combined cycle) gas and nuclear power plants, the availability factor ranges between 70 and 90%. Open-cycle gas turbines have even higher availability factors, ranging from 80 to 99%; yet this comes with a relatively rare dispatch (see Sect. 4.4).

4.3.2 Economic Characteristics

The economics of electricity generation technologies are driven by various cost categories that are briefly discussed subsequently. They may be aggregated into one

[19] Note that these values are changing over the years due to different factors, like volatile fuel prices, technological development, volatile material prices, etc. Especially in the last decades, investments and costs of some renewables significantly decreased due to technological learning and this development may continue.

single cost indicator, the so-called levelized cost of electricity generation (LCOE) as introduced below. Levelized cost of electricity can be seen as a benchmark, which allows comparing technologies over their lifetime. However, LCOEs may be less meaningful for weather-dependent renewables as they are not dispatchable. The calculation of LCOEs requires additional assumptions, and also given different application cases, a separate consideration of the cost components is useful.

Investments: investments are made in particular for building the infrastructure and can include expenses for land[20] acquisition, etc. Investments scale roughly proportional to the plant's nameplate capacity, although some economies of scale are observable. Therefore c^{inv} is typically indicated per unit of output capacity. As stated here, these are so-called overnight expenditures, i.e. the pure construction expenditures without financing expenditures during the construction period. These expenditures are often referred as capital expenditures (CAPEX). Capex are often colloquially referred to as "investment costs".

Operation and maintenance (O&M) expenditures: these include the plant's running expenditures, including standard maintenance works but excluding the fuel expenditures.

Fuel expenditures: those tend to fluctuate considerably over the years. Thus only rough indications may be provided. Note that fuel expenditures are typically given on the basis of the upper heating value. Fuel expenditures, in general, include also transport expenditures of the fuel. Consequently, fuel expenditures might be different depending on the location for technologies using the same fuel (e.g. power plant at the coast directly delivered with coal from ocean freighters versus delivery to a location within a country).

Expenditures for greenhouse gas emission allowances: besides fuel expenditures, expenses for greenhouse gas emission allowances might be considered, as it is for example the case for large power plants in Europe, which are under the regulation of the European emission trading scheme (see Sect. 6.2.4.1). Besides emission allowances, also other taxes and expenses might accrue.

Levelized cost of electricity (LCOE): they describe average cost per unit of electricity produced including all cost categories (see Sect. 4.3.3).

Technical lifetime and utilisation period: this parameter is needed to compute the LCOE, since investments must be repartitioned over the utilisation period. Whether the actual utilisation period corresponds to the technical lifetime depends on the economic context. If a continued operation is no longer profitable, decommissioning may occur prematurely. The technical lifetime is also neither strictly technical nor fully predetermined. Typically, the various components of a large plant have different lifetimes. For example, steam turbines in thermal power plants are replaced after around 20 years and gas turbines and motor engines even more

[20] In this context, it is worth noting that land is not depreciable.

frequently. The corresponding cost should be included in a detailed economic assessment or lumped into average operations and management (O&M) costs.

The data compiled in Table 4.8 reveal that the generation technologies not only differ in their technical characteristics but also in their cost components.

Notably, there are also considerable expenditure differences in investment expenditures respetively capital expenditures (CAPEX) and operational expenditures (OPEX). OCGT units are the cheapest in terms of investment expenditure; yet natural gas is considerably more expensive than hard coal or lignite – at least in Europe. Therefore a mix of various generation technologies may be economically most efficient to fulfill the energy demand at different loads. This will be discussed in Sect. 7.1.3 within a coherent theoretical framework.

The results given in Table 4.8 are dependent on the assumptions made. The chosen interest rate $i = 6\%$ in real terms reflects that generation investments are typically quite risky. Experience shows that the CO_2 price is highly uncertain. The CO_2 price p^{CO2} = 20 €/t is somewhat above average values in the European certificate trading system in the last decade, but significantly lower than today's prices. In contrast, for the fuel prices, the mid-point of the indicated range has been chosen. The full-load hours fh = 4500 for all conventional plants are those of a traditional mid-merit plant, whereas for renewables, they reflect average resource qualities across Europe.

Under these conditions, CCGT and coal-fired plants are the most cost-effective plants – which is in line with their traditional role as mid-merit plants in the European generation fleet (see Sect. 4.4.1.1 and Chap. 7). Wind onshore has LCOE in the same range, PV even lower at location with high sun availability. Yet, in economic terms, the wind (and PV) electricity is expected to be less valuable since it is not dispatchable instead of the electricity from thermal plants. On the other hand, the applied (low) CO_2 price certainly does not fully reflect the future damage cost of CO_2 emissions. For solar and offshore wind, the technology cost, on the other hand, may not reflect the latest progress in manufacturing cost – hence all these numbers should not be overinterpreted, but rather be considered first indications in economic terms.

4.3.3 Levelized Cost of Electricity

Levelized costs of electricity (LCOE)[21] are a simplifying indicator to compare the average costs of different generation technologies. They are defined as:

$$\text{LCOE} = \frac{\sum_{t=0}^{\mathrm{T}^{\text{life}}} \left(C_t^{\text{inv}} + C_t^{\text{var}} + C_t^{\text{decom}} \right) \cdot (1+r)^{-t}}{\sum_{t=0}^{T^{\text{life}}} E_t \cdot (1+r)^{-t}} \tag{4.8}$$

[21] Levelized costs of electricity are in general calculated on the basis of expenditures. From a stringent terminological perspective, these should correctly be named as levelized expenditures of electricity. However, the term levelized costs of electricity is established in the energy industry, so that it is also used in this textbook.

Thereby C_t^{inv} stands for the investments and C_t^{var} for the variable expenditures. Additionally, the expenditures of decommissioning C_t^{decom} are included and all expenditure terms are discounted using the interest rate r and the lifetime T^{life} of the power plant.

In the denominator, the annual energy quantities E_t are also discounted to provide a constant payment per unit of energy produced, which would allow an investor to recover its entire expenditures over the lifetime of a generation facility.

The terms on the right-hand side may be computed in a simplified way as follows:

$$C_0^{\text{inv}} = c^{\text{inv}} \cdot K \tag{4.9}$$

Here only the upfront investment C_0^{inv} is considered, no reinvestments for replacing larger parts such as turbines nor financing cost during the construction period. For the variable expenditures, fuel and CO_2 certificate prices (as well as any further taxes and expenses) have to be considered together with operation and maintenance expenditures:

$$C_t^{\text{var}} = \left(\frac{p_t^{\text{fuel}} + \varepsilon^{\text{fuel}} \cdot p_t^{\text{CO2}}}{\eta} + c_t^{O\&M} \right) \cdot E_t \tag{4.10}$$

Finally, decommissioning expenditures should be included; yet discounted decommissioning expenditures are almost negligible except for nuclear plants.

The annual electricity production E_t may be written as a product of the installed capacity and the so-called full-load hours fh (see Sect. 2.1). Those may alternatively be expressed as a product of the capacity factor cf and the number of hours per year H:

$$E_t = fh \cdot K = cf \cdot H \cdot K. \tag{4.11}$$

But this requires an estimate of the capacity factor – for renewables like wind and solar, it will be mostly resource dependent (see below); yet for the dispatchable generation technologies, the capacity factor will depend on the system context (which other generation technologies are present) and on technology characteristics (what are the variable costs). Chapter 7 discusses how to determine equilibrium results regarding the generation fleet and its calculation.

The results given in Table 4.8 are obviously dependent on the assumptions made. The chosen interest rate $i = 6\%$ in real terms reflects that generation investments are typically quite risky. In periods with high (risk free) central bank rates and scarce project resources, the required return on investments may even raise much higher. Experience shows that the CO_2 price is highly uncertain. The CO_2 price $p^{\text{CO2}} = 20$ €/t is somewhat above average values in the European certificate trading system in the last decade, but significantly lower than today's prices, whereas. In contrast, for the fuel prices, the mid-point of the indicated range has been chosen. The full-load hours $fh =$ 4500 for all conventional plants are those of a traditional mid-merit plant, whereas for renewables they are reflective of, they reflect average resource qualities across Europe.

Under these conditions, CCGT and coal-fired plants are the most cost-effective plants – which is in line with their traditional role as mid-merit plants in the European generation fleet (see Sect. 4.4.1.1 and Chap. 7). Wind onshore has LCOE in the same range. Yet, in economic terms, the wind electricity is expected to be less valuable since it is not dispatchable in contrast to the electricity from thermal plants. On the other hand, the applied CO_2 price certainly does not fully reflect the future damage cost of CO_2 emissions. For solar and offshore wind, the technology cost, on the other hand, may not reflect the latest progress in manufacturing cost – hence all these numbers should not be overinterpreted, but rather be considered first indications in economic terms.

4.4 Scheduling Electricity Generation—The Unit Commitment and Dispatch Problem

Traditionally, regional or national integrated utilities have run the electricity system. These had an assigned service area where they had to serve the demand. This traditional organisation model still prevails in many regions of the world, including parts of the USA. In such a setting, the utility has to serve all the demand arising in its service area – this demand being labelled frequently as "load" by engineers. We discuss different organisational structures of the electricity sector in Chap. 8, and demand is considered in detail in Chap. 3. But even under modified institutional settings and with flexible elements in electricity consumption, the challenge remains to match the time-varying demand for electricity with different generation units, some of which may also have time-varying output. This scheduling problem is frequently subdivided for conventional large power plants in the unit **commitment** and dispatch decisions. Unit commitment thereby describes the decision to turn units on and off, whereas the **dispatch** decisions consist in selecting the actual output level for the operating units.

The **scheduling** problem may be considered at different time scales, but one key operational planning approach is to schedule electricity generation on a day-to-day basis. In Sect. 4.4.1, we discuss this scheduling problem in detail to provide a flavour of the operational and mathematical challenges linked to the management of electricity systems worldwide. Section 4.4.2 then broadens the scope and discusses further issues and problems in operational planning in a less formal way.

4.4.1 Day-Ahead Operational Planning

The traditional operational planning (in a regulated market) is conceptually at first sight relatively simple: there is one entity, the (operational planning department of an) electric utility, aiming at making the best use of existing assets (generation units) to meet well-defined objectives. Its prime objective is (or at least should be) obviously to meet the demands from the customers. An additional objective from the perspective of the customers or society should be to do so at minimum cost. Further objectives may be relevant (e.g. meeting specific emission targets); yet we subsequently limit the analysis to the first two objectives in view of a concise treatment.

The first objective thereby sets a clear requirement or constraint to the operations, whereas the second objective is framed as reaching an optimum of a certain function, namely the cost of system operation. Hence we can formulate operational planning as a constrained optimisation problem (see Sect. 4.4.1.3). Alternatively, we may solve simplified versions of the problem graphically (see Sects. 4.4.1.1 and 4.4.1.2). In contrast to traditional operational planning in a regulated market, utilities in a liberalised market compete against each other and maximise their individual profit. At first glance, the traditional model appears to be no longer useful for this planning task. However, under the assumption of perfect competition and an inflexible (or nearly inflexible) demand, it can be shown that both planning problems result in an identical market equilibrium from a system perspective. Consequently, the described planning approach is also relevant (under some assumptions, see also Sec. 7.1.2) in a liberalised electricity market.

4.4.1.1 Simple Scheduling of Power Plants: Merit-Order Approach

Before providing a full mathematical description of the scheduling problem, one may consider a simplified version of the problem for thermal power plants. This so-called **merit-order approach** can be described and solved graphically and hence provides an easily accessible first-order approximation for the scheduling of thermal power plants.[22] The basic idea is that cheap generation units should be used first to meet the demand if costs are to be minimised. Since investments and other fixed costs are not influenced by operation decisions, only variable costs are considered – e.g. wind, solar and nuclear are used with highest priority, even if there has been important upfront investment. Additionally, none of the units may exceed its nameplate capacity and in the case of non-dispatchable renewables such as solar and wind, the output is also limited by the currently available input energy (e.g. no solar radiation at night).

We therefore order the available generation units by increasing variable cost and put their available capacities successively on the horizontal axis, while indicating their variable cost on the vertical axis. This is the merit order or supply stack as illustrated in Fig. 4.33.

Adding the demand level as a vertical line in the graph, we immediately obtain the units operating to meet demand (see Fig. 4.33). The same merit order may be used for each time period of the planning horizon – unless the availability of specific generation units varies over time (e.g. wind). In our simple example, the low load at night would require the operation of the wind (which is assumed to be available in this example), nuclear and coal units. During the daytime, utilisation of the coal unit would be increased and additionally, the CCGT is needed to meet demand.

[22] Note that a similar approach may also be used to analyse electricity market equilibria and therefore will be discussed in-depth in Chap. 7.

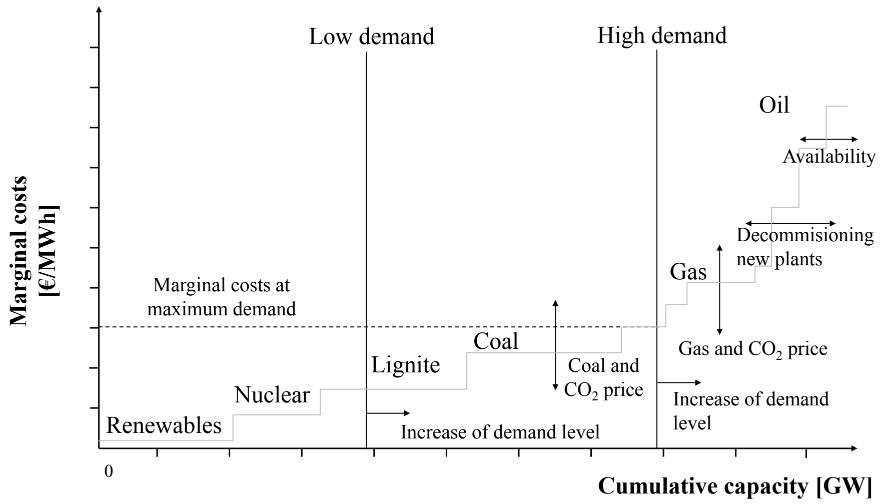


Fig. 4.33 Merit order for an exemplary power plant portfolio

This solution to the generation scheduling problem is straightforward. In general, it does not yet provide feasible operation schedules for the generation units. The reasons are operational constraints that prevent the flexible adjustment of power plant output to the requested levels. What the main restrictions are and how they may be handled is discussed in Sect. 4.4.1.3.

4.4.1.2 Simple Scheduling of Reservoir Power Plants

Hydroreservoir plants do not fit directly into the supply stack model discussed in the previous section. The reason is that their generation output does not come at any cost at first sight. This is true concerning so-called pagatory or cash-effective cost, i.e. cost corresponding to actual cash out. Yet, there is an indirect cost of producing electricity from a hydroreservoir: the opportunity cost of not being able to use the hydroenergy at some other point in time. Using water now means losing an opportunity to use the water later. The question then arises of how this opportunity cost may be determined.

This may be done by considering the profit maximisation problem for electricity production in reservoir power plants over time. To do so, the water inflow and the electricity prices for a given (future) period have to be estimated. When neglecting water inflow and reservoir restrictions, the optimal strategy is to use the available energy to produce electricity at the highest electricity prices. Three steps are necessary to identify these highest prices: (1) expected electricity prices are sorted in descending order (the so-called **price duration curve**, analogue to the sorted annual load curve, see Sect. 3.1.3). (2) Calculate the expected amount of water inflow (preferably converted in hours of full-load), (3) determine the intersection of the price duration curve and full-load hours (based on water inflow). The determined value describes the minimum expected electricity price at which the plant is

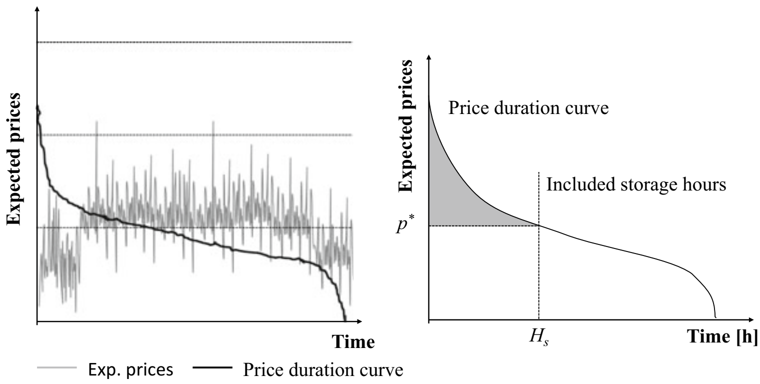


Fig. 4.34 Schematic illustration of the dispatch of a reservoir power plant in a given planning period

dispatched (see Fig. 4.34). This minimum price is also called "**water value**", reservation price or opportunity costs of water and is depicted as p^* in Fig. 4.34. In consequence, the water is used to produce at the highest prices in the given planning period. However, this procedure requires sound estimates of the water inflow as well as electricity prices.

The optimal use of pumping in **pump storage plants** (as well as the charging of other storage technologies, cf. Sect. 5.2) can be derived analogously. Therefore, the mirrored price duration curve has to be depicted starting from the y-axis, respectively, from the hydro inflow starting point H_a if hydro inflows occur (see Fig. 4.35). Additionally, the curve has to be stretched in the y-direction with 1/round-trip-efficiency (see Sect. 4.2.1) to account for efficiency losses and compressed in the x-direction with a term, which is calculated by multiplying the round-trip-efficiency by pump power and dividing this by turbine power, to account for the different capacity of the pumping process. In the case of a daily storage limit H_e, an upper price limit p_{t^*} can be derived, at which water is used for electricity production (with the water turbine) as well as a lower price limit p_{p^*}[23] at which water is pumped. In between the upper and the lower price level, the plant will not be operated. If the storage limit is not binding, the price for producing electricity corresponds to p_{s^*}. The price for pumping has to be determined by retransforming p_{s^*} by multiplying with the round-trip efficiency.

This schematic and simplified concept provides a rough understanding of the basic dispatch principles of storage. In reality, several restrictions have to be considered (and sometimes also hydrocascades), necessitating more complex dispatch models. A model which includes the most common restrictions, namely reservoir volume constraints, is discussed in the next section.

[23] As the pumping cost curve has been transformed with the round-trip-efficiency, a retransformation of this pumping price is necessary by multiplying the price with the round-trip-efficiency of the plant. In consequence, the price at which electricity is pumped is slightly less than p_{p^*}.

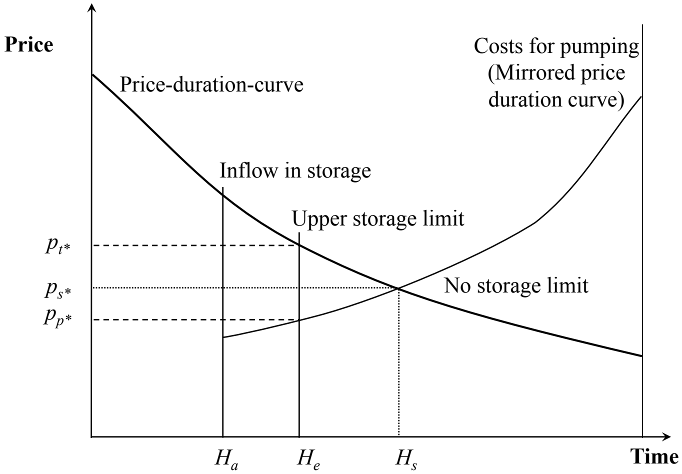


Fig. 4.35 Schematic illustration of the dispatch of a pumped hydropower plant in a given planning period

4.4.1.3 General Model of Short-Term Power Plant Scheduling

Production scheduling is a topic of relevance not only to the energy industry but also to almost any industrial production, be it for cars, washing machines or computers. Also, for services like airline travel, container shipping, or hotel stays, scheduling problems arise. Operations Research, as the branch of applied mathematics dealing with this kind of problems, has correspondingly developed multiple problem formulations and solution methods to support decision making in this field. The power plant scheduling problems have some specificities linked to the particularities of electricity, notably its limited storability and the characteristics of the generation units.

An immediate implication of the limited storability of electricity is that electricity has to be produced at precisely the point in time when it is needed. Correspondingly, production scheduling has to be done with a sufficient, at least hourly, time granularity.[24] The limited flexibility of many large-scale generation units (see Sect. 4.3.1) and the existence of limited storage capabilities (see Sects. 4.4.1.2 and 5.2) imply that the scheduling has to be done simultaneously for different time periods, including time-coupling constraints.

Consequently, the objective of the scheduling is to minimise the sum of operational costs C_{op} over all planning time steps $t \in \{1, \ldots, T\}$:

$$\min_{y_{ut}, s_{ut}, o_{ut}, L_{ut}, y_{ut}^{ch}} C_{\mathrm{op}}$$

[24] The lack of storability and the limited predictability of demand also imply that some fast-reacting reserves have to be foreseen in system operation, so that the system may adjust to unforeseen disturbances in demand or supply. This issue is neglected here, but will be taken up again in Sect. 5.1.4.2 and in Sect. 10.3.

$$C_{op} = \sum_{t=1}^{T} \sum_{u} c_u^{var} \cdot y_{ut} \cdot \Delta t + \sum_{t=1}^{T} \sum_{u} \left(c_u^{\text{start}} \cdot s_{ut} \right) \tag{4.12}$$

Costs are thereby also summed over all units u, and they comprise not only the variable cost c_u^{var} related to the output y_{ut} of the units over the time step duration Δt, but also additional cost related to start-ups s_{ut}. Starting a large thermal power plant requires additional fuel (for heating up the components) and induces extra wear and tear related to thermal stress of the components. These costs are here collapsed into the cost term c_u^{start}, also neglecting that these costs may be dependent on the duration of the preceding shutdown period. Restarting a cold unit typically induces higher cost than restarts after short cool-down periods. Further variables of the optimisation problem are y_{ut}^{ch} describing the used electricity for pumping (storage charging), L_{ut} as storage level in period t of plant u, o_{ut} as binary variable of operation (on/off).

The cost minimisation is constrained by a number of restrictions which reflect system requirements and unit flexibility limitations. Note that all variables are restricted to positive values.

1. The main system requirement is the aforementioned demand–supply balance, which has to be fulfilled in each time segment with D_t as demand:

$$\sum_{u} y_{ut} \cdot \Delta t = D_t \cdot \Delta t \qquad \forall t. \tag{4.13}$$

The for-all operator $\forall$ thereby indicates that this constraint is not only to be applied for one particular t, but rather for all time segments of the planning horizon.

2. At the unit level, the limitation of the generation output to the installed capacity K_u is the fundamental requirement, which has to be fulfilled for each unit at each time step:

$$y_{ut} \leq K_u \cdot o_{ut} \qquad \forall u, t. \tag{4.14}$$

Yet this requirement is slightly modified here by introducing the binary on/off variable o_{ut}. If the unit is not operating, i.e. when $o_{ut} = 0$, the output level is also limited to zero.

3. The preceding upper bound to unit output is complemented by a lower bound, reflecting that there is usually a minimum output level required for stable operation. This is only relevant as long as the plant is turned on ($o_{ut} = 1$); otherwise, the minimum output drops to zero. Therefore the minimum operation restriction reads:

$$y_{ut} \geq P_u^{\min} \cdot o_{ut} \qquad \forall u, t. \tag{4.15}$$

4. The binary operation variables o_{ut} and the start-up variable s_{ut} are obviously linked. The simplest way to express the relation is by writing:

$$s_{ut} \geq o_{ut} - o_{u,t-1} \qquad \forall u, t. \tag{4.16}$$

This relationship, together with strictly positive start-up costs and the objective of cost minimisation, ensures that the start-up variable s_{ut} is zero throughout except when the on/off variables are $o_{ut} = 1$ and $o_{u,t-1} = 0$.

5. For variable renewable sources, generation is not only restricted by the installed generator capacity but also by the currently available supply, which may be described as a fraction w_{ut} of the installed capacity.[25]

$$y_{ut} \leq w_{ut} \cdot K_u \qquad \forall u, t. \tag{4.17}$$

This is notably relevant for photovoltaics, wind and run-of-river hydroplants.

6. For storage plants u' (notably hydroreservoirs), the storage fill level $L_{u't}$ is treated as an additional variable, which has to fulfil a dynamic energy balance equation:

$$L_{u't} \leq L_{u',t-1} + \left(i_{u't} + \eta_{u'}^{cyc} y_{u't}^{ch} - y_{u't}\right) \cdot \Delta t \qquad \forall u, t. \tag{4.18}$$

Thereby $i_{u't}$ describes natural inflow into the storage, which is only relevant for hydroreservoirs, whereas $y_{u't}^{ch}$ describes electricity used for filling the reservoir (e.g. through pumping in case of pumped hydrostorage). Note that for convenience, the filling level of the storage is directly expressed in energy units (e.g. MWh), whereas inflows and outflows $i_{u't}, y_{u't}^{ch}$ and $y_{t,u'}$ are described as power flows (with MW as unit or similar). Correspondingly the latter have to be multiplied by the length of the time segment Δt. Additionally, the charging power $y_{u't}^{ch}$ is multiplied by the round-trip or cycle efficiency $\eta_{u'}^{cyc}$, which considers losses both in the charging and in the discharging processes (e.g. pumps and turbines).

7. The storage filling level is not only non-negative like the other variables but also limited by the storage volume $V_{u'}$

$$L_{u't} \leq V_{u'} \qquad \forall t, u'. \tag{4.19}$$

8. For thermal power plants, starts and shutdowns put thermal stress on plant components. Therefore, their operation is usually further restricted by a minimum operation time T_u^{opmin} and a minimum shutdown time T_u^{sdmin}. There are

[25] Note that w_{ut} may also be viewed as a "momentaneous" **capacity factor** for unit u at time t.

several ways of implementing these restrictions; a rather compact one is[26]:

$$\sum_{\tau=t-T_u^{\mathrm{opmin}}+1}^{t} s_{u\tau} \leq o_{ut} \qquad \forall u,t$$

$$\sum_{\tau=t-T_u^{\mathrm{sdmin}}+1}^{t} s_{u\tau} \leq 1 - o_{u,t-T_u^{\mathrm{sdmin}}} \qquad \forall u,t. \tag{4.20}$$

Further restrictions may complement these constraints to include, e.g. limited ramping for units during operation or specific constraints for CHP units. But the problem structure always remains similar: it is an optimisation problem with a linear objective function and linear constraints[27]; yet some of the variables only take discrete values, in occasion the unit commitment variables are binary variables.

Hence the problem at hand leads to a so-called mixed-integer programme, a programming class widely discussed and analysed in operations research over the last decades. These problems may be formulated in compact matrix notation as:

$$\min_{x} \quad \mathbf{c}^{\mathbf{T}} \cdot \mathbf{x}$$

$$s.t. \quad \mathbf{A} \cdot \mathbf{x} \geq \mathbf{b} \tag{4.21}$$

The vector of variables $\mathbf{x}$ thereby includes the generation quantities y_{ut} the operation and startup variables o_{ut} and s_{ut} and the storage filling level L_{ut} and charging y_{ut}^{ch}. The cost-vector $\mathbf{c}$ includes multiple instances of the coefficients c_u^{var} and c_u^i for the different time steps. Each line of the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ corresponds to the coefficients of one particular restriction.

For the solution of pure linear programmes (LPs) without binary variables, computationally efficient standard methods are available, the best-known of which is the so-called simplex algorithm. By contrast, mixed-integer linear programmes (MILPs) require more advanced techniques, e.g. the so-called branch and bound algorithm, frequently complemented by heuristic strategies. The computation time rises rapidly with the number of integer variables considered. The MILP problems are even known to be NP-hard, i.e. that no algorithm guarantees in general that the solution time only increases polynomially (and not exponentially) with the number of binary variables. Nevertheless, day-ahead unit commitment and dispatch problems with dozens of units may nowadays be solved numerically within minutes.

[26] Cf. Rajan and Takriti (2005).

[27] Nonlinearities may be approximated through piecewise linear functions.

4.4.2 From Day-to-Day Planning to Portfolio Management

The scheduling problem described in the previous section has been and is at the heart of utility operations with regional monopolies (see Table 4.9, framed box). Notably in these cases, each electrical utility has an assigned service area, with an exogenously given "load" (demand). The scheduling of power plants then clearly aims to meet load with as low costs as possible.

Since the deregulation of the electricity industry (see Chap. 6), electrical utilities compete for customers. They, therefore, do not have a fixed load, but rather the demand they have to meet is a result of trading and sales activities. Hence the previously discussed day-ahead scheduling problem is particularly relevant after (day-ahead) spot market results have been published; i.e. prices as well as purchase and sales quantities are known (see last two lines of Table 4.9).

Before closure of spot trading, some planning activities are still necessary, even in deregulated markets, e.g. linked to the procurement of fuels or the scheduling of maintenance and hydroreservoirs. But then, the operation of the power plants is not to be considered as driven by a predetermined, fixed electrical load but rather by

Table 4.9 Scheduling problems at various time horizons and in different market types

Planning horizon	**Year ahead**	**Week ahead** month ahead	**Day ahead** **before spot market auction**	**Day ahead** **after spot market auction**	**Intraday** during intraday trading	**Real time** after intraday trading
Key decisions	Maintenance scheduling; Hydro reservoir planning; Fuel procurement	Unit commitment baseload units; Hydro reservoir planning	Bid submission to day-ahead-spot market	Unit commitment; Dispatch	Change in dispatch	System balancing
Integrated utility with regional monopoly						
Objective function	Min cost	Min cost		Min cost		Min cost
Load	Fixed	Fixed		Fixed		Fixed
Generation company in deregulated markets						
Objective function	Max op. profit	Max op. profit	Max op. profit	Min cost	Max op. profit	
Load	Variable	Variable	Variable	Fix	Fixed but adjustable	

(anticipated) market prices. The power plants will operate as long as they can earn money, or more precisely: as long as their operating margin is positive.

The previously defined scheduling model may be modified to account for the setting of a liberalised electricity market. The main change is thereby in the objective function, which now reflects the **maximisation of operation profits** (EBITDA), i.e. the difference of market-based revenues and the previously considered operational costs:

$$\max_{y_{t,u}, s_{t,u}, o_{t,u}, l_{t,u} y_{t,u}^{ch}} R - C_{op}$$

$$R = \sum_{t=1}^{T} \sum\nolimits_{u} p_{t,u} \cdot y_{t,u} \cdot \Delta t. \tag{4.22}$$

Pushing one step further, one may consider the price uncertainties on the procurement and sales markets and include different available products and contracts on the procurement and sales markets. Then methods of financial portfolio management may be applied to electricity generation and storage portfolios, albeit some characteristic differences have to be considered. Also, as a prerequisite, future prices have to be available with sufficient granularity. This generally requires building hourly price forward curves – a topic discussed in Sect. 11.2.

Overall, power plant scheduling is not limited to the day-ahead planning horizon. Rather, a broad range of scheduling problems arise in both types of markets (see Table 4.9) that are briefly explained subsequently, starting in the order of the table:

Long-term scheduling (year ahead): this is notably required for maintenance planning of thermal power plants, especially baseload power plants. Also, water management for seasonal (large) reservoir hydropower plants and the management of so-called take-or-pay contracts for fuels (especially natural gas) with a minimum annual purchase quantity require considering long planning horizons due to long-term quantity constraints.

Medium-term scheduling (week and month ahead): this is primarily done to manage (pumped-) storage hydropower plants. Also, unit commitment decisions for coal-fired power plants may be taken more on a week-ahead than day-ahead basis.

Short-term scheduling (day-ahead): this encompasses the unit commitment and dispatch planning for the following day after the publication of the day-ahead spot market results (generally after 12 p.m.). This is still the most important scheduling process in the current competitive market.

Very short-term scheduling (intraday): here, replanning of (unit commitment and) dispatch is done with a time horizon of less than one day. Portfolio managers reschedule especially to cope with unexpected events like power plant outages or renewables forecast errors or participate in intraday trading.

System balancing (real-time): this encompasses the real-time compensation of load and infeed variations and power plant outages. In deregulated electricity markets, this task is the responsibility of the grid operator. He thereby dispatches reserve power, which has been contracted from power plant operators beforehand.

The planning problems increase in size, computation time and uncertainty with longer planning horizons. Long-term planning is generally done with a lower degree of details in modelling to avoid intractable or computationally burdensome problem formulations, e.g. regarding approximation of nonlinearities or the modelling of small units.

4.5 Further Reading

Kaltschmitt, M., Streicher, W., & Wiese, A. (Eds.) (2007). Renewable Energy – Technology, Economics and Environment. Berlin, Heidelberg: Springer.

The book Renewable Energy presents the physical and technical principles of promising ways of utilising renewable energies. The book gives a detailed overview of the different renewable energy technologies for electricity production and heat provision.

Nag, P. K. (2014). Power Plant Engineering. 4th edition. New Dehli: McGraw Hill Education (India).
Hegde, R. K. (2015). Power Plant Engineering. Pearson India.

Both books provide an overview of power generation technologies and conceptual knowledge about power plants engineering.

Wood, A., Wollenberg, B., & Sheblé, G. (2013). Power Generation, Operation, and Control. 3rd edition. Hoboken, New Jersey: Wiley.

This book gives – among other things – a comprehensive introduction into the economic dispatch and unit commitment problems.

Strauss, K. (2016). Kraftwerkstechnik – zur Nutzung fossiler, nuklearer und regenerativer Energiequellen. 7th edition. Berlin, Heidelberg: Springer.

The book of Strauss provides an overview of power generation technologies and indicates future development opportunities. It gives an overview of available energy sources (fossil, regenerative, nuclear), the principles of converting the respective primary energy into electricity, environmental pollution resulting from the energy conversion, and statements about efficiency, system availability and costs.

4.6 Self-check of Knowledge and Exercises

Self-check of Knowledge

(1) Explain the steam cycle of a steam turbine and illustrate the cycle process in a p–V-diagram.
(2) Explain the Joule cycle of a gas turbine and illustrate the Joule cycle in a p–V diagram.
(3) Describe a gas combined cycle and give examples of technologies used for combined heat and power plants.
(4) Explain the mass defect and what it has do to with nuclear energy.
(5) Differentiate different types of hydropower and describe different turbine types as well as their general field of application concerning head height and water flow.
(6) Formulate the basic physical principle to calculate the power of wind and explain the typical power curve of a wind power plant.
(7) Compare solar thermal power plants and photovoltaic by showing the basic principle and their key characteristics.
(8) State a minimum of three different types of inputs in biomass plants, name the three principles types of conversion and arrange which input can be used in which conversion.
(9) Characterise a minimum of three technologies with their key techno-economic characteristics.
(10) Sketch an illustrative merit-order curve of the power plant portfolio in your country of residence.
(11) Illustrate the water value of a reservoir power plant with the help of a schematic illustration of the dispatch of a hydropower plant.
(12) Formulate the general mathematical model of short-term power plant scheduling.

Exercise 4.1: Hydropower Plant

You are the operator of a hydropower plant at the Ruhr built in 1957 and not modernised since. The river Ruhr has a barrage at the power plant with a height difference of 6.5 m between the upper and lower water. The Ruhr flows through the barrage in autumn with a constant flow of 1.7 million m^3 per day. However, 5% of the water flow has to be used for a fish ladder and is therefore not available for power generation by the turbine. In addition, the turbines can be switched off, whereby the water around the turbines is channelled through side channels so that the level in the upper water basin always remains constant.

Furthermore, the following assumptions are given:

- Electrical efficiency of the power plant: 87%.
- Density of water: 1000 kg/m^3.

- Gravitational acceleration (gravitational constant): 9.81 m/s^2.

(a) Which type of turbine would you choose? What are the advantages of this selection?
(b) What is the maximum electrical output that the power plant can provide under these conditions?

Exercise 4.2: Wind Power Plant
Calculate the theoretical annual energy yield (in kWh) of the following wind turbines with standard air density at 25 °C and a yearly availability of 95%:

You have the following information to complete the task:

$$P_{\text{WEA}} = \frac{1}{2}\rho \cdot A \cdot v^3 \cdot c_p$$

ρ air density (1184 kg/m^3 at 25 °C).
A rotor area [m^2].
v wind speed [m/s].
c_p performance coefficient (system parameter).

(a) **At coastal area**: (constant) wind speed v = 8 m/s; (Performance coefficient of wind power plant at 8 m/s = 0.5)

Rotor diameter wind energy plant 1 = 71 m

Rotor diameter wind energy plant 2 = 90 m

(b) **inland**: (constant) wind speed v = 4 m/s, (Performance coefficient at 8 m/s of wind power plant = 0.4)

Rotor diameter wind energy plant 1 = 71 m
Rotor diameter wind energy plant 2 = 90 m

Calculate the full-load hours of the wind turbines mentioned (taking the produced energy at a constant wind speed), if all wind turbines are each equipped with a generator with a nominal electrical output of 2000 kW.

Exercise 4.3: Levelized Cost of Electricity

(a) Calculate the levelized costs of electricity of the following four technologies with the respective data given in the table.

The following data are given: i = 6%, CO_2 price of 15 €/t.

(b) What is the impact of an emission prices increase to 100 €/t on the LCOE?

Key indicator	Investment cost c^{inv}	O&M cost $c^{O\&M}$	Fuel cost p^{fuel}	Technical lifetime T^{life}	Full-load hours
Unit	[€/kW$_{\mathrm{el}}$]	[€ / MWh$_{\mathrm{el}}$]	[€ / MWh$_{\mathrm{th}}$]	[a]	[€/MWh]
Natural gas—CCGT	810	5	15–45	35	5000
Coal	1800	7	5–12	50	7000
Solar PV—residential	1200	24	0	25	1300
Onshore wind	1500	19	0	25	2500

Exercise 4.4: Scheduling of Power Plants

You are responsible for the power plant scheduling of a utility. You have to prepare the power plant schedule for the next day based on forecasts for load and for wind generation as depicted in the following table:

Hours	Load (MWh$_{\mathrm{el}}$/h)	Generation wind (MWh$_{\mathrm{el}}$/h)	EEX prices (€/MWh$_{\mathrm{el}}$)
1–8	1200	0	40
9–16	2300	250	35
17–24	1700	400	20

You are responsible for the scheduling of the three power plants in the portfolio with the following technical and economic data:

Power plants	Variable costs (€/MWh$_{\mathrm{el}}$)	Maximum capacity (MW)	Minimum stable operation limit (i.e. minimum output in operation) (MW)
CCGT	38	180	80
Hard coal	30	750	150
Nuclear power	5	1200	320

(a) **Determine the cost minimising schedule of the portfolio**: Use the classical merit-order approach to determine the unit commitment and dispatch for the given time steps and indicate the generation output of each technology for each time step. Determine the total costs of generation.

(b) **Determine a profit maximising schedule of the portfolio**: Your boss asks you to analyse the possibility to sell and buy electricity at the EEX spot market. Expected prices are given in the above (first) table. Determine the optimal operation (i.e. the generation in each time step). Also, indicate in which hours you buy/sell which amount of electricity at the EEX spot market. Finally, compute the savings that can be achieved in comparison with a). What are the main differences between the two solutions and explain where they come from?

(c) **Determine the cost minimising schedule of the portfolio taking minimum stable operation limits into account**: You recognise that your solution in a) is insufficient since minimum operation limits of the power plants (see second table above) have not been considered so far.

   i. Determine first the time steps in which the solution violates the minimum stable operation limits.
   ii. Compute for these time steps the optimal unit commitment and dispatch with minimum operation limit. (*Hint: Compute **all** meaningful solutions and determine the optimal solution by comparing the respective changes in the objective function value.)*

References

Agora Energiewende. (2017). *Flexibility in thermal power plants—With a focus on existing coal-fired power plants*. Agora Energiewende. Available at: https://static.agora-energiewende.de/fileadmin/Projekte/2017/Flexibility_in_thermal_plants/115_flexibility-report-WEB.pdf [Accessed May 15, 2022].

Anderson, J. (2011). *Fundamentals of aerodynamics* (5th ed.). McGraw-Hill Series in Aeronautical and Aerospace Engineering.

Balussou, D. (2018). *An analysis of current and future electricity production from biogas in Germany*. Dissertation: Karlsruhe Institute of Technology (KIT). Available at: https://publikationen.bibliothek.kit.edu/1000084909 [Accessed May 15, 2022].

BDEW. (2019). *Leitfaden "Stromkennzeichnung"*. Bundesverband der Energie- und Wasserwirtschaft. Available at: https://www.bdew.de/media/documents/Leitfaden_Stromkennzeichnung_08-2019.pdf [Accessed August 12, 2022].

Bidart, C. (2013). *A techno-economic assessment of the generation and usage of biogenic gases in Chile as a substitute of natural gas*. Dissertation: Karlsruhe Institute of Technology (KIT). Available at: https://publikationen.bibliothek.kit.edu/1000038622 [Accessed May 15, 2022].

Bundesverband Geothermie. (2021). *Lexikon der Geothermie - Geothermie*. Available at: https://www.geothermie.de/bibliothek/lexikon-der-geothermie/g/geothermie.html [Accessed May 15, 2022].

dena. (2014). *dena-Studie Systemdienstleistungen 2030*. Deutsche Energie-Agentur.

Giesecke, J., Heimerl, S., & Mosonyi, E. (2014). *Wasserkraftanlagen - Planung, Bau und Betrieb* (6th ed.). Springer.

Giraldo, J., Gotham, D., Nderitu, D., Preckel, P., & Mize, D. (2012). *Fundamentals of nuclear power*. State Utility Forecasting Group. Available at: https://www.purdue.edu/discoverypark/sufg/docs/publications/SUFG%20nuclear%20report.pdf [Accessed May 15, 2022].

Granet, I., & Bluestein, M. (2014). *Thermodynamics and heat power* (8th ed.). CRC Press.

Graw, K.-U. (1995). *Wellenenergie - eine hydromechanische Analyse*. Bergische Universität - Gesamthochschule, Lehr- und Forschungsgebiet Wasserbau und Wasserwirtschaft und Habilitation: Bergische Universität.

Gunn, K., & Stock-Williams, C. (2012). Quantifying the global wave power resource. *Renewable Energy, 44*, 296–304.

Hau, E. (2016). *Windkraftanlagen Grundlagen · Technik · Einsatz · Wirtschaftlichkeit* (6th ed.). Springer.

IAEA PRIS. (2021). In-operation & long-term shutdown reactors. Available at: https://www.iaea.org/PRIS/WorldStatistics/OperationalReactorsByType.aspx [Accessed April 3, 2021].

IEA. (2021). *Snapshot of global PV markets*. International Energy Agency.

IEA & NEA. (2015). *Projected costs of generating electricity*. 2015 Edition. International Energy Agency.

Junginger, M., & Louwen, A. (2019). *Technological learning in the transition to a low-carbon energy system: Conceptual issues, empirical findings, and use, in energy modeling*. Academic Press.

Kaltschmitt, Streicher, & Wiese (Eds.). (2007). *Renewable energy: Technology, economics and environment*. Springer.

Kaltschmitt, M., Streicher, W., & Wiese, A. (2013). *Erneuerbare Energien - Systemtechnik, Wirtschaftlichkeit, Umweltaspekte*. Springer.

Krieger, H. (2019). *Grundlagen der Strahlungsphysik und des Strahlenschutzes* (6th ed.). Springer.

Lamarsh, J., & Baratta, A. (2001). *Introduction to nuclear engineering* (3rd ed.). Prentice-Hall.

Larson, L. (2019). *The front end of the nuclear fuel cycle: Current issues*. Congressional Research Service. Available at: https://crsreports.congress.gov/product/pdf/R/R45753/1 [Accessed May 15, 2022].

Matek, B. (2013). *Geothermal energy: International market update*. Geothermal Energy Association.

Möst, D. (2006). *Zur Wettbewerbsfähigkeit der Wasserkraft in liberalisierten Elektrizitätsmärkten - Eine modellgestützte Analyse dargestellt am Beispiel des schweizerischen Energieversorgungssystems*. Peter Lang Verlag.

Murray, R., & Holbert, K. (2020). *Nuclear energy* (8th ed.). Butterworth-Heinemann.

NREL (2021). *National renewable energy laboratory – Best research-cell efficiency chart*. Available at: https://www.nrel.gov/pv/cell-efficiency.html [Accessed May 15, 2022].

Pehnt, M., & Schneider, J. (2010). Kraft-Wärme-Kopplung. In Pehnt, M. (Ed.), *Energieeffizienz* (pp. 117–146). Springer.

Pistner, C. (2012). Energie der Kerne - Physikalische Grundlagen der Kernenergienutzung. In Neles, J. & Pistner, C. (Eds.), *Kernenergie - Eine Technik für die Zukunft?* (pp. 21–39). Springer.

Quaschning, V. (2004). Technology fundamentals – Photovoltaic systems. *Renewable Energy World, 7*(1), 81–84.

Schneider, M., et al. (2020). *The world nuclear industry*, Status Report 2020, Paris. Available at: https://www.worldnuclearreport.org/IMG/pdf/wnisr2020-v2_hr.pdf [Accessed May 15, 2022].

Unger, J., Hurtado, A., & Isler, R. (2020). *Alternative Energietechnik* (6th ed.). Springer.

VGB. (2013). *Verfügbarkeit von Wärmekraftwerken 2003–2012*. VGB PowerTech e.V.

Wood, A., Wollenberg, B., & Sheble, G. (2013). *Power generation, operation, and control* (3rd ed.). Wiley.

Working Panel Waste Management. (2012). *Waste disposal for nuclear power plants*, VGB PowerTech. Available at: https://www.vgb.org/vgbmultimedia/VGB_Brosch_Entsorgung_web_ENGL_-p-6422.pdf [Accessed May 26, 2022].

World Nuclear Association. (2018). *Nuclear power reactors*. Available at: https://world-nuclear.org/information-library/nuclear-fuel-cycle/nuclear-power-reactors/nuclear-power-reactors.aspx [Accessed May 15, 2022].

Zweifel, P., Praktinjo, A., & Erdmann, G. (2017). *Energy economics. Theory and applications*. Springer.

5 Electricity Transport and Storage

Electricity can be singled out from other industrial products, including most energy products, with two properties (cf. also Sect. 2.5): (1) a specific grid is needed for the transport from producers to consumers (2) storage of electricity itself is hardly possible, and the conversion of electricity into storable energy carriers raises various challenges. Hence, this chapter addresses the following key questions:

- What principles drive the transport of electricity through the grid?
- What possibilities exist to make electric energy storable?

These questions become of significant importance when a transformation of the energy system towards the massive use of variable, site-dependent renewables is intended. This requires a replanning of the grids connecting the generation and consumption of electricity but also a reconsideration of the operation principles – based on a thorough understanding of the physical principles and the current practices. At the same time, electricity storage is gaining importance due to the variability of solar and wind infeed in time and potential mismatches with also time-variable demand. Therefore, we start by discussing electricity transport as the crucial link between generation and demand in Sect. 5.1. Storage technologies are then scrutinised in Sect. 5.2, bearing in mind that storage is a link between production and consumption when they occur at different times, just as transport creates a link in space.

Key Learning Objectives
After having gone through this chapter, you will be able to

- Describe the basic principles, the structure and the key components of electricity networks.
- Apply the key equations to compute currents, voltages and power flows in electricity networks.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_5

- Define the key concepts for system operation and the relevant ancillary services.
- Characterise the major storage technologies relevant for electricity systems.

5.1 Electricity Transmission and Distribution

Electricity is transported through power lines, i.e. metallic conductors, organised in electric networks or electricity grids. Current flowing in these networks may be either direct current (DC), i.e. current keeps (mostly) the same flow direction over time. Or, the network is operated with alternating current (AC), i.e. current changes flow directions multiple times per second. Subsequently, we start by discussing in Sect. 5.1.1 the basics of electricity networks, including why AC is predominantly used in today's electricity grids. We also line out the structure of current electricity networks. Section 5.1.2 then discusses the physical principles of power flows, while Sect. 5.1.3 provides a closer look at the components of electricity grids. Finally, system operation is addressed in Sect. 5.1.4.

5.1.1 Basics of Electricity Networks

5.1.1.1 DC Versus AC in Electricity Transmission

Today's European electricity network is operated with **alternating current** (**AC**), which goes back to the ideas of Nicola Tesla, who invented electrical motors and transformers in 1888 in the USA. However, the first electrical grid was built in 1882 by Thomas A. Edison to light his Pearl Street laboratory in Manhattan, based on **direct current** (**DC**). In hindsight, these two basic concepts (AC vs. DC) competed to be the dominant one for widespread electrification; the AC concept became the preferred transport technology, but why?

The **electric power** P transported in an electrical circuit is prima facie for DC the product of **voltage** V and **current** I.

$$P = I \cdot V. \tag{5.1}$$

The loss of power P_{loss} that occurs during that transport depends directly on the line resistance R, thus on the line thickness, the length of the line and the material used, as well as the electric current flowing I. Resistors (and also **power lines**) resist the flow of electric current. Therefore, electrical energy is needed to "transport"

Table 5.1 Electric resistivity of materials

Material	ρ [Ω mm^2/m] at 20 °C
Aluminium	2.7×10^{-2}
Carbon steel	2×10^{-1}
Copper	1.7×10^{-2}
Fresh water	2×10^{8}
Typical insulator	$10^{12 \ldots 16}$

Source Own illustration based on Helmenstine (2019)

current through the resistance. A share of electrical energy is lost by heating the resistor, in our case the power line. This **loss of power** P^{loss} can be determined by[1]

$$P^{\text{loss}} = I^2 \cdot R, \tag{5.2}$$

with R depending on the line thickness A (cross-sectional area of the conductor typically in mm^2), the material used characterised by the parameter ρ (**electric resistivity** of a material measured in Ω mm^2/m, cf. Table 5.1) and the length l of the conductor (in m):

$$R = \rho l / A. \tag{5.3}$$

Consequently, the loss of energy depends on the square of the current through the line.

It is advantageous to use high voltage and low current to transmit a certain electric power P (cf. Eq. 5.1), given that the resulting losses rise according to Eq. (5.2) with the square of the current. With transformers invented by Nicola Tesla, voltage could be transformed easily. With the transformation of voltage levels, power transport became possible with lower losses at longer distances. In consequence, AC power networks prevailed.[2]

Today's electricity systems are mainly operated at alternating current, but why with three phases? Less conductor material is necessary to transmit electrical power in a **three-phase system** compared to a single-phase or two-phase system at an equal operating voltage. A three-phase supply requires three conductors and hence just 1.5 times as many wires as a single-phase AC power line with two conductors (phase and neutral), since with three symmetric phases the neutral line has zero power flow and may thus be omitted. The three-phase system can hence

[1] The following calculations are only valid for DC lines (neglecting reactive power). However, they can also serve as rough estimation for AC power lines and can be used for a basic understanding.

[2] Contrarily to the "large-scale" power grids, on-board electrical systems in cars, trucks, ships, airplanes, etc. with short transport distances for electrical energy (just inside the vehicle) are in general DC systems at voltage levels between 12 and 48 V. Also, many of the home appliances actually use DC and most renewable installations produce direct current. Therefore, some cost savings might occur if future (distribution) grids are based on DC.

transport three times as much power as a single-phase AC power line and the capacity to conductor material ratio is two times higher. Additionally, three-phase systems can easily produce a rotating magnetic field with a specified direction and constant magnitude. Electric motors are then designed to follow this rotating field, which simplifies their design.

A new impetus was given to DC technology by the development of semiconductor-based converter technology in the 1970s. This created the possibility of converting three-phase AC power to DC power with low losses. For the integration of DC lines in AC systems, so-called converter stations are needed at both ends of the DC line to convert the three-phase current from the transmission grid into direct current and back again. The high-voltage direct current (**HVDC**) technology offers advantages in transmission over very long distances since only line losses related to active power occur. With AC lines, also reactive power (cf. Sect. 5.1.2) arises that induces additional losses. Disadvantages of HVDC systems are higher construction costs than for alternating current due to the need for converter stations and losses occurring at converter stations. Moreover, the HVDC technology is currently only available as a point-to-point connection over long distances and not as a meshed grid as is the case for AC networks (cf. also Sect. 5.1.3.1).

5.1.1.2 Structure of Current Electricity Networks

At the end of the nineteenth century, single power plants usually supplied large cities and agglomerations without interconnecting larger areas. With a continuously increasing electricity demand at the beginning of the twentieth century, consumers required a more stable operation of the electricity systems. A European wide interconnection of high-voltage alternating current (**HVAC**) lines started in the late 1930s. After World War II, today's 380 kV transmission network was established and extended throughout Europe. Transmission networks in Europe are designed to transport the electrical energy as efficient as possible considering economic factors, network safety and redundancy. A large interconnected grid offers the following benefits:

- Consolidation of load, resulting in significant equalising effects.
- Pooling of generation and equalising of renewable infeed, resulting in lower generation costs (in the total system).
- Common provisioning of reserves, resulting in lower costs for reserve power (cf. Sect. 5.1.4.2).
- Increasing market liquidity, due to a larger number of participating units (cf. Sect. 8.3).
- Mutual assistance in the event of disturbances.
- A larger power plant capacity in the meshed system, increasing frequency support and hence security of supply.

Four voltage levels can be distinguished in today's European electricity grid (cf. also Fig. 5.1): extra high voltage, high voltage, medium voltage and low voltage.

Originally, **transmission grids** consisted of high voltage levels of 110 kV (sometimes also 150 kV, e.g. in the Netherlands, Belgium and France). But they

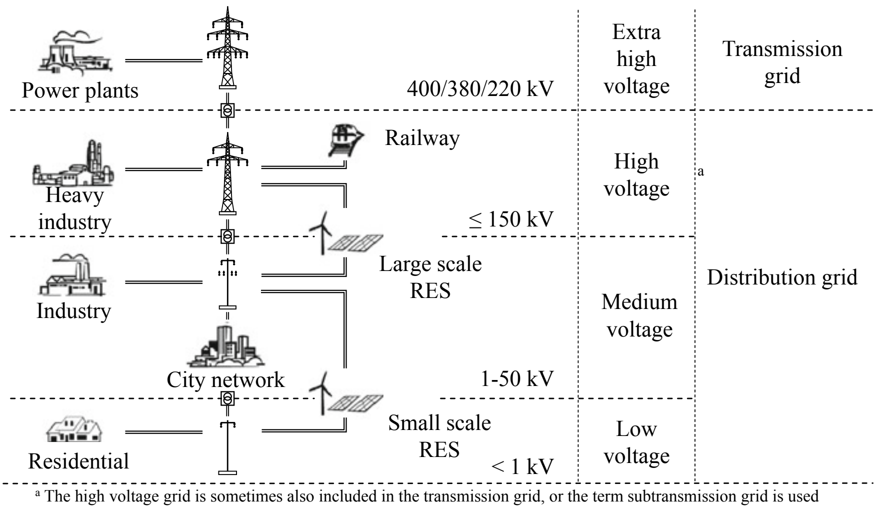


Fig. 5.1 Hierarchical structure of the European electricity system and main generators and consumers

have mostly been replaced by **extra high voltage levels** of 220/380/400 kV designed to enable higher security of supply by interconnecting larger areas. The directions of electricity flows through these grids may vary over time, depending on the different regions' demand and supply situations. Moreover, the intermeshed grid also allows for the transport of electricity over longer distances. Since the liberalisation of the electricity markets in the 1990s, trading of electricity and concomitantly the transport of electricity have significantly increased (see Chap. 10), resulting in congestions, especially at interconnectors with limited capacities between European countries.

The **high voltage level** of 110 kV/150 kV (depending on the country) is used for the transport of large quantities of electricity. These networks are usually ring feeders to address the N-1 security criterion (cf. Sect. 5.1.4.1) and have a similar function as the medium- and low-voltage distribution networks. Some 110 kV networks are also intermeshed. Furthermore, they connect heavy industry and railways that require high voltage levels.

Grids at **medium voltage level** are operated in a voltage range of 1–50 kV, with a different voltage level from country to country, often between 10 and 20 kV. Medium-voltage grids are connected to higher voltage levels and receive the energy from the high-voltage grid and/or from decentralised installed capacities (e.g. RES installations). They distribute the energy at the local level in urban and rural areas with industrial and commercial sites being often directly connected to these grids. As such, they are also called distribution grids. With the increasing share of decentralised generation, especially from renewable power plants, the direction of the energy flow is sometimes reversed in some areas.

Low voltage levels below 1 kV are used for the final distribution of electricity. Electrical energy is provided from the upstream higher voltage grids and transformed to low voltage at local substations or is received from small-scale decentralised sources, such as photovoltaics. Residential buildings and small commercial customers are connected to these grids. The voltage level in Europe is in general 230 V.

Transformers step up and down the voltage between the different voltage levels. Power in conventional power plants is provided at a relatively low voltage between about 2.3 kV and 30 kV. The voltage at the generator is then transformed to a higher voltage and injected at the extra high and high voltage levels for transmission over long distances. Medium to large-sized renewable energy power plants such as wind or solar parks feed in their generated electricity at high or medium voltage level, depending on their size. Small-scale renewable energy installations feed in at medium-voltage or low-voltage levels (cf. Fig. 5.1).

DC power lines complement the above-mentioned structure of the power grids for subsea transmissions, long-distance transports and interconnections between asynchronous grids. Submarine HVDC systems are generally used to connect the electricity grids of islands and countries separated by seas, for example, between Germany and Scandinavia, between Great Britain and mainland Europe, as well as Southern Europe and North Africa. Also, within Germany, HVDC lines are expected to transport electricity over longer distances from North to South in the future. HVDC links can thereby also be used to control problems in the grid with AC electricity flow. As mentioned in Sect. 5.1.1.1, DC lines are so far only realised as point-to-point connections and not as meshed networks like AC grids. Meshed DC power grids are currently discussed to link offshore parks in the North and East Sea. However, further research is necessary for their implementation.

5.1.2 Physical Principles of Power Flow

5.1.2.1 Fundamental Physical Laws

Several fundamental physical laws are relevant for a fundamental understanding of electricity networks and a power flow analysis. In the following, a basic overview shall be given of Kirchhoff's laws, Ohm's law and its generalisation to AC components, the representation of power lines with the help of an equivalent circuit diagram, and AC and DC power flow analysis.

Flows in electrical networks cannot be directed like in other mediums, such as e.g. water. Instead, power flows in transmission and distribution networks follow **Kirchhoff's laws**. According to Kirchhoff's laws, power flows do not only follow the direct connection between two network nodes but also parallel connections in the meshed grid. These generally unintended flows are called loop flows. Power flows within the transmission network depend on the **network topology** and the physical characteristics of transmission lines. Kirchhoff's laws are described by two equations that deal with the current and potential difference of electrical circuits: they are based on the conservation of charges (Kirchhoff's current or point law) and

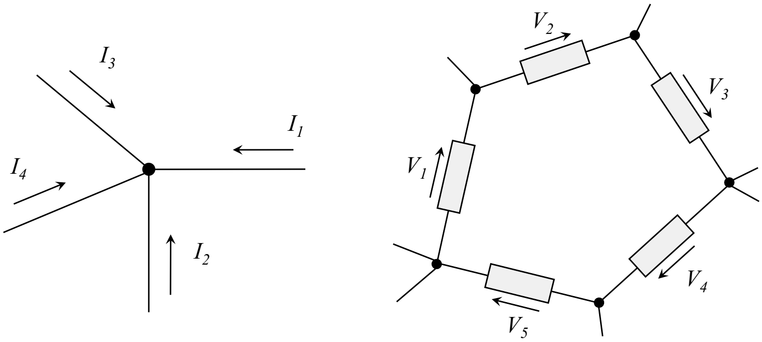


Fig. 5.2 Kirchhoff's circuit laws: two equalities that deal with the current and potential difference (voltage)

energy (Kirchhoff's voltage or mesh law) in electrical networks (cf. also Fig. 5.2). The principle of conservation of electric charges implies that at any node in an electrical circuit, the sum of currents flowing into that node is equal to the sum of currents flowing out of that node (cf. Fig. 5.2 left side). Or in other words: The algebraic sum of currents in any node in a network of conductors is zero, which is expressed by the following equation[3]:

$$\sum_{l=1}^{L} I_l = 0. \tag{5.4}$$

Kirchhoff's voltage law, sometimes also called Kirchhoff's mesh (or loop) law, implies that the directed sum of the electrical potential differences (voltage) around any closed loop in a network is zero (cf. Fig. 5.2 right side). Analogously to Kirchhoff's current law, it can be stated as

$$\sum_{l=1}^{L} V_l = 0. \tag{5.5}$$

In a DC system, the current flowing through a conductor between two points is directly proportional to the voltage difference between the two points. Introducing the resistance R, this is expressed by the following mathematical equation known as **Ohm's law**

$$V = I \cdot R. \tag{5.6}$$

[3] The formula is valid for DC networks; however, it is also valid for AC networks under some circumstances, such as no charge storage effects occurring in the nodes and lines.

Thereby I is the current through the conductor in units of amperes (abbreviated: A), V is the potential difference across the conductor which is called voltage and measured in units of volts (V) and R is the resistance of the conductor in units of Ohms (**Ω**).

In AC circuits, current and voltage are often out of phase, i.e. their sinusoidal changes are somewhat shifted in time (cf. Fig. 5.3). Such a shift is called **phase angle** φ. Assume the AC voltage over time is described by the function $V(t) = V_0 \cdot \sin(\omega \cdot t)$, with V_0 the voltage amplitude and ω the **angular velocity** (the **frequency** then is $\frac{\omega}{2\pi}$). Then the current flow is given by $I(t) = I_0 \cdot \sin(\omega \cdot t - \varphi)$.

As a result of the alternating current and the phase shifts, the laws valid for DC current are not directly applicable for AC. Kirchhoff's and Ohm's laws are dependent on the assumption that currents are stationary (i.e. DC) or that they occur only in ideal conductors. This means that whenever current flows into one end of a device, it simultaneously flows out of the other end. In other words, the devices are considered to be concentrated components. This assumption is not valid for high-frequency AC circuits, where this simplification is no longer applicable. For practical applications, the presentation of general Kirchhoff's rules has been

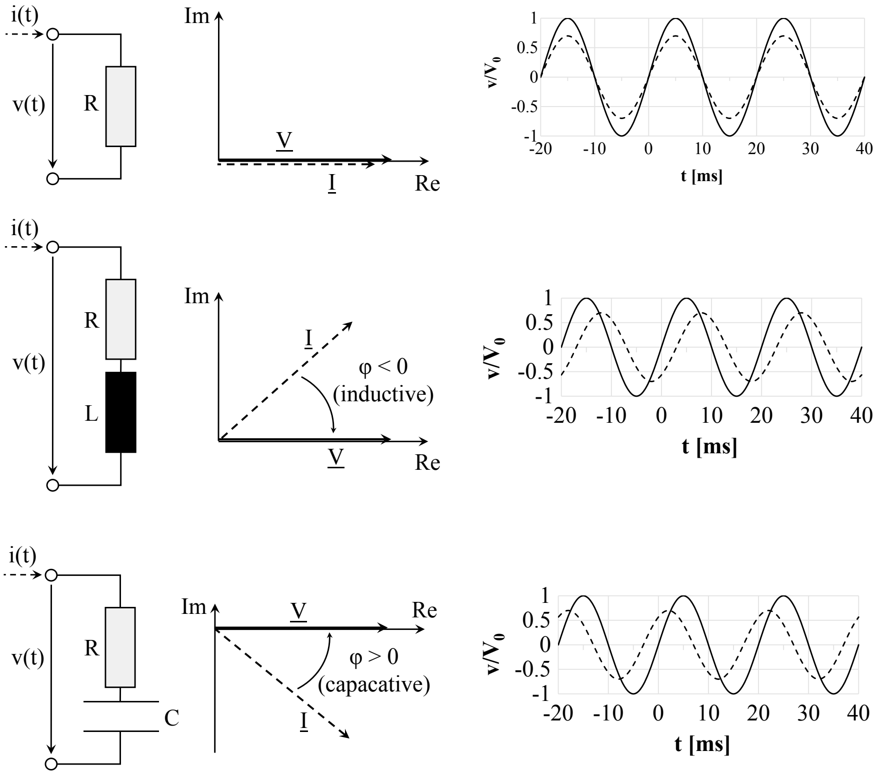


Fig. 5.3 Equivalent circuit elements including resistors, inductors and condensers, their corresponding phasors and the sinusoidal changes of voltage and current over time

variously modified and extended. The applicability of Kirchhoff's laws may often be improved by considering virtual inductances distributed along the conductors. These are treated as imaginary circuit elements that produce a voltage drop equal to the rate of change of the flux.

For AC circuits in general, a so-called **equivalent circuit** is often used to depict a given real circuit and to simplify the calculations. The **equivalent circuit** preserves all relevant electrical characteristics of a given circuit but is composed of ideal elements. Those do not exist in such perfection in reality yet facilitate the description of real behaviour and the mathematical treatment. To simplify real circuits, components are approximated by the following basic circuit elements (cf. also Fig. 5.3):

- Ohmic resistance R (ideal electrical resistance).
- Condenser C (pure capacitive reactance).
- Inductor L (pure inductive reactance).

Connections between these elements are assumed to be ideally conductive. This means they are without resistance and all components are considered time-invariant, meaning their parameters do not change with time. This allows the depiction of real-world circuits as an idealised network model within a graphical illustration, as shown in Fig. 5.3.

Ohm's law (as defined above) does not directly apply here since that form only refers to resistances (value R) and not complex impedances, which may include **capacitance** (C) or **inductance** (L).

But equations for time-invariant AC circuits correspond to Ohm's law if variables are generalised to **complex numbers** and the current and voltage waveforms are described through complex exponentials.[4] Under the assumption that voltage and current are steady sinusoids (as is the case for large-scale electricity networks during regular operation), time-independent complex variables $\underline{V}$ and $\underline{I}$ can be introduced. These complex values of voltage and current can be depicted as so-called **phasors** (phase vectors) in the complex plane (cf. Fig. 5.3). The **phase angle** φ describes here the phase difference between voltage and current.

The complex generalisation of resistance is **impedance**. This is usually symbolised as $\underline{Z}$. The basic circuit elements can then be expressed as follows:

$$\begin{aligned} \underline{Z} &= R + j0 &&= R && \text{ohmic resistance} \\ \underline{Z} &= 0 + j\omega L &&= \omega L \cdot e^{j\frac{\pi}{2}} && \text{inductive reactance} \\ \underline{Z} &= 0 - j\frac{1}{\omega C} &&= \frac{1}{\omega C} \cdot e^{-j\frac{\pi}{2}} && \text{capacitive reactance.} \end{aligned} \tag{5.7}$$

[4] A key property of complex numbers is the equality: $e^{j\varphi} = \cos\varphi + j\sin\varphi$. Note: when inserting $\varphi = \pi$, Euler's identity is obtained: $e^{j\pi} = -1$. Given that the imaginary unit j (denoted i outside the electrical engineering community) is defined through the equation $j^2 = -1$, this is also consistent with $e^{j\frac{\pi}{2}} = j$.

Now **Ohm's law** can be written in the more general form

$$\underline{V} = \underline{I} \cdot \underline{Z}, \tag{5.8}$$

where $\underline{V}$ and $\underline{I}$ are the complex scalars for voltage and current, respectively, and $\underline{Z}$ is the complex impedance. The real part of $\underline{Z}$

$$\mathrm{Re}\{\underline{Z}\} = R \tag{5.9}$$

corresponds to the Ohmic, active or effective resistance. The imaginary part of Z

$$\mathrm{Im}\{\underline{Z}\} = X \tag{5.10}$$

corresponds to the **reactance**. When $\underline{Z}$ is complex, only the real part is responsible for dissipating heat. The reciprocal value of the impedance $\underline{Z}$ (complex resistance) is the **admittance** $\underline{Y}$ (complex susceptance)

$$\underline{Y} = \frac{1}{\underline{Z}} = G + jB. \tag{5.11}$$

The real part of the admittance is the **conductance** G. The imaginary part of the admittance is called the susceptive part of admittance or **susceptance** B.

Also, the equation for power flows (Eq. 5.1) may be generalised to complex numbers for AC circuits:

$$\underline{S} = \underline{I}^* \cdot \underline{V}. \tag{5.12}$$

Thereby, $\underline{S}$ is the so-called apparent power, and $\underline{I}^*$ is the complex conjugate to $\underline{I}$, i.e. the complex number with same real but negative imaginary part: $\underline{I}^* = \mathrm{Re}\{\underline{I}\} - j\mathrm{Im}\{\underline{I}\}$.

For a given circuit element, the complex voltage may be replaced according to the generalised Ohm's law (Eq. 5.8). With the decomposition of the complex impedance $\underline{Z} = R + jX$ (cf. Eqs. 5.9 and 5.10), a decomposition of the apparent power for a network element in active power and reactive power is obtained. This **power triangle** is shown in Fig. 5.4, thereby the symbols without underscore ("_") stand for the magnitude of the corresponding AC quantities.[5]

In case of a generator, the generator output (**apparent power** S) in volt-amperes (VA), may decomposed in the usable power (**active power** P) in watts, and the wasted or stored power (**reactive power** Q) in volt-amperes-reactive (var). Both power components have to be transported through the grid, resulting in an active (or real) and a reactive power flow. Over time, the reactive power is yet transported back and forth inside the circuit, so it is not usable by consumers. Depending on the

[5] Note that for AC voltage and current, the magnitude ("effective value") is related to the amplitude of the corresponding sinus wave (depicted in Fig. 5.3) through a factor $1/\sqrt{2}$. E.g. for voltage: $V = V_0/\sqrt{2}$.

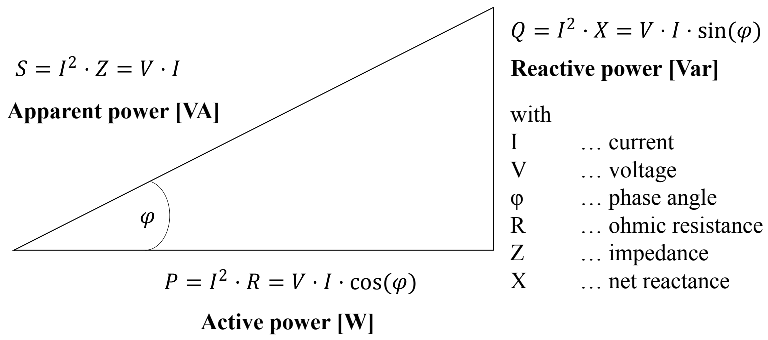


Fig. 5.4 Power triangle represented in complex space

phase angle, a reactive power flow is either (predominantly) induced by a condenser (negative phase angle) or an inductor (positive phase angle). Applied to a transmission line, the active power obtained from the power triangle corresponds to the power loss as indicated in Eq. (5.2). The ratio of the active power to the apparent power $P/S = \cos(\varphi)$ is called **power factor**.

5.1.2.2 Stationary Power Flow Computations for Symmetric Three-Phase Systems

Power flow or load flow analyses are essential for monitoring the electricity grid, the determination of the current state and as a basis for reliability analyses of the electricity network. A load flow or **power flow analysis** is a numerical analysis of the flow of electric power in an interconnected system. A power flow computation usually makes use of simplifications, such as a one-circuit representation and the so-called per-unit system (cf. below). It focuses on key parameters of AC power systems, such as voltages, voltage angles, real power and reactive power, including also losses in lines and transformers. In power flow analysis, the power systems are analysed in regular steady-state operation.

Depending on the objective or planning purpose, different levels of detail can be distinguished: for short-term planning, monitoring and as a basis for reliability analyses of electricity networks, a high level of detail is essential, including voltages, voltage angles, real power and reactive power. In this application, generator injections and loads are generally given at the different nodes in the electricity network. For long-term planning, such as planning future expansion of power systems or analysing economic effects, a higher aggregation with some approximations is adequate, resulting, e.g. in a so-called DC power flow approach with a reduced computation time. With the reduced calculation time, a combination with dispatch models is possible. This allows for optimising generation and load levels at the different nodes of the electricity network and calculating power flows. This is typically relevant for analysing economic aspects of electricity networks, including market design issues, such as market splitting or nodal pricing (see Sect. 10.6.1).

In the following, the relevant basic principles and assumptions for AC power flow modelling are described. The model is simplified to an approximated DC power flow model in the subsequent chapters, enabling the combination with optimal dispatch models, to obtain so-called optimal power flow models (cf. Sect. 7.1.3).

Today's (worldwide) power systems are generally operated as three-phase systems (cf. Sect. 5.1.1.1) consisting of three alternating current circuits, with a shift in the sinusoidal oscillations corresponding to a phase angle of 120° (i.e. 360°/3). In conventional thermal and hydro power plants (see Sects. 4.1 and 4.2.1), the three phases are produced through three separate windings on one generator shaft, turned by 120°. The generator **frequency** is typically 50 Hz (e.g. in Europe) or 60 Hz (in the US). Three-phase systems may also be operated with a fourth conductor, the so-called neutral wire, especially in low-voltage distribution. A neutral conductor enables the supply of three separate single phases at a constant voltage. Furthermore, a three-phase system is usually more cost efficient than an equivalent single-phase or two-phase system at a given line-to-ground voltage as less conductor material is necessary to transmit electrical power (cf. Sect. 5.1.1.1).

Symmetry in the three-phase system: for the power grid in regular operation, it can be assumed that the three-phase system is built (e.g. regarding position of lines) and operated symmetrically. If symmetry prevails, three-phase systems can be represented by a one-phase equivalent circuit, which is the basis for power flow calculations resulting in less computing time.

To calculate **AC power flows** on a power line with a line length up to 250 km, the so-called π-equivalent model (for details cf. Bergen and Vittal 2000 and Schwab 2015) is used under the assumption of a symmetric and stationary state. Longer power lines are sectioned into a series of π-equivalents within the calculation. The π-equivalent model (cf. Fig. 5.5) of a power line with concentrated elements can be used as a starting point for the derivation of the power flow equations. The interested reader is referred to Schwab (2015) and Schweppe et al. (1988), who describe the way from the electrotechnical fundamentals to the calculation of the complex apparent power S, the **active** (or real) **power** P and the **reactive power** Q as given in Eq. (5.13).

$$\begin{aligned}
\underline{S}_{nm} &= g_{nm}V_n^2 - g_{nm}V_nV_m\cos\theta_{nm} - b_{nm}V_nV_m\sin\theta_{nm} \\
&\quad + j\left(b_{nm}V_nV_m\cos\theta_{nm} - g_{nm}V_nV_m\sin\theta_{nm} - V_n^2\left(b_{nm}^q + b_{nm}\right)\right) \\
P_{nm} &= g_{nm}V_n^2 - g_{nm}V_nV_m\cos\theta_{nm} - b_{nm}V_nV_m\sin\theta_{nm} \\
Q_{nm} &= b_{nm}V_nV_m\cos\theta_{nm} - g_{nm}V_nV_m\sin\theta_{nm} - V_n^2\left(b_{nm}^q + b_{nm}\right).
\end{aligned} \tag{5.13}$$

The line characteristics comprise the conductance g_{nm}, the inductive susceptance b_{nm} and the (halved) capacitive susceptance $b^q{}_{nm}$ of the line as well as the thermal line limitation describing the technical transfer capability of the line from

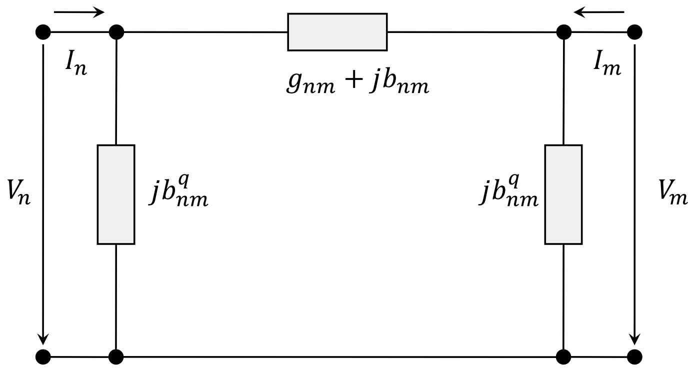


Fig. 5.5 π-equivalent of a long power line

n to m. Conductance and susceptances can be calculated using the resistance R_{nm}, the reactance X_{nm}, the shunt[6] capacitance C^q_{nm} and the angular velocity ω (cf. Eq. 5.8):

$$g_{nm} = \frac{R_{nm}}{R^2_{nm} + X^2_{nm}}; \quad b_{nm} = \frac{-X_{nm}}{R^2_{nm} + X^2_{nm}}; \quad b^q_{nm} = \frac{\omega C^q_{nm}}{2}. \tag{5.14}$$

The **power flow** within a **meshed system** is generally determined in 4 steps. In the first step, the electricity network is described with equivalent circuit elements. In the second step, a system of equations is set up, which is then solved in the third step. In the final step, it is analysed whether the complex power flow $\underline{S}$ violates the thermal line limitations and whether the voltage V remains within prespecified voltage boundaries.

In the power flow problem, three types of nodes (or buses)[7] are distinguished:

1. Generation nodes labelled as PV-nodes: For generator nodes, it is assumed that the active power generated ${P^G}_n$ and the voltage magnitude V_n are known. Voltage phase and reactive power are the output quantities and are determined within the calculation.
2. Load nodes labelled as PQ-nodes: For load nodes, it is assumed that the real power ${P^D}_n$ and reactive power ${Q^D}_n$ at each load node are known. Output values of the calculation are voltage value and voltage phase.
3. Slack node: For the slack node, it is assumed that the voltage magnitude V_n and voltage angle θ_n are known. Active power and reactive power are determined so that there is a balance in the whole electricity network between power injection

[6] In electrical engineering, the term shunt is used in general to designate an alternative current path in a circuit—similarly like shunts designate bypasses in surgery. The capacitance here is not an actual electric component but part of the equivalent circuit representation of the electrical line.

[7] The term "bus" is employed here in reminiscence of the busbars which are thick conductors in transformer substations connecting different lines.

and load, including network losses. In general, the slack node should be a high-powered generation node, which can be used to balance the system. Often the largest power plant is chosen as the slack node.

Both the voltage magnitude and voltage phase are unknown for each load node and must be solved for. In the case of generator nodes, the voltage phase and reactive power are unknown. Active and reactive power must be solved for the slack node. In a system with n nodes, there are hence $2n$ unknown variables, which have to be determined within a system of equations. There are two power balance equations for each node of the electricity network. The real power equation is:

$$P_n = V_n \sum_{m=1}^{N} V_n (G_{nm} \cos \theta_{nm} + B_{nm} \sin \theta_{nm}), \tag{5.15}$$

with P_n being the injected power at node n, G_{nm} being the real part of the element in the node admittance matrix (cf. below) corresponding to the nth row and mth column, B_{nm} being the imaginary part of the element in the node admittance matrix corresponding to the nth row and mth column and θ_{nm} being the difference in voltage angle between the nth and mth node. The reactive power balance with Q_n as injected net reactive power is given by:

$$Q_n = V_n \sum_{m=1}^{N} V_n (G_{nm} \sin \theta_{nm} + B_{nm} \cos \theta_{nm}). \tag{5.16}$$

These equations are applied to all the nodes in a power flow system and thus allow for determining the voltages, currents and power flows in the system based on the parameters contained in the node admittance matrix or Y matrix, which summarises the parameters G_{nm} and B_{nm}.

The node **admittance matrix** or Y matrix is an $N \times N$ matrix of complex numbers describing a meshed grid with N nodes. The admittance matrix is constructed for the type of power flow analyses envisaged here based on the single line diagram of a power system, a simplified notation for representing a three-phase power system. As indicated in Sect. 5.1.2.1, the admittance $\underline{Y}$ is the reciprocal of the impedance $\underline{Z}$. The admittance matrix $\underline{\boldsymbol{Y}}$ consists of the real part $\mathrm{Re}\{\underline{\boldsymbol{Y}}\}$ corresponding to conductance $\boldsymbol{G}$, and the imaginary part $\mathrm{Im}\{\underline{\boldsymbol{Y}}\}$called susceptance $\boldsymbol{B}$.

$$\underline{\boldsymbol{Y}} = \boldsymbol{G} + j\boldsymbol{B} \tag{5.17}$$

For a power system with N buses, the admittance between the bus under consideration n and another bus m, connected to n, can be written as $\underline{Y}_{nm} = G_{nm} + jB_{nm}$.

In the node admittance matrix used for power flow calculations, the off-diagonal elements $\underline{Y}_{nm,n \neq m}$ equal to the negative admittances of the corresponding branch connecting buses n and m (cf. Eq. 5.14): $\underline{Y}_{nm,n \neq m} = -(g_{nm} + jb_{nm})$. The diagonal elements $\underline{Y}_{nm}$ are computed as sum of the admittances of all branches connected to bus i plus the susceptances of the shunts close to the node in the π-equivalent representation of the line: $\underline{Y}_{nn} = \sum_{n,n \neq m} \left(g_{nm} + jb_{nm} + jb^{q}_{nm} \right)$. In realistic systems, the Y matrix is relatively sparse. This is because each node in a real power system is usually connected to only a few other nodes through the transmission lines (and not to all other nodes).

The resulting power flow equations are nonlinear. In general, the Newton-Raphson method[8] is used today to solve the equation system. When power injections and withdrawals are known, calculation time is quite fast also for large-scale problems. Yet, the power flow equations are both non-convex and non-concave, so that convergence is not guaranteed. Furthermore, optimal power flow models (cf. Sect. 7.1.3) using these equations and including several (or plenty of) time steps may be rather time consuming for large electricity networks. For applications with a longer time horizon, such as the principle design of electricity networks and their economic analysis, some simplifications can reduce calculation time.

5.1.2.3 Linear Approximations for Stationary Power Flow

In AC power flow analyses, generator injections and loads are in general given. For long-term planning, such as planning future expansion of power systems or analysis of economic effects, a higher aggregation with some approximations may be adequate, as a reduction in computation time is necessary. Especially when several time steps and scenarios (with different infeed and loads at the nodes of the electricity network) have to be considered, some simplifications are needed. Of course, the simplifications reduce the accuracy of the power flow calculation. Nevertheless, the deviation of the solution is very small and often negligible. Hence, several authors (cf. e.g. Schweppe et al. 1988) and (Overbye et al. 2004) conclude that a so-called DC power flow approximation is adequate for the long-term analysis of the design of electricity networks as well as for economic analyses including the derivation of nodal prices (cf. Sect. 10.6.1).

Furthermore, **DC power flow** models may relatively easily be integrated with an optimal power plant dispatch model (cf. Sect. 4.4). Instead of using given generation and load levels at the different nodes, the optimised dispatch and corresponding power flows can then be determined in the model (cf. Sect. 7.1.3).[9] For single lines in a meshed system, experiences with large-scale systems show that the deviation of power flows between AC and DC power flow calculations is on

[8] The Newton–Raphson method is an algorithm which approximates the zeroes of a real-valued function with the help of the tangent line (making use of the first derivative). The intercept of the tangent line is used to calculate a new approximation of the zero. The new approximation is again used with the corresponding tangent line of the function to come again with a new tangent line closer to the zero. This iteration is iterated until a termination criterion is reached. In comparison to classical interval nesting for finding the zeroes of a function, the Newton–Raphson methods needs in general less iterations.

[9] The power plant dispatch can also be determined in AC power flow models; however, in general, a heuristic approach is necessary to determine the optimal solution.

average significantly lower than 1%. For lines with a high share of reactive power and low real power flows, deviations can be considerably higher. Hence, depending on the objective of the electricity network analysis, the appropriate power flow calculation has to be chosen.

DC models provide a linear approximation of the power flows in AC models that is particularly valid for extra-high-voltage lines, i.e. the transmission system. The following assumptions are thereby made to reduce calculation time:

1. Voltage angle differences θ_{nm} are rather small, so that $\cos\theta_{nm} \approx 1$ and $\sin\theta_{nm} \approx \theta_{nm}$. Changes in voltage angle then have little effect on reactive flows but a significant effect on active flows.
2. Voltages at nodes are assumed to be equal at each node (for simplification, voltage magnitudes V are often standardised to per unit (p.u.) calculations, i.e. they are expressed relative to the reference voltage level. In DC models then, the voltages are all equal to 1).
3. Finally, it is assumed that line losses can be ignored (as $X \gg R$) and hence $P_{nm} = -P_{nm}$. Especially, this assumption is less appropriate for lower voltage grids.

As a consequence of these assumptions, the reactive power flows Q_{nm} are much smaller than the active power flows and the active power flows simplify to:

$$P_{nm} = V^2 b_{nm} \theta_{nm}. \tag{5.18}$$

Plugging this into the nodal power balance equation corresponding to Kirchhoff's current law (cf. Eq. 5.4) respectively the AC formulation of Eq. (5.15), one obtains:

$$P_n = V^2 \sum_{m=1}^{N} b_{nm}(\theta_n - \theta_m). \tag{5.19}$$

The main advantage of these simplifications is the high speed in solving the problem. This allows this approach to be used for large electricity networks, including determining an optimised power plant dispatch like in large energy system models (cf. Sect. 7.1). Also, the formulation of models for determining optimal capacity extensions for generation and grid capacities is possible with this approach. The reader should be aware that DC models approximate the actual power flow in electricity networks and deliver a good starting value for AC power flow calculations.

Power losses can be included in the DC approximated model to overcome some of the drawbacks of DC approximated models and to provide a more accurate representation. This can be achieved by using the second-order Taylor series expansion of the cosine function to approximate the losses of a line:

$$\cos \theta_{nm} = 1 - \frac{{\theta_{nm}}^2}{2} \tag{5.20}$$

Taking into account that the resistance is much smaller than the reactance,[10] the Ohmic loss P_{nm}^{loss} can be approximated by (for further details cf. Todem 2004):

$$P_{nm}^{\text{loss}} = \frac{1}{V^2} R_{nm} (P_{nm})^2 \tag{5.21}$$

This approximation can be applied for high-voltage power grids. However, in distribution grids with lower voltage, losses are significantly underestimated.

Based on a DC power flow model, so-called power transfer distribution factors (PDTF) can be determined (cf. Sects. 7.3 and 10.6.1).[11] Power transfer distribution factors show the linearised impact of a transfer of power and describe the (incremental) distribution of power transfers between two regions on the connecting lines. Thereby areas, zones or single nodes may be considered as regions. The PTDF values provide a linearised approximation of how the flows on the transmission lines and interfaces change in response to a change in injection at one node and a corresponding withdrawal at another node. PTDFs depend on the **topology** of the electric power system and the characteristics of the transmission lines. According to Baldick (2003), PTDFs change when an outage of a line occurs, if a controllable element reaches its control limits, and also as the pattern of injections and withdrawals change the loadings on the lines in the system. Referring to Baldick (2003), the PTDFs are relatively insensitive to the levels of injections and withdrawals for a fixed topology. This allows the usage of PTDFs as a simplified method to represent electricity flows in electricity network analysis. Furthermore, PTDFs allow the grouping of different nodes together, resulting in a reduced PTDF matrix.

5.1.3 Electricity Network Components

In general, power lines, transformers, switches, protections and further elements can be distinguished as equipment of electricity networks. As power lines and flexible AC components (FACTS) are of significant importance for the regular operation of future grids with high shares of renewables, the remainder of this section focuses on these elements.

[10] For a 380-kV-overhead line, the resistance R is approximately 0.04 Ω/km and the reactance approximately 0.4 Ω/km.

[11] AC models could also be used to determine PDTFs. In that case, a reference power flow pattern is selected first, and then, the AC power flow equations are linearised around that reference point using a Taylor series expansion. This leads to an improved accuracy of the linear approximation compared to the DC approximation.

5.1.3.1 Power Lines

The main transmission technologies will be introduced in the following sections. This includes high-voltage alternating current (HVAC) and high-voltage direct current (HVDC) systems, while other technologies such as gas-insulated lines[12] and superconductive lines are not addressed. The interested reader is referred to Schwab (2015) and Oeding and Oswald (2011) for a detailed technical overview.

High-Voltage Alternating Current (HVAC)

Transmission flows through **HVAC lines** depend on the physical laws of Kirchhoff. Thus, HVAC line flows cannot be directly controlled as they are subject to line impedances. The limited controllability may also result in so-called loop flows in AC grids.

Transmission lines are typically not operated at their full **thermal capacity limit**.[13] In case of a transmission line failure in the network, other lines need to carry the additional load to avoid a system outage. Hence, an electricity system is generally operated in a way as to withstand the outage of any single technical equipment. This operation guideline is called the N-1 criterion (cf. also Sect. 5.1.4.1). In other words, the N-1 criterion means the rule according to which elements remaining in operation within TSO's responsibility area after a contingency must be capable of accommodating the new operational situation without violating operational security limits, cf. also Network Code on System Operation (EC 2017a, b). Additionally, the transport of active power via a line is limited by the maximum physical transmission capacity, or in other words, the thermal capacity limit. The reaching of the thermal capacity limit is mainly dependent on the thermal losses induced by currents from active and reactive power. In consequence, not only active power contributes to reaching the thermal limit. The need for reactive power is varying depending on the workload of the line. In general, inductive reactive power and Ohmic losses increase quadratically with line load. An exemplary illustration for a power line is given in Fig. 5.6.

The high transmission losses over longer distances due to Ohmic resistance are a major disadvantage of HVAC lines (see Sect. 5.1.1.1). In general, overhead lines differ from underground cables as different materials (with different conductance) are used. A regular HVAC overhead line faces losses of up to 15% over a distance of 1000 km.

As mentioned, the capacity of a power line depends on the thermal limits of the line. The transmission capacity of overhead lines is restricted by a maximum current, which is determined such that compliance with safety regulations is

[12] Gas-insulated lines (GIL) allow higher transmission voltages and power ratings. The conductor core is placed in an isolating gas within a metal tube. GILs are applied in switching stations and in urban areas as well as in areas where overhead lines are not usable due to spatial or optical reasons.

[13] The capacity of a power line is usually restricted by the thermal capacity limit. A higher temperature of the cable (resulting from thermal losses of power transport in the line) implies an expansion of the conductor cable and thus leads to greater sagging of the cable. Correspondingly, wind conditions and outside temperature have an effect on the possible transport volume of a power line.

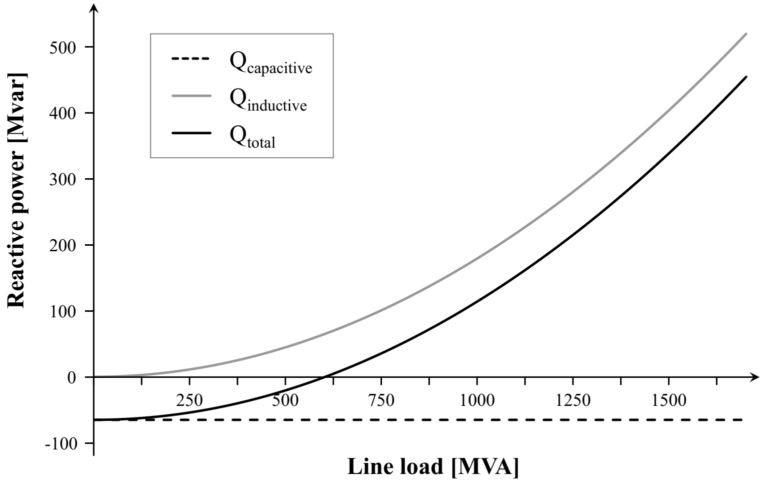


Fig. 5.6 Reactive power behaviour of a 380 kV line (100 km)

guaranteed under all circumstances. Given the material and wire diameter of the conductor, the transmission capacity is limited by the conductor temperature. Generally, an operating temperature greater than 80 °C has to be avoided as the thermal expansion of the overhead line would be too large. A flashover to soil or vegetation could occur. In addition to the dissipated heat induced by the resistance to the electric current, the line is cooled (very rarely also warmed) by the surrounding air – depending on the weather conditions. The main factors of cooling are ambient temperature, wind speed, solar radiation and precipitation. Based on worst-case assumptions, the maximum power line limit is determined according to the European norm EN 50182. When a so-called **dynamic line rating** is implemented, the static worst-case-based limitations do not need to hold for all days of the year. Instead, temperature, wind speed, solar radiation and precipitation, and the temperature of the line are monitored in real-time. In favourable weather conditions, e.g. with strong wind or low temperature, the maximum limit of the overhead line can be increased compared to a normal situation. Dependent on the weather conditions, the thermal limit can be modified based on the active monitoring of the overhead line. This may enable a capacity increase on power lines by up to 50% compared to conventional operation.

HVAC underground cables: HVAC underground cables have a lower Ohmic resistance in comparison to overhead lines due to their larger cable cross section and the usage of copper as a conductor material. This may induce up to 50% less transmission losses for HVAC underground cable systems in comparison to HVAC overhead lines and results in reduced operation costs (cf. ICF consulting 2003). Yet underground cables face higher capacitive currents reducing the transmission capacity of the line. In general, these currents need to be compensated to keep up the transmission capacity when reaching a critical length. Compensation stations have to be installed approximately every 20 km (cf. ICF consulting 2003), which is an economic disadvantage compared to overhead lines. In terms of reliability, there is no clear evidence, and studies show contrasting results, mainly because long-term experience for HVAC underground cables is still lacking. While in some studies, higher failures are documented for underground cables, other studies show lower failures.

High-Voltage Direct Current (HVDC)

In contrast to three-phase HVAC lines, high-voltage direct current (**HVDC**) **lines** are point-to-point connections. HVDC lines have the advantage that they do not require reactive power and only active power is transmitted. Ohmic losses of HVDC lines are 25–35% lower compared to equivalent HVAC lines. At both ends of an HVDC line, converter stations are required to connect to the HVAC grid and conversion losses occur of up to 1.7% per station. As costs around 135–250 million Euro arise per converter station, the HVDC technology is getting profitable in comparison to HVAC lines for distances beyond around 600 km (cf. Fig. 5.7). The drawbacks of higher converter station costs and conversion losses have to be outweighed by the benefits of lower line losses. Two main HVDC technology concepts can be distinguished:

1. **Line-commutated converters** (LCCs) or current source converters use thyristors. Thyristor valves have become important for high-voltage direct current (HVDC) as thyristors can switch power on the scale of megawatts. Most HVDC systems in operation use line-commutated converters. The converter behaves approximately as a current source on the AC side, injecting both grid-frequency and harmonic currents into the AC network and hence is considered a current source inverter.
2. **Self-commutated converters** or **voltage source converters** (**VSC**) are operated with isolated gate bipolar transistors (IGBT), which can be turned on and off (in

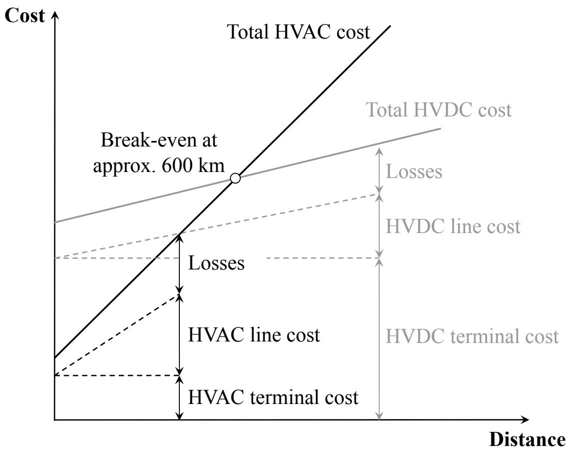


Fig. 5.7 Simplified illustration of the costs of a HVAC and a HVDC overhead line dependent on the distance. *Source* Own illustration based on ABB (2019)

contrast to line-commutated converters). This additional controllability gives the advantage that the IGBTs can be switched on and off many times per cycle to improve the harmonic performance. As the converter no longer relies on the AC system for operation, it is self-commutated. A voltage source converter can start without AC grid, or in other words, is able to support a black start.

HVDC lines allow for full power flow control carrying high power of more than 1000 MW. The technology is often used for interconnecting different synchronous systems, e.g. sea cables between Scandinavia and Continental Europe are HVDC lines. HVDC technology is preferred for sea cables at lower distances as HVAC compensation technologies are impractical at sea. In addition, in case of a failure in neighbouring grids, the HVDC line prevents cascading to the domestic one. Point-to-point connections of HVDC lines are today state-of-the-art, while meshed HVDC systems are still under research. The development of circuit breakers, so-called switchgears, is currently ongoing, enabling the operation of meshed HVDC grids.

Today, HVDC lines are used for interconnecting countries, especially using sea cables, and connecting offshore wind farms. Further applications are for long-distance transports of electricity, e.g. in countries like China. The German Network Development Plans also foresee the installation of HVDC lines from the north to the south of Germany.

5.1.3.2 Flexible AC Transmission Components

A **flexible alternating current transmission system** (**FACTS**) is a system composed of non-rotating equipment used for the AC transmission of electrical energy and is generally a power electronics-based system. These are essential components of smart grids which improve the controllability and enhance the power transfer capability of the network. The primary purpose of FACTS is to supply the electricity network as quickly as possible with inductive or capacitive reactive power to improve transmission quality and efficiency. With higher controllability of power flows, further (cost-intensive) extensions of power systems can be avoided. FACTS provide a better adaption to different grid conditions and improve the usage of the existing grid infrastructure.

In general, different types of parallel and series-connected FACTS can be distinguished:

1. Parallel-connected FACTS: Thyristor-controlled reactor (TCR), thyristor-switched capacitor (TSC), static VAr compensator (SVC) and static synchronous compensator (STATCOM). Parallel-connected FACTS provide inductive or capacitive reactive power in a node and are used for voltage regulation in undisturbed operation.
2. Series-connected FACTS: Thyristor-controlled series capacitator (TCSC), thyristor-switched service capacitator (TSSC), thyristor-controlled series reactor (TCSR), thyristor-switched series reactor (TSSR) and static synchronous series compensator. Series-connected FACTS are used to control power flows of

single lines, e.g. the split of power flows on different lines. Additionally, they are used to increase/decrease the voltage difference along the line.

5.1.4 System Operation

5.1.4.1 System Security

The availability of electricity is an essential requirement for a well-functioning economy. Consequently, power systems are designed to provide a high level of security of supply. This comes with the challenge of ensuring a permanent equilibrium of electricity supply and demand as electricity cannot be stored in large amounts. Not adequately maintaining this equilibrium can result in supply interruptions, which negatively affect firms and households.

Before deregulation, integrated energy utilities were responsible for a stable and secure supply of electricity. With deregulation, generation, transmission, distribution and sales have been unbundled and especially transmission system operators (TSO) as well as distribution system operators (DSO) are now mainly responsible for a stable and **secure supply** of electricity (cf. Sect. 6.1).

To cope with system security and to ensure the ability of a system to withstand disturbances, different concepts are relevant:

1. **N-1 Security**
2. **Reliability**
3. **Availability**.

While N-1 security is mainly an operational principle that may be implemented in the day-to-day operation and longer-term planning, reliability is a descriptive concept that may be measured through various indicators at the system level. **Availability** finally is instead an indicator at plant or component level and as such has already been introduced in Sect. 4.3.1. There it has been considered for generation plants, yet it may be conceptualised and measured similarly for grid components like HVDC cables or transformers.

N-1 Security

Securing the system against all possible outage events (contingencies) is clearly impossible, yet on the other hand, system-wide outages are very costly. Therefore, the fundamental operation rule for electricity networks calls for securing against all foreseeable disturbances. But what is foreseeable?[14] In current operation concepts, it is typically assumed that the probability of two or more independent faults or

[14] The disaster at the Fukushima Daiichi nuclear power plant in Japan was not the kind of "unforeseen" event. Available evidence indicates that a tsunami like the one in March 2011 could happen once every 1000 years or less. No precautions had yet been taken to protect the plant against such an event. The ex ante appraisal of events is hence not always aligned with the objective information available and perceptions may be biased also in view of the necessary investment for having sufficient redundancy.

failures simultaneously taking place is too low to be considered relevant. When the system withstands a failure of any single component, the system is called "N-1 secure". Any single failure of one of its *N* components could be compensated by the system and the operation can be continued. Likewise, in a system that is "N-2 secure" (or "N-3 secure"), the system can withstand a simultaneous failure of two (three) components. Consequently, N-1 security means that the system is redundantly operated, which helps to withstand the disturbance. If the N-1 criterion is violated, an N-1 secure system status has to be reached again in the shortest time with the help of switching operations in the electricity grid. To allow N-1 secure operation, conventional power systems are designed such that the N-1 principle is fulfilled for the maximum grid load. If the grid load is lower, security of supply is even higher, as the electricity network is not necessarily in an insecure system state after a disturbance and may tolerate further disturbances.

But, how is the N-1 criterion checked in today's electricity network? Transmission system operators are obliged to ensure the N-1 security of the electricity transmission system. During operational planning, the TSOs have to predict the physical flows of the electricity network, check if any overloadings or voltage violations ("**congestions**") occur and manage them if there are any. This is notably done one day ahead of actual operation "today for tomorrow". Before the year 2000, the localisation of generation was very stable and predictable most of the time. However, this has changed with the deregulation of electricity markets and the integration of large amounts of variable renewable generation. Hence, accurate congestion detection requires carrying out power flow forecasts that demand cooperation among neighbouring TSOs in Europe, including notably a standardised data exchange. This cooperation is now organised through Regional Security Coordinators (RSCs). After all the data concerning power flows between TSO regions have been exchanged, each TSO, with the support of the corresponding Regional Security Coordinator, can construct a power flow model that represents the most probable state for the 24 h of the next day. One day ahead, with the so-called **Day-Ahead Congestion Forecast** (**DACF**), a security analysis is performed that simulates the failure of any line or generator in the considered system. This analysis is processed for 24 timestamps and the different TSO areas. The aim here is to generate a comprehensive 24 h overview of the security risks on the grid for the following day. Hence power flow computations for all relevant N-1 cases are performed based on the power flow equations discussed in Sect. 5.1.2.2. The last step involves analysing the constraints detected for the following day and identifying remedial actions to solve them. At the latest after closing of the intraday markets, e.g. 15 min before "real-time", (but also earlier if congestions occur), the TSO as system responsible party can call for **redispatch**, if no other measures (especially grid topology changes) can solve the congestion. Redispatch requests issued by the TSO require power plants to adjust their real power to avoid or eliminate congestion (cf. Sect. 10.6.2). Hence, it is a measure overriding the power plant schedule resulting from the final market result.

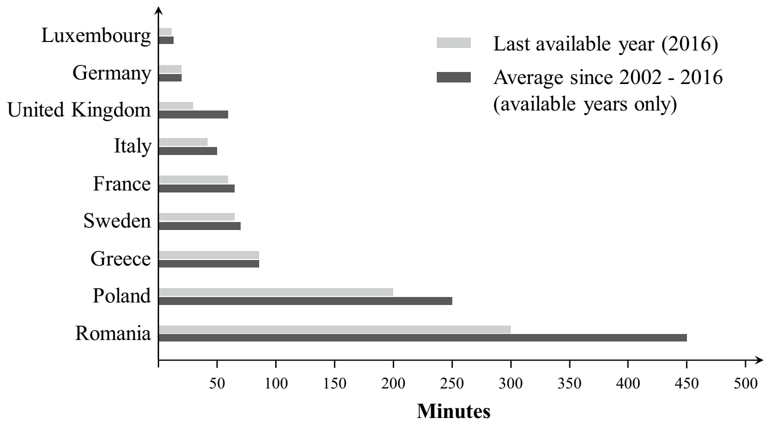


Fig. 5.8 Comparison of the System Average Interruption Duration Index (SAIDI) of European countries. *Source* Own illustration based on data from CEER (2018)

Reliability

Reliability as an indicator for security of supply is quantified based on the duration of supply interruptions. Reliability is measured with the help of the **System Average Interruption Duration Index** (**SAIDI**). SAIDI is calculated as the weighted sum of all supply disruption within one year multiplied with the number of concerned customers and put in relation to the total number of customers (in one area/one country). The SAIDI is compared for different European countries in Fig. 5.8.

Additionally, the **System Average Interruption Frequency Index** (**SAIFI**) is commonly used as a further reliability indicator by electric power utilities. SAIFI is the average number of interruptions that a customer would experience and is calculated based on the total number of customer interruptions divided by the total number of customers served. In Germany, the SAIFI is below 1, meaning that there is less than one interruption on average per customer and year.

Besides SAIDI, the reliability of whole network areas is quantified with the help of the so-called **Average System Interruption Duration Index** (ASIDI), which is a sum of weighted interruptions of supply at transformer stations in the distribution grid. For Germany, the ASIDI value for unplanned outages in the medium-voltage grid is about 10 min per year. These outages correspond to about 80% of all unplanned outages.

A further indicator, which is especially useful for analysing how different customers are concerned by a possible outage, is the so-called **Energy Not Served** (**ENS**). This indicator measures the amount of electricity demand – measured in MWh – that is not met by generation in a given year (for a customer or customer group). France indicates 2320 MWh of ENS in 2016, including load shedding, while Germany provides no numbers. It may also be computed for future years as "Expected Energy Not Served" (EENS). It then combines both the likelihood and the potential size of any shortfall. This indicator is used to assess the security of supply as well as to set a reliability standard.

When it comes to defining a normative standard for supply interruption, there is no internationally agreed level. Acceptable durations of supply interruptions strongly depend on external factors such as climate and density of population and how much an economy is willing to pay for a reliable supply. Hence, reliability is in general different between developed countries and developing countries. Also, in the context of increasing shares of distributed renewable generation, the question arises whether consumers are willing to compromise on reliability for reductions in cost or increases in renewable generation.

In Germany, the following interruptions of supply are accepted:

- No interruption of supply is tolerated in the transmission grid. As described in the section before, the N-1 criterion is used to withstand disturbances. Several measures (such as topology optimisation, redispatch up to disconnecting selected loads) are taken to avoid a blackout.
- Across all voltage levels, a SAIDI below 1 h per year is targeted, or even maintaining the current level of less than 30 min - a level of supply security that many other European countries are not attaining.

In the context of electricity grid regulation, regulators may set targets in terms of acceptable reliability and may penalise grid operators if they do not achieve these targets (cf. Sect. 6.1.3).

5.1.4.2 Ancillary Services for Secure System Operation

As load and generation have to be balanced for each moment in time, so-called ancillary services (also called system services) are necessary for a secure system operation. Hence, ancillary services are the services and functions needed in the electric grid to allow stable operation and to facilitate and support the continuous flow of electricity. The term ancillary services refers to various functions required to maintain grid stability and security.[15] These generally include:

- **Frequency control and active power provision**
- **Voltage control and reactive power provision**
- **Short circuit management**
- **Restoration of supply**
- **Coordination and management of system operation**.

Frequency control: frequency control is carried out by the transmission system operators maintaining a balance between load and generation. Frequency has to be kept in Europe at 50 Hz within tolerable limits. An increased load (or a decreased generation due to a shortfall) reduces the frequency.

[15] To what extent and how these services may be procured on markets is discussed in Sects. 10.3 and 10.4.

This situation can be compared with a bike rider riding up a hill. If the bike rider wants to maintain his speed of the plain, he has to pedal stronger or provide more power. In the electricity network, additional power is necessary if the frequency has to be kept at the same level after a load increase or generation decrease. A reduced load or increased generation results in an increased frequency. This situation would be comparable with a bike rider going down a hill. Hence, he has to reduce his pedalling power if he wants to maintain his speed.

Instruments for frequency control for TSOs are the instantaneous reserve, the frequency containment reserve, two frequency restoration reserves and the replacement reserve (the latter four summarised under the terms of operating reserves or control energy), interruptible loads and frequency-dependent load shedding. They are discussed in the following.

Instantaneous reserve is available due to the inertia of the system, primarily provided by the power generators. In the current system, large-scale conventional power generators automatically provide this very short-term reserve through the rotating masses of their turbines and generators. With increasing shares of generation from inverter-based renewable plants, new concepts for preventing short-term frequency drops are yet required in the future (cf. Sect. 12.2).

Operating reserves (or reserve power) are the generating (and sometimes load management) capacities available within a short interval of time to meet demand in the case of any imbalance (resulting from unpredictable or imperfectly predictable events, such as forecast errors for load and variable renewables, as well as from power plant failures). Operating reserves are mainly distinguished according to the time intervals after an unexpected failure that they are in use. The terminology for the different operating reserves varies and in some countries, a differentiation between spinning and non-spinning reserve was used, while in others, it was called primary, secondary and tertiary, reserve.

In Europe, the terminology and concepts have been harmonised (cf. EC 2017a, b):

1. **Frequency containment reserves** (**FCR**) are used for the continuous control of frequency and should contain the frequency after an incident/imbalance within the synchronous area. Frequency containment is a joint action of all TSOs in the area, activated typically within 30 s at the latest and was formerly called primary reserve or spinning reserve. The power system is designed so that, under normal conditions, the frequency containment reserve always has at least a capacity of 3000 MW in continental Europe, corresponding to a parallel outage of two large power plants in Europe.
2. **Automatic frequency restoration reserve** (**aFRR**): The purpose of (automatic) frequency restoration reserve is to return the frequency to its normal range and to free up activated frequency containment reserves. It is applied within a control area (in case of an imbalance in the area) with a full activation time of 5 min in continental Europe (and between 2 and 15 min in other parts of Europe). Formerly, it was often called secondary reserve.
3. **Manual frequency restoration reserve** (**mFRR**): Activation is done manually and should free up activated frequency containment and automatic frequency

restoration reserves after 15 min. Formerly, it was often called tertiary reserve or non-spinning reserve.

4. **Replacement reserve** (**RR**): Activation is done manually with an activation lead-time exceeding 15 min. This form of reserve is only applied in some countries and serves to release faster reserves. Formerly, it was often called slow tertiary reserve.

Figure 5.9 summarises the four different operating reserves according to their time in use.

Interruptible loads are major demand units connected to the high and extra-high-voltage transmission grid. The load of these units is preferably very constant. Interruptible loads can reduce their consumption at short notice and for a predefined duration and thus, the production processes must allow for these flexibilities. Industries with large loads may offer this as a (remunerated) service for situations where the normal reserves are found to be insufficient. In Germany, the transmission system operators organise (together with the operating reserve) a tender for a defined interruptible capacity, today of 1500 MW for immediate interruptible loads (activated based on frequency changes within 1 s) and 1500 MW for fast interruptible loads (activated within 15 min).

Frequency-dependent load shedding is a measure used by the transmission system operators to avoid a total blackout of the power system. Frequency-dependent load shedding (also called rotational load) is an intended power shutdown, where electricity supply is stopped for selected areas reducing total load. If

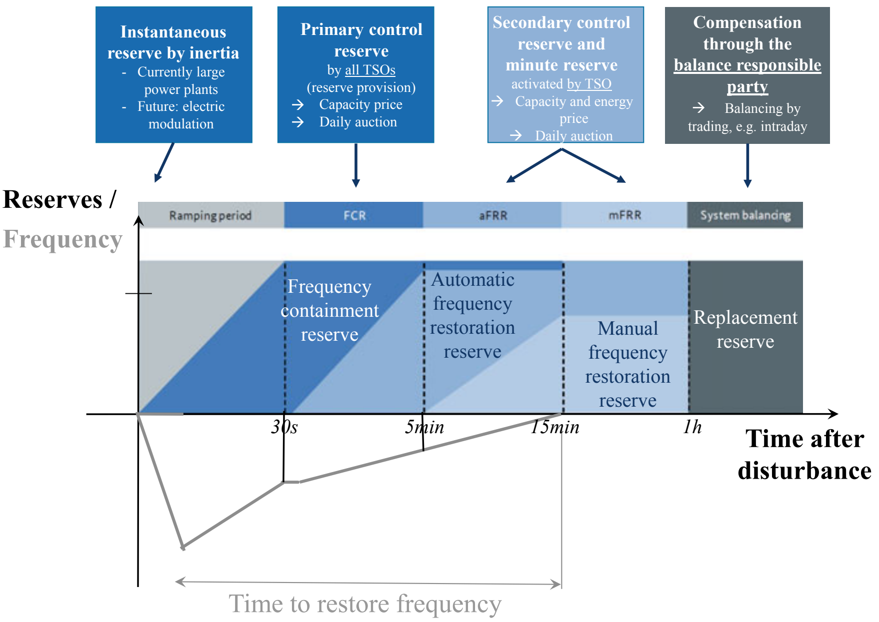


Fig. 5.9 Harmonised European reserve terminology

this load shedding is necessary for a longer time span, the shutdown is successively applied to different areas, so it is also called rolling blackout or feeder rotation. Frequency-dependent load shedding may be limited to a specific part of the electricity network or be more widespread and affect entire countries; by rotating from one area to the next, the individual blackout time of consumers is yet limited. In continental Europe (and other developed countries), frequency-dependent load shedding is extremely rare. However, it may be more common in developing countries or even a normal daily event when electricity generation capacity is underfunded or infrastructure is poorly managed.

Voltage control: the transmission and distribution system operators have the task to keep the voltage within a defined tolerable limit in their area of responsibility – the maximum allowable deviation is usually $\pm 10\%$ at the connection points for end-users. In transmission grids, voltage stability is intimately related to the reactive power balance, as Ohmic resistances are small compared to inductances and condensers.[16] Therefore voltage stability in transmission grids is ensured through a controlled feed-in of reactive power. In lower voltage distribution grids, voltage drops may also result from Ohmic losses, yet both high-voltage transmission grids and distribution grids need reactive power. They use it to compensate their reactive power behaviour on power lines, which can be capacitive in case of a low utilisation or inductive in case of high utilisation. As reactive power cannot be transported over large distances, local compensation is required. As an ultimate measure to cope with voltage instability, the grid operator may initiate a voltage-induced redispatch or a voltage-induced load shedding to maintain the voltage within the admissible band and prevent a voltage collapse. Capacitive and inductive reactive power can be provided by (conventional) generation units or by reactive power compensators [e.g. flexible AC transmission systems (FACTS)], while the provision by renewable power generators is still under research and demonstration.

Short circuit management: short circuits are typically severe faults and are managed mainly by activating circuit breakers, installed notably in conjunction with transformers. The circuit breakers are triggered by the overcurrent induced by the short circuit and provided by the conventional large-scale generators. Those automatically increase their production when circuit resistance is reduced through the short circuit.

Restoration of supply: after a (large-area) blackout, the transmission system operator and the distribution system operator must restore the electricity supply within the shortest time. This necessitates the coordination of sufficient generation injection together with selected electricity grids and loads. Therefore, power units are needed that have of **black start** capabilities, i.e. they may start their own operation without an external source of electricity. They can even establish the standard frequency in the grid by themselves. In general, selected conventional power plants and pump storage power plants are used to enable restoration of supply.

[16] This may be derived from the AC power flow Eq. (5.11) under the assumption $X \gg R$, which is also used in the linear DC power flow approximation (cf. Sect. 5.1.2.3).

Also, renewable power plants with self-commutated inverters may provide black start capabilities. However, due to the small sizes of the units, the organisation and integration of the corresponding processes are still under research and development.

Coordination and management of system operation: finally, transmission and distribution system operators are responsible for a stable system operation. Therefore, they monitor the well-functioning of electricity networks, perform a power flow analysis including monitoring the N-1 criterion, request redispatch and curtailment of renewables and coordinate the provision of the operating reserve with other neighbouring system operators.

Traditionally, ancillary services have been provided by conventional generators. However, larger shares of variable renewables require that they also take over system responsibility and system services. With the development of smart grid technologies, a shift in the **equipment** enables the provision of ancillary services. Thereby **load-commutated power inverters** can be differentiated from **self-commutated inverters**: **Load-commutated power inverters** need the electricity grid for operation. The power line controls the commutation of power (conversion from DC to AC) so that, if there is a failure in the power grid, the renewable system cannot feed power into the line. In general, thyristors are used for the inverter, which are pretty robust and cheap. **Self-commutated inverters** integrate electronic switches, significantly reducing reactive power consumption and current harmonics (cf. previous section). However, the switching process used in self-commutated inverters tends to induce isolated operation after a power failure. This has to be considered from a safety perspective (in a grid-connected system) and appropriate precautionary measures need to be taken. Due to the stricter harmonics regulations and the recent sharp price reductions for these devices, today's trend is towards self-commutated inverters. Furthermore, they enable a better provision of ancillary services. However, further research is needed for the system integration of renewable energies, including ancillary services.

5.2 Storage

5.2.1 Basics

Storage systems can bridge the temporal gaps between electricity production and consumption. As electricity can hardly be stored directly, typically, electricity storage systems transform electricity into a storable energy carrier. The most well-known form of energy storage in the electricity system is to store potential energy by using pump storage power plants (see Sect. 4.2.1). Pump storage power plants have been used for about 100 years and are still the dominating technology for storing electricity. In 2019, pump storage power plants' installed capacity was about 160 GW worldwide and about 55 GW in Europe (International Hydropower Association 2020, p. 45). The total global storage capacity of pump storage power plants is estimated at almost 9 TWh (International Hydropower Association 2018).

Electricity storage systems can use electricity procured during times with low (or even negative) wholesale electricity prices (e.g. due to high feed-in of electricity produced by renewable energies and low demand) to provide electricity in times of high wholesale electricity prices. In addition to this business model, storage systems can also provide different (ancillary) services (see Sect. 5.1.4.2), e.g. frequency control reserve, reactive power or black start capability, congestion management and reduce curtailment of renewable energies.[17]

5.2.2 Technologies

In the following, different energy storage systems having electricity as an input energy carrier are presented. There are various possibilities to classify the multiple forms of such electricity storage systems, e.g. with regard to their principle of storage (electric, electrochemical, mechanical, thermal) (see Fig. 5.10).

Other possibilities to classify energy storage technologies are their energy capacity or their **energy-to-power (E/P) ratio** (kWh/kW). The latter is also sometimes called the duration of discharge (cf. Fig. 5.11), whereas the inverse, the **power-to-energy ratio** is for batteries also labelled C-rate or charge/discharge rate. E.g. a C-rate of 2 implies that the battery may be charged or entirely discharged within half an hour. For short-term energy storage, systems with high efficiencies are available, yet they can only store a limited amount of energy. Available options for storing large amounts of energy are characterised by either limited availability or comparatively low round-trip efficiencies (cf. e.g. Heffels 2015, pp. 23–37).

In the following, key storage technologies are discussed by increasing duration of discharge, i.e. starting from very short-term storage (see Fig. 5.11). An

Electric	Chemical and electrochemical	Mechanical	Thermal
• Super-capacitators (SC) • Superconductive magnetic energy storage (SMES)	• Power-to-gas • Power-to-liquid • Power-to-chemicals • Batteries	• Pumped storage • Compressed air storage • Flywheels	• Power-to-heat

Fig. 5.10 Different principles of electricity storage. *Source* Own illustration based on Agora Energiewende (2014, p. 36)

[17] In Pérez-Díaz et al. (2015), trends (e.g. the provision of different ancillary services) and challenges (e.g. the development of new scheduling strategies) with regard to the operation of pump storage power plants are presented.

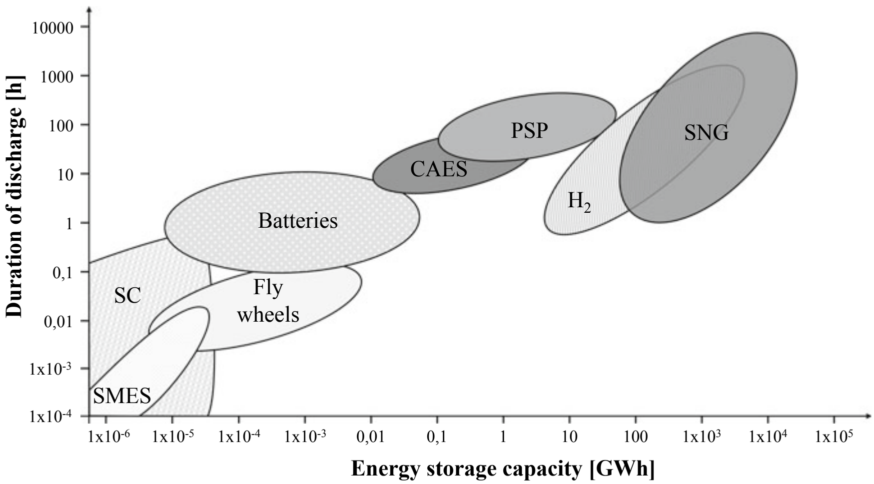


Fig. 5.11 Electricity storage systems arranged according to their duration of discharge and storage capacity. *Source* Own illustration based on Sterner and Thema (2017, p. 654)

introduction into different energy storage technologies and a comparison of these technologies according to technoeconomic characteristics can be found inter alia in Fuchs et al. (2012), IEA (2014a), IRENA (2017), Kaschub (2017, pp. 48–50), and Mongird et al. (2019).

Supercapacitors (SC) use electrostatic fields, **superconductive magnetic energy storage** (SMES) systems use magnetic fields to store electricity directly. The key benefits of both technologies are high power densities, high efficiencies, low discharge times, features that enable them to store and release electric energy extremely fast, but the power can only be released for a very short duration (cf. e.g. IEA 2014b, p. 16).

Flywheel energy storage systems are another example of a short-term energy storage system. In such a system, energy is used for spinning a mass at high speeds. By accelerating the flywheel with the help of a motor, kinetic energy is stored that can be used to produce electric energy with the help of a generator. Flywheel energy storage systems show relatively high efficiencies of about 85% (cf. e.g. IRENA 2017, p. 62) and low discharge times. Such an installation can provide a relatively high power output but only for a very short period (cf. e.g. Babrowski 2015, p. 26).

So-called primary **batteries** can only be used once; they are not rechargeable. In electricity systems, secondary batteries (also called accumulators) are therefore much more relevant, as they can be recharged. Typically, such rechargeable batteries are composed of many electrochemical cells that can convert chemical energy into electricity (discharging the battery) and the other way round (charging the battery). Each cell is made up of different components: a cathode, being the positive electrode, an anode, being the negative electrode, a separator between these two electrodes and an electrolyte enclosing the separator. An accumulator's main principle is that ions, which are characterised by having an electrical charge, can pass from one electrode to the other through the electrolyte. Such a passing is not possible for electrons, they flow via the electric network.

For many years, **lead-acid batteries** were the dominating accumulators in use. In a lead-acid battery, the anode and the cathode are lead plates, and as an electrolyte, sulfuric acid is used. Higher round-trip efficiencies (up to 95% compared to maximal 85% for a lead-acid battery (cf. e.g. EASE/EERA 2013, p. 20, IRENA 2017, p. 67 and 86) and higher cycle stabilities, meaning that they can be more often charged and discharged, make **lithium-ion batteries** more attractive. In the last years, the investment costs for such batteries have decreased drastically, as lithium-ion batteries are used in many energy storage applications (like in portable applications and electric vehicles[18]). There are many different forms of lithium-ion batteries, which vary in the materials used for the anodes, cathodes, and electrolytes (cf. e.g. IRENA 2017, pp. 63–81). For instance, the cathode of such a battery might consist of a lithium nickel manganese cobalt oxide or a lithium cobalt oxide, the anode might be made out of carbon. Right between these two electrodes, a separator, which is permeable for lithium ions, is placed. Lithium-ions are released at the cathode during the process of charging the battery, and they travel through the electrolyte to the anode. The external power source induces the electrons to flow through the external circuit to the anode, as they cannot pass the separator. The battery is now charged, and as soon as power is needed, it can be discharged: lithium-ions flow back to the cathode, where lithium and the electrons coming via the outer circuit will be deposited again (for more information about the operation of lithium-ion batteries, see Stadler et al. (2017, pp. 281–304)). Although being a very promising technology, lithium-ion batteries raise a number of issues, of which the safety and the ageing of the battery systems are of particular importance. Different hazard sources within lithium-ion batteries exist, mainly because some of the materials used are flammable. Concerning the ageing of the battery, the expected lifetime of about 15 years is influenced by different parameters – typically, the ageing effects are manifold and can be differentiated in the so-called calendar ageing and the so-called cycle ageing (cf. Kaschub 2017, pp. 58–63). Cycle ageing considers the realisable number of charging and discharging activities (number of cycles), which depends among other things on the depth of discharge (DOD). Higher DODs seem to lead to disproportionate ageing. On the other hand, calendar ageing depends inter alia on the state-of-charge (SOC), a too high SOC seems to accelerate the ageing process. Furthermore, battery ageing is generally higher at higher temperatures. Finally, it has to be mentioned that the end of life (EOL) of a lithium-ion battery does not mean that the battery cannot be used anymore. Instead, the end of life (EOL) of a lithium-ion battery is typically defined as the point in time when the battery has degraded so far that only a certain fraction

[18] This shows that the development of storage technologies for applications in other sectors can have strong effects for applications of storage technologies in the energy sector.

of its original capacity remains. Due to higher requirements for the use in the automotive sector, this value is typically higher in the case of mobile compared to stationary applications (80% compared to 70% according to e.g. Kaschub 2017, p. 60), which might provide the opportunity to use disused batteries of electric vehicles for certain stationary applications (second life application). There is intensive research for other materials to be used in future batteries, e.g. because some of the materials used at the moment are extracted under questionable conditions regarding environmental and social aspects.

So-called **flow batteries** are another type of rechargeable (secondary) batteries. A specific feature of flow batteries is that they have two external tanks containing two electrolytes, chemical components dissolved in liquids. The size of these tanks, which can relatively easily be varied, determines the energy stored in the battery. The fluids in the tanks are pumped into the battery cell, where again the anode and the cathode are separated by a membrane. In the cell, reduction and oxidation processes take place (therefore, these batteries are often called redox flow batteries); only the ions can move through the membrane, the electrons flow via the circuit (cf. e.g. ESA 2018). Flow batteries currently reach efficiencies of about 70% (IRENA 2017, p. 92).

In the last years, the number of installed battery storage systems increased in several countries like Germany and the USA. This growth can partly be explained by the fact that such systems can directly be installed at the location of the electricity consumer (so-called behind-the-meter energy storage) to increase the share of consumption of the self-produced electricity.

The main principle of **compressed air energy storage systems** (**CAES**) is to compress air and store this compressed air, e.g. in an underground reservoir. If the heat produced when compressing the air is not stored, the system is called diabatic compressed air energy storage (D-CAES) (cf. left side of Fig. 5.12). In such a system, the compressed air has to be reheated in order to produce electricity in a turbine. Here, an additional fuel has to be used, e.g. gas is to be burned. If the heat is stored and released back to the air when producing electricity, a so-called adiabatic compressed air energy storage (A-CAES) system is implemented (see right side of Fig. 5.12), where no additional fuel is needed. This implies higher overall efficiency since energy losses related to the dissipation of compression heat are avoided. Worldwide only two diabatic compressed air energy storage plants exist, one in Germany, one in the USA. Whereas these plants have round-trip efficiencies of up to 55%, adiabatic compressed air energy storage systems are supposed to reach efficiencies of up to 70% (cf. e.g. Haubrich 2006, pp. 5–6). In both existing systems, salt caverns are used to store the air under a pressure of up to 75 bar, but only in the US installation, the waste heat of the gas turbine is used to warm up the air again, resulting in a higher efficiency.

An established and widespread type of energy storage in the electricity system are **pump storage power plants** (PSP, see also Sect. 4.2.1). Here, energy is stored as potential energy by pumping water from a lower reservoir to an upper reservoir. Obviously, the more water can be pumped and the higher the water can be pumped the more energy can be stored. Sometimes, different reservoirs are combined and

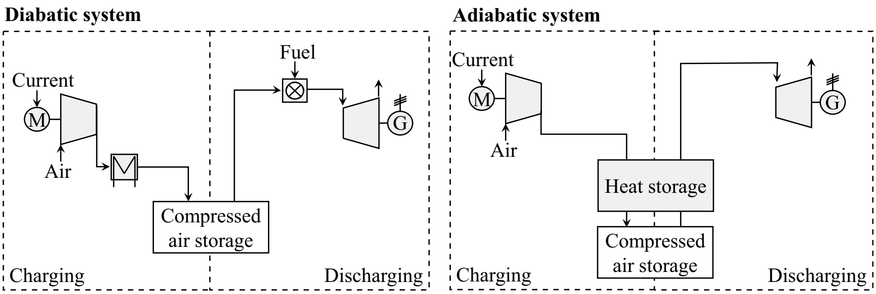


Fig. 5.12 Schematic representation of a diabatic (left) and an adiabatic (right) compressed air energy storage. *Source* Own illustration based on Calaminus (2007)

even pumped-storage hydroelectricity plants are designed in cascades or combined systems, in order to make optimal use of given inflow and topology conditions along with high flexibility in operation. Pumped-storage plants usually have relatively high round-trip efficiencies of about 80%, but no more fundamental technological improvements are expected to be realised (IRENA 2017, p. 52). If the reservoir has only (natural or man-made) water inflows without the option to pump water, the plants are called reservoir-type hydropower plants. These two types of storage systems are very flexible, have fast load gradients and provide black start capabilities (cf. e.g. Sterner et al. 2017, pp. 694–696). On the other hand, in many countries, possibilities for further development seem to be limited, as appropriate sites are scarce, and the local acceptance of new projects is low, e.g. due to interferences with the landscape.

In an electricity system dominated by fluctuating renewable energy sources like wind and PV, storage systems are needed to store energy for longer time periods, e.g. to bridge the time span of several days with hardly any wind and PV feed-in (dark calm). This can be realised by transforming electrical energy into chemical energy. A prominent example is the so-called **power-to-gas** (**PtG**) technology, which uses (green) electricity to split water into (green) hydrogen and oxygen via electrolysis. The produced **hydrogen** (H_2) can then be fed into the existing gas grid up to a specific concentration or in pure form into a future hydrogen grid. Application possibilities include internal combustion engines and fuel cells in the transport sector, electricity production in a hydrogen (gas) turbine, uses in the energy intensive industry (e.g. iron and steel, chemical industry) or the transformation into methane (**synthetic natural gas**, short: **SNG**) by methanation (cf. Fig. 5.13).

The electrolysis is an endogenous reaction, which needs energy. In contrast, the methanation represents an exothermal reaction, for which a catalyst is needed (cf. Heffels 2015, pp. 29–37):

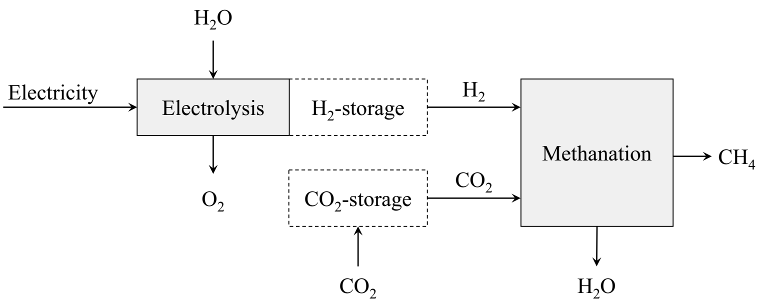


Fig. 5.13 Schematic representation of a power-to-gas process. *Source* Own illustration based on Sterner (2009, p. 106)

$$\text{electrolysis: } 2H_2O \rightarrow 2H_2 + O_2$$
$$\text{methanation: } CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O.$$

Beside hydrogen, the methanation process needs CO_2 as an input factor. For the provision of CO_2 different sources could be used, e.g. CO_2 produced by biogas plants or in industrial processes. Furthermore, extraction of CO_2 from ambient air (**direct air capture**, DAC) could be an alternative, albeit this technology still is associated with rather high costs.

The main advantage of PtG is the enormous capacity of the natural gas grid, which can be used. On the other hand, the efficiency to produce hydrogen, respectively synthetic natural gas is still relatively low (about 75/60%, cf. e.g. Heffels 2015, pp. 134–135), and the investment costs are rather high. Furthermore, electrolysers used for the described purpose have to be able to be used in a rather flexible mode of operation.

In addition to PtG, other possibilities of transforming electrical energy into other forms of energy can also be used. The technical term **power-to-X** (PtX) comprises many different forms, like chemicals (PtC) or heat (PtH). Power to chemical stands for the use of electricity to produce chemical products, e.g. chemical feedstocks. Furthermore, electricity could also be used to produce heat by electric heating units (PtH), like heat pumps or even heating rods. Without retransforming the heat into electricity, this helps to substitute conventional energy carriers that would have been used to produce the needed heat otherwise, with retransforming to decouple weather-dependent electricity production and demand. The storage of heat can be differentiated into sensible heat storage (here the temperature of the storage materials changes), latent heat storage (here the phase of the storage materials changes) and thermochemical heat storage (here reversible thermochemical reactions are realised) (cf. e.g. Bauer et al. 2012).[19] Although PtX might be an interesting opportunity to use electricity produced during hours with a lot of feed-in from

[19] Efficient thermal energy storage systems are of utmost importance for CSP as they allow to at least partly decouple electricity production from solar radiation (see Sect. 4.2.3.1).

renewables, it is not described herein in more detail because typically the chemical energy and heat will not be retransformed into electricity again.

Overall, the transformation of energy systems worldwide towards increased use of fluctuating renewable energy carriers will require increased use of storage (see Chap. 12). Multiple technologies are available today, yet still, further technological progress in terms of efficiency and costs is needed to provide a good complement to the fluctuating infeed of renewables.

5.3 Further Reading

Various textbooks are dealing with the fundamentals of power system analysis.

Glover, J. D., Overbye, T. J., & Sarma, M. S. (2017). Power system analysis and design. 6th edition. Boston: Cengage learning.

The book by Glover et al. provides an in-depth introduction in English.

For German speaking readers, the following volumes may be of interest:

Crastan, V. (2015). Elektrische Energieversorgung 1: Netzelemente, Modellierung, stationäres Verhalten, Bemessung, Schalt- und Schutztechnik. 4th edition. Berlin, Heidelberg: Springer.

Schwab, A. J. (2019). Elektroenergiesysteme – Erzeugung, Übertragung und Verteilung elektrischer Energie. 6th edition. Berlin, Heidelberg: Springer.

The first book focuses on the mathematical modelling of power grids, the second provides a more detailed introduction to the different grid elements and technologies.

The first textbook dealing extensively with the pricing of electricity based on the power system equations is now a classical text:

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., & Bohn, R. E. (1988). Spot Pricing of Electricity. Norwell: Kluwer Academic Publishers.

Regarding energy storage technologies related to power systems, few books provide a concise overview of different technologies. This is undoubtedly due to the broad range of technologies and the rapid pace of development, particularly regarding battery technologies over the last decade.

Sterner, M., & Stadler, I. (Eds.) (2017). Energiespeicher - Bedarf, Technologien, Integration. 2nd edition. Berlin: Springer Vieweg.

This collective volume in German includes contributions on different storage technologies and their role in future energy systems

.

5.4 Self-check of Knowledge and Exercises

Self-check of Knowledge

1. What are the main characteristics of the current electricity networks?
2. Write down the mathematical formulations for Ohm's and Kirchhoff's laws and describe them.
3. Describe the concepts of active, reactive and apparent power and illustrate how they are related.
4. What are the key characteristics of the main components of electricity networks and for what purposes are they used?
5. Reliability, availability and N-1 security are key concepts for system operation. Explain these concepts and define the ancillary services that are used to support system operation.
6. What physical principles may be used to store electricity and what are the available technologies that apply these principles? What key parameters may be used to characterise different storage technologies?
7. What are the key advantages and disadvantages of batteries, pumped hydro storage and power-to-gas technologies in storing electricity?
8. Calculate the efficiency of the transformation of electricity into methane and back to electricity using a gas turbine.

Exercise 5.1: Power Triangle and Power Factor

1. A load has a reactive power demand of 300 kvar and 0.4 MW active power demand. Please sketch the corresponding power triangle and label it.
2. What is the power factor of the triangle considered in the previous task?
3. If you aim at a power factor of 0.99 and you can compensate the reactive power freely, what would be your apparent power? Please compare it to the apparent power in the first task.
4. Is it a capacitive or inductive load?

Exercise 5.2: Meshed Electrical Circuits

1. Simple example:
   A current of 1 A flows through a current divider $(R_1 = 1\,\Omega,\ R_2 = 2\,\Omega)$ (cf. Fig. 5.14).

   (a) What is the joint resistance?
   (b) What are the currents in the branches?

2. Application example:
   The offshore wind park "Nordsee Ost" has an installed capacity of 228 MW and is connected to the onshore grid as depicted in Fig. 5.15. Imagine that the park would feed in maximally. Given the maximum capacity of each line, could the

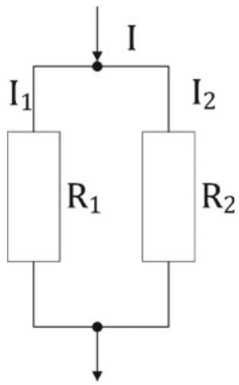


Fig. 5.14 Schematic representation of a current divider (parallel resistances)

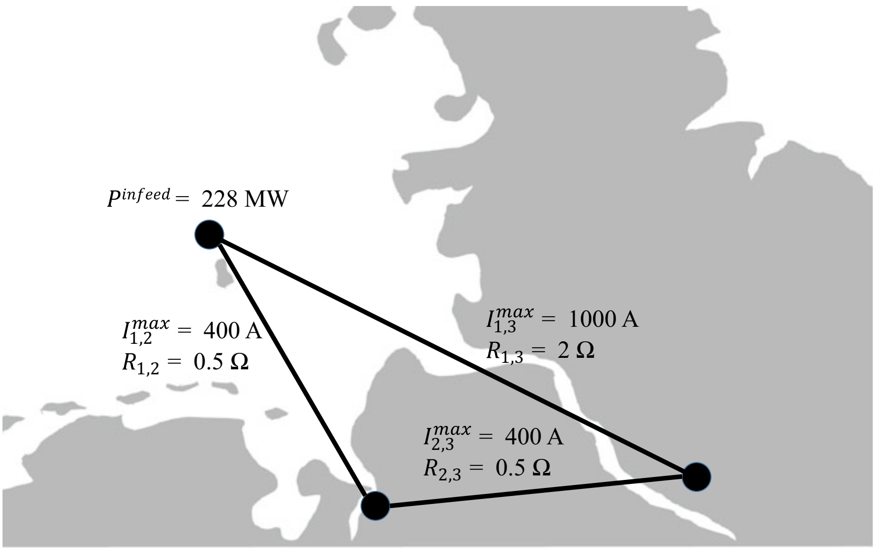


Fig. 5.15 Schematic representation of a connection of an offshore wind park via two lines

wind park be extended by 38 MW? Hint: The offshore park is connected to the 380 kV level.

Exercise 5.3: Power Flow Analysis

We consider the power flow in the system depicted in Fig. 5.16. The graph shows the results for the base case considered in the following questions (1) to (7).

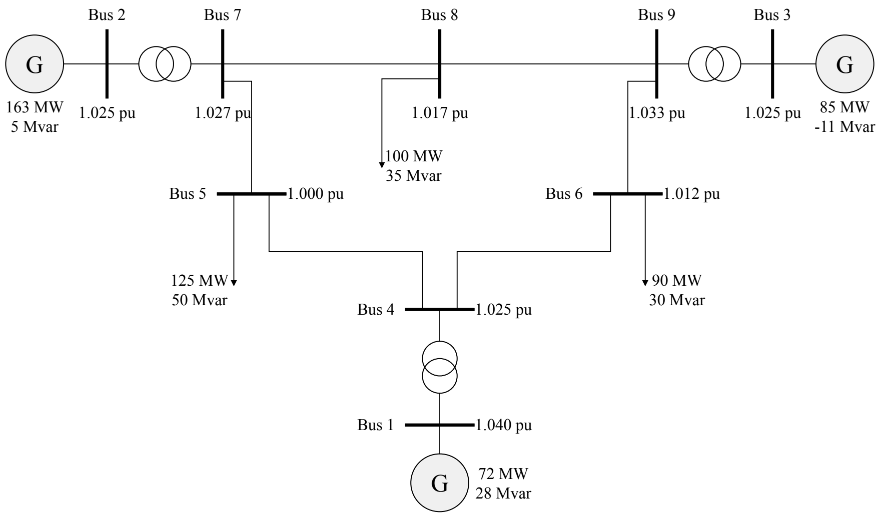


Fig. 5.16 Exemplary system for power flow calculation

1. How much active power is generated?
2. How much active power is consumed?
3. If there is a difference, why?
4. Quantify the transmission losses.
5. What does "p.u." mean?
6. Compute the power factor cosφ for each load.
7. Is bus 2 a PU or a PQ node?
8. Change of active power at bus 5:
   How is the voltage level at bus 5 affected by a load increase?

References

ABB. (2019). *HVDC – Economic and environmental advantages*. Available at: https://new.abb.com/systems/hvdc/why-hvdc/economic-and-environmental-advantages [Accessed August 1, 2019].

Agora Energiewende. (2014). *Stromspeicher in der Energiewende, Untersuchung zum Bedarf an neuen Stromspeichern in Deutschland für den Erzeugungsausgleich, Systemdienstleistungen und im Verteilnetz*, Berlin. Available at: https://static.agora-energiewende.de/fileadmin/Projekte/2013/speicher-in-der-energiewende/Agora_Speicherstudie_Web.pdf [Accessed August 1, 2019].

Babrowski, S. (2015). *Bedarf und Verteilung elektrischer Tagesspeicher im zukünftigen deutschen Energiesystem Produktion und Energie (7)*. KIT Scientific Publishing and Dissertation: Karlsruhe Institute of Technology (KIT).

Baldick, R. (2003). Variation of distribution factors with loading. *IEEE Transactions on Power Systems, 18*, 1316–1323.

Bauer, T., Steinmann, W., Laing, D., & Tamme, R. (2012). Thermal energy storage materials and systems. *Annual Reviews of Heat Transfer, 15*, 131–177.

Bergen, A. R., & Vittal, V. (2000). *Power system analysis* (2nd ed.). National Taipei University of Technology (Taipei Tech).

Calaminus, B. (2007). *Innovatives Druckluftspeicherkraftwerk der EnBW*. Presentation at Energietage Hannover. Available at: https://docplayer.org/22773261-Innovatives-druckluftspeicherkraftwerk-der-enbw.html [Accessed June 11, 2022].

CEER. (2018). *CEER benchmarking report 6.1 on the continuity of electricity and gas supply*. CEER. Available at: https://www.ceer.eu/documents/104400/-/-/963153e6-2f42-78eb-22a4-06f1552dd34c [Accessed June 11, 2022].

EASE/EERA. (2013). *Technical annex, European energy storage technology development roadmap towards 2030*. Available at: https://ease-storage.eu/wp-content/uploads/2015/10/EASE-EERA-recommendations-Annex-LR.pdf [Accessed June 11, 2022].

EC. (2017a). *Commission Regulation (EU) 2017/1485 establishing a guideline on electricity transmission system operation*. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2017:220:TOC [Accessed January 11, 2021].

EC. (2017b). *Commission Regulation (EU) 2017/2195 establishing a guideline on electricity balancing*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2017.312.01.0006.01.ENG&toc=OJ:L:2017:312:TOC [Accessed January 11, 2021].

ESA. (2018). *Flow batteries*. Available at: http://energystorage.org/energy-storage/storage-technology-comparisons/flow-batteries [Accessed December 18, 2018].

Fuchs, G., Lunz, B., Leuthold, M., & Sauer, D. U. (2012). *Technology overview on electricity storage − Overview on the potential and on the deployment perspectives of electricity storage technologies*. WTH ISEA. Available at: https://www.researchgate.net/publication/299425278_Technology_Overview_on_Electricity_Storage_-_Overview_on_the_potential_and_on_the_deployment_perspectives_of_electricity_storage_technologies [Accessed June 11, 2022].

Heffels, T. (2015). *Kraftwerks- und Speicherbedarf bei hohen Anteilen erneuerbarer Energien*. Dissertation: Karlsruhe Institute of Technology (KIT). Available at: https://publikationen.bibliothek.kit.edu/1000048819 [Accessed August 19, 2022].

Helmenstine, A. M. (2019). *Table of electrical resistivity and conductivity*. Available at: https://www.thoughtco.com/table-of-electrical-resistivity-conductivity-608499 [Accessed September 30, 2021].

IAEW, et al. *Verbesserte Integration großer Windstrommengen durch Zwischenspeicherung mittels CAES*. RWTH IAEW. Available at: https://www.tib.eu/de/suchen/id/TIBKAT:533612632/ [Accessed June 11, 2022].

ICF Consulting. (2003). *Overview of the potential for undergrounding the electricity network in Europe*. Prepared for the DG TREN/European Commission. ICF Consulting. Available at: https://ec.europa.eu/energy/sites/ener/files/documents/2003_02_underground_cables_icf.pdf [Accessed June 11, 2022].

IEA. (2014a). *Technology roadmap energy storage*. IEA. Available at: https://iea.blob.core.windows.net/assets/80b629ee-597b-4f79-a236-3b9a36aedbe7/TechnologyRoadmapEnergystorage.pdf [Accessed June 11, 2022].

IEA. (2014b). *Energy storage technology roadmap, technology annex*. IEA. Available at: https://large.stanford.edu/courses/2014/ph240/galvan-lopez2/docs/AnnexA_TechnologyAnnexforweb.pdf [Accessed June 11, 2022].

International Hydropower Association. (2018). *The world's water battery: Pumped hydropower storage and the clean energy transition*. IHA. Available at: https://assets-global.website-files.com/5f749e4b9399c80b5e421384/5fa80177d7fd2673249b3471_the_worlds_water_battery_-_pumped_storage_and_the_clean_energy_transition_2_1.pdf [Accessed June 11, 2022].

International Hydropower Association. (2020). *Hydropower status report 2020*. IHA. Available at: https://hydropower-assets.s3.eu-west-2.amazonaws.com/publications-docs/2020_hydropower_status_report.pdf [Accessed June 11, 2022].

IRENA. (2017). *Electricity storage and renewables: Costs and markets to 2030*, Abu Dhabi. Available at: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2017/Oct/IRENA_Electricity_Storage_Costs_2017.pdf [Accessed June 11, 2022].

Kaschub, T. (2017). *Batteriespeicher in Haushalten unter Berücksichtigung von Photovoltaik, Elektrofahrzeugen und Nachfragesteuerung Produktion und Energie (23)*. KIT Scientific Publishing and Dissertation: Karlsruhe Institute of Technology.

Mongird, K., et al. (2019). *Energy storage technology and cost characterization report*. Available at www.pnnl.gov/main/publications/external/technical_reports/PNNL-28866.pdf [Accessed: June 11, 2022].

Oeding, D., & Oswald, B. R. (2011). *Elektrische Kraftwerke und Netze* (7th ed.). Springer.

Overbye, T. J., Cheng, X., & Sun, Y. (2004). A comparison of the AC and DC power flow models for LMP calculations. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*. Available at: https://documents.pserc.wisc.edu/documents/publications/papers/2003_general_publications/overbyehicss2004.pdf [Accessed June 11, 2022].

Pérez-Díaz, J. I., Chazarra, M., García-González, J., Cavazzini, G., & Stoppato, A. (2015). Trends and challenges in the operation of pumped-storage hydropower plants. *Renewable and Sustainable Energy Reviews, 44*, 767–784.

Schwab, A. J. (2015). *Elektroenergiesysteme* (4th ed.). Springer.

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., & Bohn, R. E. (1988). *Spot pricing of electricity*. Kluwer Academic Publishers.

Stadler, I., Riegel, B., Ohms, D., Cattaneo, E., Langer, G., & Herrmann, M. (2017). Elektrochemische Energiespeicher. In M. Sterner & I. Stadler (Eds.), *Energiespeicher – Bedarf, Technologien, Integration* (2nd ed., pp. 229–326). Springer Vieweg.

Sterner, M. (2009). *Bioenergy and renewable power methane in integrated 100% renewable energy systems—Limiting global warming by transforming energy systems*. Kassel University Press and Dissertation: U Kassel.

Sterner, M., Stadler, I., Eckert, F., & Thema, M. (2017). Speicherintegration in einzelnen Energiesektoren. In M. Sterner & I. Stadler (Eds.), *Energiespeicher – Bedarf, Technologien, Integration* (2nd ed., pp. 685–767). Springer Vieweg.

Sterner, M., & Thema, M. (2017). Vergleich der Speichersysteme. In M. Sterner & I. Stadler (Eds.), *Energiespeicher – Bedarf, Technologien, Integration* (2nd ed., pp. 645–682). Springer Vieweg.

Todem, C. (2004). *Methoden und Instrumente zur gesamtsystemischen Analyse und Optimierung konkreter Problemstellungen im liberalisierten Elektrizitätsmarkt*. Dissertation: TU Graz.

6 Regulation: Grids and Environment

Even in liberalised energy markets, grids are still natural monopolies, which need to be regulated. Another reason for government intervention into power systems are so-called external effects, which are defined in economics as impacts of one individual's action on another individual without corresponding market transaction. Environmental damages are a blatant case of external effects, and hence, government intervention is necessary to obtain efficient solutions. Therefore, the chapter aims at answering the following key questions:

- Why do simple market-based approaches not work for electricity?
- What are the alternatives to regulation?
- Which emissions stem from the production of electricity?
- What are the main environmental problems related to electricity systems?
- Which instruments are applicable for fighting environmental problems?
- Which instruments are used for limiting climate change?

Section 6.1 addresses the need and the possibilities for regulation of the electricity grid. The environmental challenges of electricity systems and the possible policy responses are then discussed in Sect. 6.2.

Key Learning Objectives
After having gone through this chapter, you will be able to

- Describe different forms of grid regulation and their practical challenges.
- Understand so-called Ramsey-prices.
- Describe emissions caused by electricity generation and the corresponding environmental impacts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-030-97770-2_6.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_6

- Describe the different phases of life cycle assessment.
- Understand policy instruments to limit climate change and corresponding binding agreements.

6.1 Grid Regulation

The mainstream economic theory claims that competitive markets will deliver outcomes that may not be outperformed systematically by any form of government intervention. This statement has been formalised through the two fundamental theorems of welfare economics, cf. e.g. Varian (2014). Under well-defined conditions, market outcomes are **Pareto efficient**; i.e. it is impossible to improve one individual's situation through government intervention without at least one other individual being made worse-off. Welfare economics does not claim that such market results satisfy any pre-established fairness condition nor that there is a unique Pareto efficient outcome. Yet, mainstream economics argues that it is unnecessary to interfere in market competition and price formation mechanisms to obtain fair welfare distribution based on the second fundamental theorem of welfare economics. Instead, such distributional issues may be handled separately through redistribution measures, preferably lump-sum transfers. These do not interfere with market price formation and hence do not distort the allocative efficiency of market mechanisms, i.e. the incentives for the most efficient use of scarce resources like energy.

For this key result on the efficiency of market-based approaches to hold, a certain number of assumptions have however to be fulfilled (see also Sect. 7.1.1). One assumption is that there are no natural monopolies, or more precisely, no subadditive and irreversible cost structures over the relevant range of outputs. This condition is violated for electricity grids, as will be discussed in more detail in Sect. 6.1.1. The resulting key regulatory recipes are subsequently addressed in Sect. 6.1.2. Section 6.1.3 focuses on the nowadays widely applied so-called performance-based grid regulation. The pricing of network services and the resulting challenges are discussed in Sect. 6.1.4, notably with regard to distributed renewable energies.

6.1.1 Fundamentals of Electricity Market Regulation

Electricity grids are one example of a number of networks that form monopolistic bottlenecks. Other examples for such networks are distribution grids for natural gas and water, telecommunications and railway infrastructure.

A **monopolistic bottleneck** is a **natural monopoly** with irreversible or **sunk costs**. The essential characteristic of a natural monopoly is **subadditive costs**. The following inequality describes subadditive costs:

$$C(q_1 + q_2) < C(q_1) + C(q_2). \tag{6.1}$$

Thereby $C(q)$ is the cost function for the production of the quantity q of a given good. The statement of inequality (6.1) is then that the costs for producing the total quantity $q_1 + q_2$ are lower than the sum of the costs when q_1 and q_2 are produced separately.[1] This has to be true for the whole relevant range of output. The definition of cost subadditivity may be generalised to the case of multiple goods with joint production:

$$C\left(\sum_i q_i\right) < \sum_i C(q_i). \tag{6.2}$$

If costs are subadditive, a single monopolistic company will be more cost-efficient than any number of competing utilities – the sector is subject to a "natural" monopoly. This is the case with electricity grids: providing a network connection to any given number of customers within one region will always come at a lower cost if done through one network rather than through several, partly overlapping networks. In electricity generation, there are by contrast economies of scale only up to a specific size, e.g. for coal plants up to the current upper limit of about 1000 MW nameplate capacity. Beyond that size, there are no subadditive cost structures to be expected. Hence, competition is likely to function for sufficiently large markets like those of most European countries.

Following Baumol et al. (1982) and others, a natural monopoly by itself does not require government regulation. Regulation is only necessary if costs are not only subadditive but also (at least partly) irreversible. Once a cable is buried in the ground, a large part of the cost is "**sunk cost**", not only in the literal sense but also in the economic sense of not being recoverable. Even if the cable would be dug out again and sold on the market, only a small fraction of the initial cost could be recovered. An example of a market with a natural monopoly but without (or only minor) sunk costs is the airlines market. A large carrier will benefit from subadditive costs since it can use a hub-and-spoke network to provide connection services between multiple destinations. Nevertheless, the market is "contestable"; i.e. new entrants may try to break monopolies and oligopolies since they have only limited irreversible cost. If their business model turns out to be unprofitable, they may still get back essential parts of their upfront investment into aeroplanes by reselling them on a relatively liquid secondary market.

[1] Note that economies of scale, defined through: $C(\lambda \cdot q) < \lambda \cdot C(q)$ for $\lambda > 1$, are a sufficient condition for subadditive costs, but not a necessary one over the whole output range. Similarly, marginal costs below average costs over the entire output range are another sufficient but not necessary condition for subadditive costs.

In the presence of a monopolistic bottleneck, governments may still choose among different regulatory alternatives (cf. Viscusi et al. 2005): laissez-faire, franchise bidding and state ownership are basic choices, yet they are hardly applied so far to the electricity sector, given that they are generally believed to induce either excessive monopoly rents, contractual problems or low efficiency, or several of these problems. Therefore, the focus in the following is rather on the regulatory approaches in place in Europe and other parts of the world with competitive electricity markets.

The key elements of the regulation in place are:

- Non-discriminatory access to the monopolistic bottleneck, i.e. the electricity grid,
- Unbundling between the monopolistic bottleneck and the competitive parts of the sector, notably generation and retailing,
- Price regulation of the monopolistic bottleneck.

These issues are discussed in the next section.

6.1.2 Non-discriminatory Grid Access, Unbundling and Price Regulation

Subsequently, we start by reviewing the key issues related to non-discriminatory grid access and unbundling in Sects. 6.1.2.1 and 6.1.2.2. Then we discuss the basic alternatives of cost-based and incentive-based price regulation for the monopolistic bottleneck in Sect. 6.1.2.3. This has to be complemented by discussing further regulatory challenges, like quality regulation in Sect. 6.1.3. The question of grid tariff structures may be addressed directly by regulation, but may also be left at least partly to the discretion of grid operators. It is therefore left to Sect. 6.1.4.

6.1.2.1 Non-discriminatory Grid Access

As discussed above, the electricity grid is a monopolistic bottleneck where competition does not work. In the past, economists considered the vertically integrated electricity sector to be a natural monopoly as a whole. Yet, since the 1970s, first economists and then practitioners started to distinguish between the grid as a monopolistic bottleneck and the remaining segments of the industry such as generation, trading and retail services (cf. Joskow 2007). Those do not exhibit subadditive cost structures, and thus, competition is possible in these fields.

Yet, electricity generators need the electricity grid to deliver their product to their customers. Hence, competition is only possible in the field of generation, if access to the grid is possible for all competitors – and the same holds for trading and retail services. Non-discriminatory access to the monopolistic bottleneck is thus a fundamental prerequisite for functioning electricity markets. This has to be stipulated by law (e.g. the German energy act EnWG) and is to be enforced by a regulatory authority (such as the Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway (BNetzA) in Germany). This regulator will ensure non-discriminatory pricing by grid companies and also the absence of non-price discrimination measures (e.g. restrictions in grid access). In European markets, a key element for operationalising non-discriminatory grid access are the so-called balancing groups. These virtual entities allow grid users (energy companies, larger industrial facilities, etc.) to bundle their generation and sales activities within one grid area. The grid operator then asks the grid users to provide schedules for infeed and offtake within its grid area so that it can manage the grid accordingly (see Sect. 8.4).

6.1.2.2 Variants of Unbundling

Besides the absence of privileged grid access, grid operators should also not be able to provide other types of advantages to some other (related) player exposed to competition. To prevent such abuse of a monopoly position, a separation between grid and competitive businesses is necessary. This separation imposed by regulation is commonly labelled “unbundling” and in practice, regulators impose different forms of **unbundling**:

Accounting unbundling obliges companies active in both the grid and the competitive part of the sector to keep separate accounts for these activities and, consequently, to have separate balance sheets. This is a minimum requirement for appropriate price regulation.

Informational unbundling includes the requirement that information obtained by a grid operator is not used by the competitive parts of the same company. Otherwise, the company may obtain a competitive advantage compared to its competitors, e.g. through more detailed information about customers (e.g. load profiles). Informational unbundling implies that both business processes and IT systems must be separated between the grid and the competitive business segments.

Functional unbundling (**Management unbundling**) requires that managers in the grid business may not be involved in competitive business segments and vice versa. This prevents information sharing and strategic decision-making in view of overarching company interests.

Legal unbundling is achieved by making any grid operator a separate legal entity (company). As a separate entity, the grid operator has an obligation to keep separate accounts and also informational and functional unbundling are more easily established and monitored.

Ownership unbundling requires that grid businesses and competitive businesses like generation, trading and retail (supply) have different owners. With different owners, all incentives for privileging a competitor should vanish.

The European regulation has gradually increased the requirements for unbundling in the electricity and gas sector. In the first electricity market directive 96/92/EC, only accounting separation between generation, transmission and distribution was required. The so-called acceleration directive 2003/54/EC then required informational, functional and legal unbundling for transmission and distribution grid operators, allowing exemptions to legal unbundling for small distribution grids. The political and regulatory debate around the third internal energy market package (notably directive 2009/72/EC) focused on whether ownership unbundling should be enforced for transmission system operators (TSOs). As a result of a compromise, the European Commission did not impose full ownership unbundling throughout Europe. Instead, directive 2009/72/EC provides two alternative forms of unbundling, labelled Independent System Operator (ISO) and Independent Transmission Operator (ITO), which allow integrated companies like EDF, E.ON or RWE, to hold shares in a transmission system operator, yet with very limited possibilities to influence managerial decisions in the TSOs, which became own legal entities. In practice, many formerly vertically integrated utilities like E. ON and Vattenfall (in Germany) have sold off their transmission assets in the years around 2010.

An issue that has not received comparable attention is whether the distribution grid should be fully unbundled from the retail business. Both the presumed larger market power of generation companies compared to retail companies (also called suppliers) and the lower transaction costs (relative to the company size) may be invoked to justify the focus on transmission unbundling. Yet, with the advent of distributed generation and smart grids, the neutrality of distribution grid operators may become more important. On the other side, coordination between generation, grid and consumption at a decentralised level becomes more important (cf. e.g. Friedrichsen 2015). These trade-offs require further investigation in view of a grid regulation that enables the transition to a sustainable low-carbon energy system (see also Sect. 12.3).

The challenge of regulation at a European scale is further complicated by the broad variety of existing company structures. For example in France, one major distribution grid operator (Enedis, subsidiary of EDF) serves more than 95% of all customers. In Germany, by contrast, there are more than 800 different distribution grid companies of very unequal size.

6.1.2.3 Price Regulation: Cost-of-Service Versus Incentive Regulation

The absence of a well-functioning market for grid services leads to the need to supervise or fix the rates charged by network operators to their customers. The large variety of proposed regulation schemes may be broadly classified into two categories:

- **Cost-of-service regulation and**
- **Incentive regulation**, also called **performance-based regulation**.

These categories may be further subdivided, as shown in Fig. 6.1.

As indicated in Fig. 6.1, the price regulation should be complemented by a (direct or indirect) quality regulation. These and further aspects relevant for practical implementation will be analysed after discussing the basic concepts in the subsequent sections.

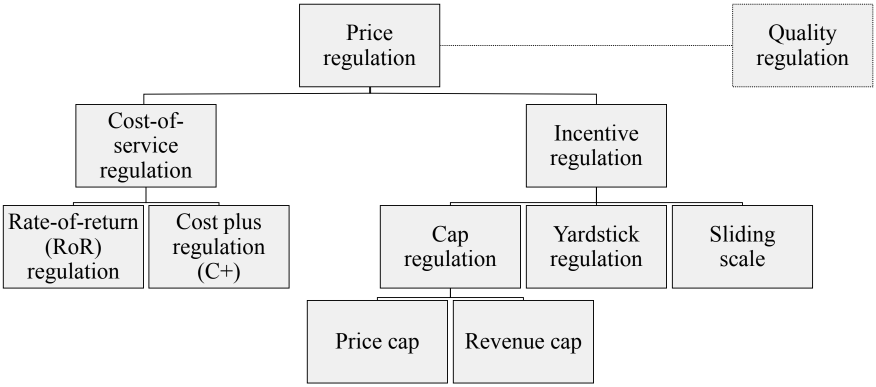


Fig. 6.1 Types of price regulations

Cost-of-Service Regulation

In a cost-of-service regulation, the cost incurred by the provider for delivering the service is taken as the basis for the price regulation by the regulator. Two variants may be distinguished:

Cost-plus regulation, in its simplest form, determines the allowable revenues R_t^{C+} based on the sum of (expected) operational expenditure O_t and capital expenditure based on depreciation C_t^{capex}. On top of this sum, a profit margin a is conceded:

$$R_t^{C+} = \left(O_t + C_t^{\text{capex}}\right) \cdot (1 + a). \tag{6.3}$$

The cost-plus approach was rather popular in Europe before liberalisation but has been mostly replaced by incentive regulation schemes. Operational expenditure includes all recurring expenses, e.g. for staff, raw materials, operation inputs such as energy or insurances and maintenance. It is frequently abbreviated as OPEX, and correspondingly, for capital expenditure the term of CAPEX has been coined. These include the expenses for machinery, equipment like poles and lines, buildings and computer systems. Regulation thereby usually does not consider the cash flows but the annual depreciation. The sum of both is then labelled TOTEX. The cost of capital (interest payment for debt and return on equity) is usually implicitly considered through the choice of the profit margin.

Rate of return regulation has been traditionally employed in the USA. Thereby, the focus is more on the recovery of the capital cost. With a pre-specified rate of return r on the capital employed K_t, the allowable revenues are computed as:

$$R_t^{\text{RoR}} = O_t + C_t^{\text{capex}} + r \cdot K_t. \tag{6.4}$$

Both regulation schemes do not provide an incentive to reduce costs since increases in costs can be passed through to the customers. Even worse, there are incentives for the grid operators to inflate their capital base in the case of a rate of return regulation with a guaranteed return exceeding the actual cost of capital of the operator. This effect has been first described in detail by Averch and Johnson (1962). For the cost-plus regulation, a similar result holds: increases in operational expenses directly increase the operator's profit. Excessive use of capital and correspondingly higher depreciations are also advantageous for the operator in the chosen formulation if the marginal costs of capital are smaller than the profit margin multiplied by the depreciation rate.

Therefore, cost-of-service regulations will only provide welfare-optimal results if the regulator has perfect knowledge both of the actual cost and the cost reduction possibilities of the grid operator. Then she will discard any expenditure on capital and operational goods that are not "used and useful". In reality, information asymmetry prevails and the grid operators have more knowledge than the regulator. By imperfectly applying the "used and useful" criterion, the regulator will aim to limit the operator's profits, yet still no incentives for dynamic efficiency improvements are provided.

Incentive or Performance-Based Regulation

Given the limitations of cost-of-service regulation, economic theory and regulatory practice have come up in the 1980s with the concept of incentive regulation. Littlechild (1983) proposed price cap regulation for the to be privatised British Telecommunications (BT), but the concept was soon after also applied to electricity and gas infrastructures and other sectors.

Price cap regulation in its basic form may be described by the formula:

$$p_t = p_0 \cdot (1 + \mathrm{RPI}_t - (t - t_0) \cdot X). \tag{6.5}$$

Accordingly, the maximum price p_t in year t is determined based on a starting price p_0 in year t_0 adjusted for the inflation through the retail price index RPI_t (measured relative to t_0) and an expected annual productivity gain, labelled X in the original literature. The scheme is therefore also known as RPI-X-regulation.

The starting price p_0 is usually determined based on cost information, whereas expectations regarding inflation and productivity gains are often derived from mere extrapolations of historical statistics.

Revenue cap regulation is a variant of the price cap approach for multi-product firms. Electricity grid operators are multi-product firms because they provide grid connection to customers at different grid levels. Yet, the provision of these services is done using a (partly) common infrastructure. Therefore, it is both simpler in application and more reflective of cost structures if not the prices of single products are regulated, but instead the overall revenue. This leads to the basic formulation of a revenue cap as follows:

$$R_t^{\mathrm{RCp}} = R_0 \cdot (1 + \mathrm{RPI}_t - (t - t_0) \cdot X). \tag{6.6}$$

The essential advantage of such a regulation scheme is that the revenues of the grid operators in year t are decoupled from their cost during the same period. Hence, the grid operator has no incentive to inflate their costs, rather they have a strong incentive to minimise costs, given that their profit will correspond to the difference between the fixed revenue cap and their actual costs. Therefore, the term “incentive regulation” has been coined for this type of regulation.

Both practical experience and more advanced theoretical reasoning however indicate that the application of this regulatory recipe is not that straightforward: given that estimates of future productivity gains are subject to considerable uncertainty, the application of a specific price or revenue cap formula should be limited to a pre-specified regulation period. Otherwise, there is a substantial risk that the operator obtains huge excess profits if the possible productivity gains are underestimated. But also the opposite risk has to be taken seriously, namely that the operator ends up in financial distress if ex-ante estimates of productivity gains turn out to be too optimistic.

Therefore, the duration of a regulation period is usually set to three to five years in practice. A new reference price or revenue level is set for the next regulation period based on cost considerations. This may however induce a so-called **ratchet effect**: since the regulated entities, i.e. the grid operators, anticipate that any cost decrease they achieve during one regulation period will induce lower (i.e. more ambitious) reference levels in the following regulation periods, they will not engage in cost reductions that require continued additional efforts. The shorter the regulation period, the more pronounced is this ratchet effect. Consequently, longer regulation periods provide more substantial efficiency incentives – at the expense of higher risks as stated above.

With any limited duration of a regulation period, the necessity of setting a new reference level also implies problems similar to those incurred in cost-of-service regulations. Notably, the criterion of “used and useful” expenditures will have to be applied by the regulator who will have to struggle with the problems of asymmetric information.

Yardstick competition as proposed first by Shleifer (1985) is a possibility to avoid some of the difficulties of conventional revenue or price cap approaches. The revenue level is thereby set by using an average cost value of similar firms as a reference. This mechanism also decouples the revenue level of the firm from its own cost level and thus provides incentives for continued cost reduction efforts. Yet, it is only applicable if an adequate number of sufficiently similar firms is available for comparison purposes. One may argue whether Germany’s more than 800 distribution grid operators are sufficiently homogenous to enable an adequate comparison. But for one single TSO per country (e.g. France), this approach will not work.

Sliding scale regulation is a possibility to combine cost-based regulation with incentive schemes (cf. e.g. Schmalensee 1989; Laffont and Tirole 1993; Lyon 1996). Its basic linear version may be seen as a weighted combination of a cost-based and an incentive-based scheme. Formally, this may be written

$$R_t^{\mathrm{SSc}} = (1-\alpha)\cdot R_t^{C+} + \alpha \cdot R_t^{\mathrm{RCp}}. \tag{6.7}$$

The sliding scale parameter α then gives the weight of the incentive-based scheme in the overall revenues. The higher its value, the more emphasis is laid on incentivising the grid operator, since cost overruns will only be recovered to a lower extent through revenue increases.

6.1.3 Practical Challenges of Performance-Based Regulation

There are numerous aspects in the application of the aforementioned concepts of incentive regulation. Subsequently, we limit the discussion to three major points: the issue of heterogeneous network operators, the question which parts of the costs should be subject to performance-based regulation and the need for a quality regulation.

6.1.3.1 Heterogeneity of Network Operators and Benchmarking

In principle, multiple network operators may contribute to alleviating the regulator's problem of asymmetric information. Indeed, the regulator may use the multiplicity of observations on different grid operators to derive benchmark performance measures. This is done in a rather straightforward way in the yardstick competition approach described above. There the average of (similar) grid operators is taken as a reference value.

Beyond that, more sophisticated approaches may be applied that consider that not all grid operators are alike. Usually, the term **benchmarking** is used to designate such methods.

Benchmarking is a method for coping with grid operators' heterogeneity by adjusting the cost for some observable characteristics of the different grid operators. With the help of benchmarking, inefficiencies in the cost structures of grid operators are to be identified. Hence, it may be used in conjunction with yardstick competition to set an adjusted revenue level, as is the case in Norway since 2007. Or it may also be used in connection with revenue cap approaches to determine firm-specific paths for productivity gains – this has been the practice in Germany since 2009. Different statistical techniques may be used to arrive at the benchmarks, among which the most frequently used are **Data Envelope Analysis** (DEA) and **Stochastic Frontier Analysis** (SFA). Whereas DEA is a nonparametric approach related to optimisation problems, SFA is more related to regression methods (cf. e.g. Bogetoft and Otto 2011).

European regulators have repeatedly applied these techniques to electricity grid operators, yet results have been heavily criticised, primarily by practitioners (cf. also Kuosmanen et al. 2013). One of the most frequent criticisms is that the results may be strongly impacted by outliers and not robust against specification errors. For example, the German regulator has responded to these criticisms by taking a best-of-four approach when assigning an efficiency level to each grid operator. The four alternative benchmark values are obtained by combining the DEA and SFA methods with two different cost bases (see next section).

6.1.3.2 Cost Base for Regulation

Applying the aforementioned methods raises two additional major questions in practice that have not been discussed so far. Both are related to the basic concept of the cost of a firm.

The first key issue is whether the incentive regulation schemes discussed previously should be applied to the total cost of a firm, i.e. TOTEX (see Sect. 6.1.2.3 for terminology), or only to the variable part, i.e. OPEX.

From a welfare perspective, the consideration of the total cost is the more obvious choice. Yet, companies tend to argue that they may not be able to influence their capital expenditure in short to medium run, given the long lifetime of assets. Thus, a comparison based on TOTEX may disadvantage some firms over a longer period – although this may be softened in a revenue cap scheme by requiring only limited productivity gains per year for firms that are found to be inefficient in a benchmarking exercise. On the other hand, a pure OPEX-based incentive regulation provides distorting incentives: the network operators will aim to reduce their operating expenditures whereas CAPEX is not touched – in extremis, there will be an excessive substitution of benchmarked OPEX by CAPEX, which are subject to a cost-based regulation.

When including the capital expenditures in the incentive scheme, the definition of these expenditures is a key aspect. But even under other regulatory regimes, the so-called **regulatory asset base** (RAB) is pivotal in regulation. What items are included in that asset base, what lifetimes are assumed for the computation of depreciations and are the capital expenditures derived directly from the depreciation according to the company accounts? The answers to these questions have a major impact on the profitability of the regulated firms, their incentives to invest and the outcomes of the regulatory benchmarking.

6.1.3.3 Quality Regulation

Whereas companies tend to cut costs under a proper incentive regulation scheme, they have an incentive to inflate their expenditures under cost-based regulation (see Sect. 6.1.2.3). Consequently, it is unlikely that the quality of service will be put at risk under cost-based regulation schemes. However, this type of regulation does not guarantee by itself the most efficient use of revenues for upholding the quality of service. Yet, the fear of being blamed for some partial or full blackout is likely to be a strong motivation for maintaining at least some quality level.

In incentive regulations, this motivation may not be strong enough to counterbalance profit maximisation goals. Therefore, an incentive regulation has to be complemented by some form of quality regulation. Usually, three dimensions of quality are distinguished:

- **network reliability**, i.e. the absence of interruptions,
- **network performance**, e.g. in terms of voltage stability or power harmonics,
- **service quality**, e.g. response time to customer requests and complaints.

The emphasis is thereby on the first one. Yet this quality regulation has to face three challenges: the time lag between investment (cuts) and quality impacts, the stochasticity of outage events and the choice of appropriate quality indicators.

As quality indicators for **network reliability**, the indicators defined in Sect. 5.1.4 are commonly used by regulators (cf. CEER 2015):

- SAIFI: System Average Interruption Frequency Index,
- SAIDI: System Average Interruption Duration Index,
- CAIDI: Customer Average Interruption Duration Index,
- ENS: Energy not supplied.

The concepts mainly differ in the weighting of different interruption events affecting various customers (see Sect. 5.1.4). Moreover, the exact definitions applied by different regulators differ in several aspects, notably whether short interruptions (e.g. below 3 min) are counted or whether events attributed to "force majeure" such as extreme weather situations are included.

From a welfare maximisation perspective, energy not supplied should be weighted by the marginal utility it provides to customers. In system planning, this is also known as the **value of lost load** (**VOLL**). Rough estimates range between 2 and 10 €/kWh for industrialised countries. Empirically this value could be determined by measuring the willingness to pay of customers for non-interruption of service.[2] However, this is hardly put into practice so far, since the willingness to pay varies a lot between customer groups and is likely to be time-varying and undoubtedly challenging to measure (cf. Kjolle et al. 2008; Shivakumar et al. 2017).

Suppose ENS (or another quality indicator) is considered to be a sufficiently accurate indicator. Then rational grid operators will make consistent and optimal choices if they are penalised for all service interruptions according to the corresponding willingness to pay of the customers. This is even true in the presence of time lags and stochasticity since full rationality implies anticipation of future penalties with best available probability estimates. Yet assuming such an entirely rational behaviour is rather heroic than realistic, and therefore, quality penalties and rewards are typically restricted to some single-digit percentage of the revenue cap. This is of particular relevance for small grid operators, where **stochasticity** has a higher relative impact due to the absence of levelling effects related to the law of large numbers (cf. Schober et al. 2014).

[2] More precisely, this should be qualified as a (monetary) willingness to accept service interruptions (cf. e.g. Horowitz and McConnell 2003 for the distinction between willingness to accept and willingness to pay).

6.1.4 Principles of Network Pricing

So far, prices for grid services and the corresponding revenues have been discussed either under the assumption that there is just one product delivered by the grid operator (and thus one price) or alternatively with an exclusive focus on the total revenue and not on its split according to different products or customer groups. But in practice, the question of who should be charged for grid usage and on what basis is very relevant since electric grids connect multiple producers with many consumers. A fundamental question is whether exclusively consumers should pay for their usage of the grid infrastructure or whether producers should also be charged.

To derive adequate answers to these questions, Sect. 6.1.4.1 starts by explaining the rather general, formal concept of Ramsey prices as price differentiation between products or customer groups under the assumption that each customer only buys one product. This is followed by a more informal discussion of different products and services delivered to each customer and the corresponding price components in Sect. 6.1.4.2. In Sect. 6.1.4.3, an application of Ramsey pricing to a stylised network is discussed, whereas Sect. 6.1.4.4 discusses the role of different price components on the example of a low-voltage grid, including notably also so-called prosumers. The practical implications for network tariffs, notably focusing on future smart grids, are then outlined in Sect. 6.1.4.5.

6.1.4.1 Ramsey Pricing

The so-called Ramsey rule goes back to the British mathematician Frank Ramsey (1903–1930). A related rule also named after Frank Ramsey is applicable in the context of optimal taxation issues. The common underlying question is how a monopolist should fix prices (or tax rates) to maximise welfare under a given budget constraint.

This problem arises in the case of a grid operator since the basic rule for efficient pricing, namely to set prices equal to marginal costs, does not fulfil the budget constraint. As stated in Sect. 6.1.1, the need for regulation of grid operators arises from the fact that grid costs are subadditive. Even the stronger condition that marginal costs are below the average per-unit cost holds in the case of grid operators. This implies that setting prices equal to marginal costs will result in revenues for grid operators, which do not cover their costs. If there is a single product with a single price, the problem may be depicted as shown in Fig. 6.2. The first-best solution of setting prices equal to marginal cost results in the equilibrium indicated by point F. Yet at the corresponding price p^*, the grid operator incurs losses equivalent to the hashed area, corresponding to the difference between average cost and price multiplied by the equilibrium quantity.

The optimal solution satisfying the budget constraint, i.e. without losses incurred by the grid operator, is given in the one-product case by point S. This **second-best** solution (cf. Sect. 6.2.3.1) corresponds to setting the price equal to average per-unit cost.

If multiple products i (grid services) are sold to different customers, the second-best solution is not that obvious. It may be obtained by maximising the

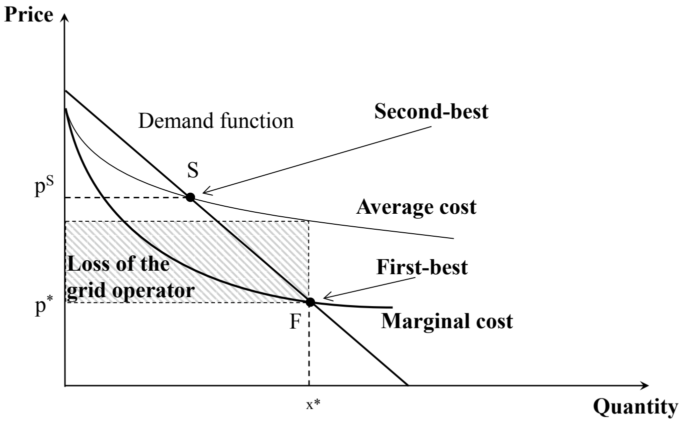


Fig. 6.2 First-best and second-best solutions for grid tariffication

economic surplus (welfare) W under a budget (non-loss) constraint. The economic surplus may be derived as the sum of consumer and producer surpluses over all products. This is equivalent to obtaining the integral willingness to pay for the delivered quantities q_i and deducing joint production cost C:

$$W = \sum_{i=1}^{N} \int_{0}^{q_i} p_i(q_i)\mathrm{d}q_i - C(q_1 \ldots q_i \ldots q_N). \tag{6.8}$$

Thereby the inverse demand functions $p_i(q_i)$ are used to represent the willingness to pay (and thus marginal utility of customers) for the different products.

The budget constraint of the grid operator then corresponds to:

$$\sum_{i=1}^{N} p_i q_i - C(q_1 \ldots q_i \ldots q_N) = 0. \tag{6.9}$$

Using the Lagrange multiplier approach and determining first-order optimality conditions, one obtains after some rearrangements the Ramsey pricing rule:

$$\frac{p_i - c_i(q_1 \ldots q_i \ldots q_N)}{p_i} = \frac{k}{|\varepsilon_i|} \quad \forall i. \tag{6.10}$$

The left-hand side thereby corresponds to the relative markup over the marginal cost $c_i = \partial C / \partial q_i$. And this markup is found to be inversely proportional to the price elasticity of demand $\varepsilon_i = \frac{\partial q_i}{\partial p_i} \cdot \frac{p_i}{q_i}$. The proportionality constant k thereby has to be chosen such that the budget constraint is fulfilled.

According to this Ramsey rule, those costs that are not directly attributable to one product should hence not be distributed equally or proportionally to the customer groups, but rather those should pay most that have the lowest evasion or substitution tendency. This rule minimises the so-called dead-weight losses associated with government intervention in markets, although it may result in prices and distribution effects that are perceived as unfair. The implications of applying (or not applying) this rule will be discussed in Sect. 6.1.4.5.

6.1.4.2 Capacity, Energy and Other Prices for Grid Usage

We may push the cost allocation and pricing one step further by arguing that a single customer does not obtain one single service from the grid but rather different services depending on the hour of the day or the year when she or he consumes or produces electricity.

A widely applied principle is that consumers are only charged the costs of the grid voltage levels they are using. This way, a large industrial customer, who is directly connected to the high-voltage grid, will not be charged the cost of the low-voltage grid. Since the electricity a large industrial customer consumes does not transit the low-voltage grid, this has been true in conventional power systems with little small-scale distributed generation, leading to reverse flows in the grids. Households and other small electricity users connected to the low-voltage grid are contrarily charged the cost of the low-voltage grid and a fraction of the costs of higher voltage grid levels – since large proportions of the energy they are using is transported via the higher level grids.

On the other hand, most grid costs depend on the maximum grid capacity needed over the year or even several years since this grid capacity is a key design choice in the grid and determines its investment. The consequence for grid tariffs is that customers should be charged for their contribution to maximum capacity on the one side and for the energy they get transported through the grid on the other side. And if there are costs that are independent of power and energy provided, those should be charged separately in a lump-sum connection fee.

Conventional grid tariffs in many countries around the world are designed in line with these general considerations: they are two- or three-part tariffs including[3]:

- an annual base fee,
- a capacity fee and
- an energy fee.

However, these basic tariff elements need further reflection to obtain an adequate tariff system. An interesting aspect regarding the capacity fee is whether it should be paid based on the individual maximum capacity or according to the contribution to the maximum system load. The first approach is currently implemented in

[3] Also retail prices frequently are made up of the same components as discussed in Sects. 3.1.6 and 7.4.7. This is a consequence of the fact that even in deregulated electricity markets with unbundling, retail contracts mostly also include the payment of the grid charges.

Germany, the second in the UK, and justifications may be given for both. Another aspect that limits the applicability of analytical solutions based on marginal calculus is the frequent indivisibility of investments in the grid – e.g. building a new transmission or distribution line comes at some fix cost per km, almost independently of the actual capacity of the line. A further, very practical aspect has so far been the lack of automatic meter reading systems ("smart meters") in the case of households and other low-voltage grid users. With the conventional electromagnetic meters, the capacity used cannot be measured easily and is therefore frequently not charged. This has led to tariff schemes with relatively high energy fees.

Therefore, a combination of economic optimisation calculus, engineering rules of thumb and considerations of practicability and acceptability is needed to determine an adequate network tariff system.

6.1.4.3 Application: Ramsey Pricing in an Electrical Network

We consider a stylised network with two voltage levels: at the higher level, i.e. the transmission grid, there are three grid nodes with the generators and loads connected as depicted in Fig. 6.3. The lower-level distribution grid is only considered for one of the transmission grid nodes. The key parameters for the grid users are summarised in Table 6.1.

Since the distribution grid is only used by the grid users located in area A2.1, potential costs of this lower-level grid are attributable to the two user groups, solar panels (A2.1) and small and medium consumers (A2.1). Yet the costs of the transmission grid cannot be attributed clearly to any of the grid users since Kirchhoff's laws govern the power flows (see Sect. 5.1.2.1). Moreover, the power flows may vary over time or according to meteorological conditions. This raises the question how to determine appropriate reference capacities. Yet, we assume this question to be solved and consider the reference capacities given in Table 6.1 to reflect properly the grid usage at some reference cost or price level $c_0 = p_0$. This reference cost level may be thought of as some pure energy procurement cost (respectively sales revenue in the case of generators), i.e. based on a wholesale market price disregarding grid costs (see Sect. 7.1). It is as such applied to all grid users in the example. As indicated in Table 6.1, the reference cost level is set at 60 €/MWh. The total grid cost is assumed to be independent of the produced and consumed quantities and equal to 1,000,000 €/h on average. Dividing this cost by the total dimensioning capacity Q, grid costs correspond to 20 €/MW/h.[4]

The application of Ramsey pricing now aims at an optimal distribution of these latter costs to the different grid users. This requires in a first step the reformulation of Eq. (6.10) so that the unknown p_i is expressed as a function of the single unknown variable k:

[4] We use hourly values to keep the numbers simple. Those can easily be transformed into annual numbers by multiplying with the number of hours of a year, i.e. 8760. As the basis of our calculations is an average hour, we also disregard the distinction between capacity, energy and other grid price components as usual in simple models of Ramsey pricing.

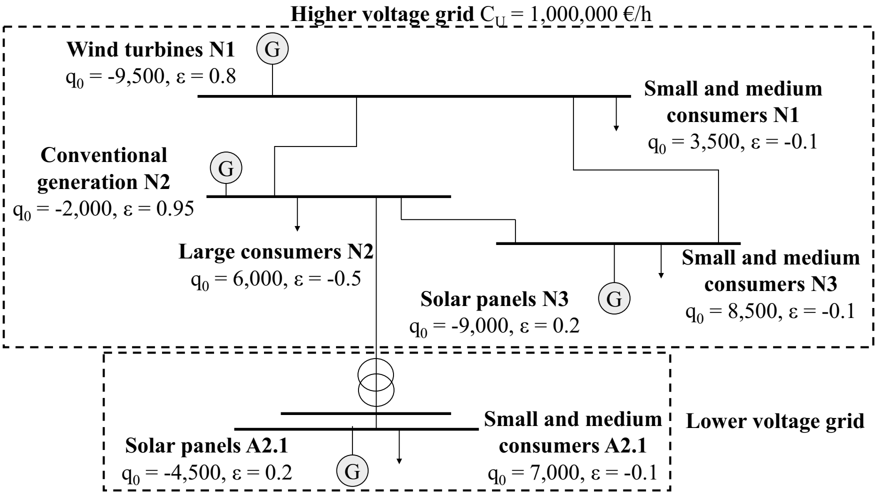


Fig. 6.3 Schematic representation of a two-level grid

Table 6.1 Key characteristics of grid users, grid costs and reference costs

Node higher voltage grid	Area lower voltage grid	Grid user	Grid usage [reference capacity in MW (q_{i0})]	Price elasticity (ε_i)
N1	–	Wind turbines	−9,500	0.8
	–	Small and medium consumers	3,500	−0.1
N2	–	Conventional generation	−2,000	0.95
	–	Large consumers	6,000	−0.5
	A2.1	Solar panels	−4,500	0.2
	A2.1	Small and medium consumers	7,000	−0.1
N3	–	Solar panels	−9,000	0.8
	–	Small and medium consumers	8,500	0.1
All		Total (abs. values)	$Q = 50,000$	
Overall cost/market value	Reference	$c_0 = p_0$ = 60 €/MWh		
	Grid	C = 1,000,000 €/h, i.e. C/Q = 20 €/MW/h		

$$p_i = \frac{c_0}{1 + \frac{k}{\varepsilon_i}} \quad \forall i. \tag{6.11}$$

Thereby we have used the substitution $|\varepsilon_i| = -\varepsilon_i$ which is valid for consumers who (mostly) have negative price elasticities (see Sect. 3.1.4). Yet with this reformulation, the approach is also valid for producers where grid fees would reduce market revenues. Additionally, the dependency of the quantities q_i on the prices p_i needs to be described. Here an iso-elastic formulation is assumed:

$$q_i = q_{i0}\left(\frac{p_i}{p_0}\right)^{\varepsilon_i} \quad \forall i. \tag{6.12}$$

The reference prices p_0 are set to the same value c_0 as discussed before. Yet the revenue of the grid operator from customer i is only based on the difference between the price p_i and the reference price p_0, as the reference price reflects wholesale costs/revenues. Hence the necessary condition for grid revenue adequacy becomes after inserting the two previous expressions and rearranging:

$$\sum_{i=1}^{n}(p_i - p_0)q_i = \sum_{i=1}^{n}\left(\frac{1}{1 + \frac{k}{\varepsilon_i}} - 1\right)\left(\frac{1}{1 + \frac{k}{\varepsilon_i}}\right)^{\varepsilon_i} p_0 q_{i0} = C. \tag{6.13}$$

This is now a single equation for the single unknown k. It cannot be solved analytically, yet a numerical solution is straightforward using a spreadsheet program like Microsoft Excel or another computation software. This is true since each term of the sum on the left side is found to be monotonously increasing with k, as long as k is positive and strictly smaller than the absolute value of the smallest negative price elasticity ε_i. Hence, the sum increases monotonously from (close to) zero to infinity and there is a single optimal value k. The corresponding calculations are implemented in the spreadsheet *RamseyPricing.xlsx* contained in the electronic appendix to this chapter. Using the function "search target value", the user can determine the optimal solution given in Table 6.2.

From the results, it is obvious that grid users with low price elasticities pay much higher markups than those with high price sensitivity – or producers must accept much higher discounts. In our example, the small and medium consumers bear a 42.5% surcharge, whereas it is only 8.5% for the large consumers who have a five-time higher price elasticity. Wind and conventional producers in our example are even more price-sensitive and therefore only face discounts of 5.3% and 4.5%, respectively. As stated in Eq. (6.10), the markups and discounts are directly proportional to the inverse of the price elasticities.

In this vein, it hence seems justifiable that producers are (almost) exempt from grid fees.[5] Yet even for consumers, the practical application of Ramsey pricing faces serious difficulties: first, the price elasticities are generally not known and not

[5] Note that we have considered wholesale market prices to be exogenously given, independent of generator decisions. This is obviously not true in reality. Yet an endogenous treatment of market

Table 6.2 Results of Ramsey pricing for the application example

Node higher voltage grid	Area lower voltage grid	Grid user	Final network price (in €/MW/h)	Relative markup (%)
N1	–	Wind turbines	57.0	−5.3
	–	Small and medium consumers	104.4	42.5
N2	–	Conventional generation	57.4	−4.5
	–	Large consumers	65.6	8.5
	A2.1	Solar panels	49.5	−21.3
	A2.1	Small and medium consumers	104.4	42.5
N3	–	Solar panels	49.5	−21.3
	–	Small and medium consumers	104.4	42.5
Calibration parameter		$k = 0.04255$		

easily measurable. This is particularly true as the relevant elasticities regarding irreversible and long-term investment in grid infrastructure are long-run price elasticities. Their estimation requires long price series, which are prone to structural breaks. The second issue is (perceived) fairness. This is not an issue in the framework of mainstream economics since distributional effects should be handled separately from pricing issues according to the fundamental theorems of welfare economics (see the introduction to Sect. 6.1). Yet as grid operators are either state-owned or regulated utilities, the decision on grid tariffs is also a political one. In particular in the presence of very limited empirical evidence, the case for Ramsey pricing as a form of discriminatory pricing is challenging to defend in the public debate.

Our example of Ramsey pricing could be extended to the lower grid level contained in Fig. 6.4. Yet, we leave it aside to limit complexity and focus in the following subsection on the challenges arising in low-voltage grids, notably in connection with so-called prosumers.

6.1.4.4 Application: Network Tariffication in Low-Voltage Grids

The heterogeneity of grid users in a future **smart grid** further complicates identifying efficient and cost-reflective prices as those discussed previously under the concept of **Ramsey pricing**. Notably, the presence of prosumers raises new challenges.

equilibrium would require a substantial extension of the classical Ramsey pricing model, making use of the market equilibrium models introduced in Sect. 7.1.

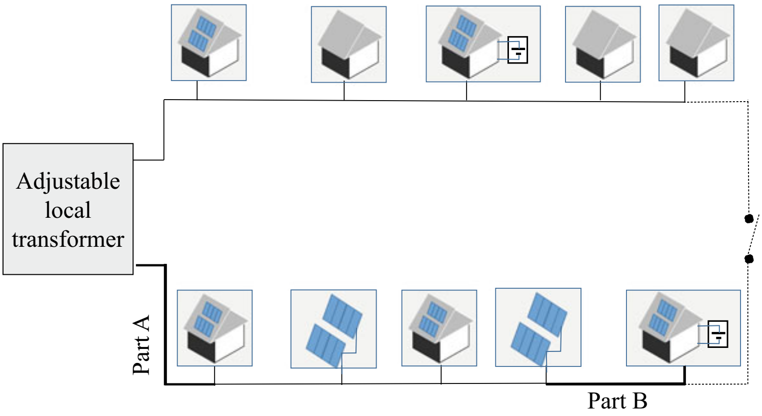


Fig. 6.4 Schematic representation of a low-voltage grid

Prosumers are a rather new type of grid users – the term has been coined to designate customers that not only consume electricity but also operate on-site generation facilities, notably PV or CHP systems. Moreover, they may have some flexibility in their generation and consumption pattern, e.g. through the use of storage possibilities.

To illustrate the arising challenges for network tariffication, we consider an exemplary low-voltage grid as depicted in Fig. 6.4.

This grid consists of

1. One transformer as a connection point to the higher voltage grid levels
2. Several consumption points (households)
3. Several production units (PV-panels) associated with households
4. Several production units (PV-panels) not associated with households
5. Some storage devices (batteries) associated with production units.

The grid is constructed as a closed loop, but the switch at the end of the line is open in standard operation mode. It is only closed if there is an interruption somewhere else in the circuit, so that **N-1 secure** operation is possible. In what follows, we yet disregard faults and other contingencies and focus on regular operation.

Without a complete formal treatment, we subsequently aim to explain the challenges of finding the correct "prices for the grid services". We do so by starting with simple configurations and adding complexity step by step.

(a) **Transformer plus consumers**: this is a simplistic version of a grid to start dealing with the question: How should we fix the grid tariff in an efficient and cost-reflective way? The answer begins by stating that the load peak in the system mainly drives the transformer size. Hence, each consumption point should pay a share of the transformer cost proportional to the (expected) contribution to the transformer peak load. In the case of consumers with similar consumption profiles, the contribution to the transformer peak load will be proportional to the households' own peak load. Hence, there should be a (part of the) capacity charge paid per kW of (non-coincident) peak load in the household. If the households have different consumption profiles, this is yet no longer the optimal pricing scheme. Then, the capacity charge should be paid on the basis of the (average) contribution to the system peak load, i.e. the **coincident peak load** (cf. Sect. 3.1.6).

As long as consumption is purely random and effectively not controlled by the consumer nor the consumer can avoid the grid charge through relocation, the height and structure of the capacity charges do not impact the economic efficiency – in the Ramsey formula, the price elasticity is equal to zero. Yet as soon as consumers may react, e.g. by rescheduling load, capacity charges have to reflect contributions to system load peaks.

(b) **Energy instead of capacity pricing**: traditional grid tariffs in the low-voltage grid often charge the users with an energy fee instead of a capacity fee – not least since traditional meters only measure the energy flowing through the wire and do not keep track of load maxima or loads per time interval. Charging the customers proportional to their energy consumption instead of their contribution to system load again makes no difference in efficiency, as long as the consumers have no capability to react. Such a pricing then also does not affect overall costs (and thus economic efficiency) even if it may be perceived as unfair, i.e. the resulting distribution of costs is not in line with preconceptions of equitable cost sharing. But suppose households have opportunities to adjust their consumption. In that case, a time-independent energy price does not provide an adequate signal for doing so – as it does not incentivise consumers to shift their load away from the peak load period.

(c) **Adding connection lines**: if the previous system consisting of a transformer and several consumption points is now complemented by the (underground or overhead) lines necessary to supply the individual households, the question arises of who should pay for these lines. In Fig. 6.4, part A of the line serves all households of this feeder, while part B is only needed in regular operation to supply one household. In the logic of cost-reflective pricing, the costs of part A should hence be borne by all households, those of part B yet only by the household attached to it. This discriminatory pricing would be efficient since it would put more substantial incentives on capacity reductions by the consumers beyond part B of the line, where line upgrades are more costly for the customers. Yet this perspective is only correct if we consider the location of the transformer as fixed. This is undoubtedly true in the short to medium term but may not hold in the long run. Then also, the justification for discriminatory pricing in terms of economic efficiency gets weaker.

(d) **Addition of (renewable) generation units**: What changes in terms of efficient grid pricing if we now include some generation units in the system? A differentiation is necessary to answer this question:

If the generation follows the load and reduces the peak load on the transformer and the lines, it contributes to grid cost savings. It may then earn remuneration from the grid – corresponding to the grid capacity price multiplied by the (expected) reduction in grid peak load.

Another situation occurs if the generation is driven by natural variability as is the case for PV and wind. In that case, it is unlikely that the distributed generation will predictably reduce the load peak.[6]

In grids with high shares of renewables, even the power flow may be reversed during periods with high renewable infeed. As long as the transformer capacity is not exceeded (and voltage limits are not violated) by the reverse flows, grid costs are not driven by the renewable infeed and generation should therefore not be allocated any grid costs – except for a payment covering the transportation losses in the grid, which we omitted so far.

As soon as the transformer (or another critical element) is yet more heavily used during the period of maximum reverse flow than in the peak load situations, the long-run costs of the grid are driven by the maximum reverse flow. Hence cost-reflective pricing would imply charging the generators in the grid for their (expected) contribution to peak reverse flow and conversely paying the consumers for the relief they provide in that situation.

(e) **Curtailment of (renewable) generation**: the situation described at the end of the previous section (reverse flows determine capacities of grid elements) would yet provoke adaption reactions from renewable producers: instead of paying high grid capacity charges they would curtail their production in the peak reverse flow situations, as long as the revenues from sales on the wholesale market (or under a renewable support scheme) are lower than the capacity charges. In the optimum, capacity charges would then be paid by consumers and generators with higher per kWh charges for the consumers. This is a consequence of their lower price elasticity (i.e. their higher willingness to pay for unlimited grid access) and the Ramsey pricing rule (see Sect. 6.1.4.3). Obviously, also the grid should be extended as long as the aggregate willingness to pay exceeds the grid construction costs.

(f) **Addition of storage or other flexibility providers**: already the addition of limited (curtailment) flexibility in the previous step has led to differentiated grid charges applicable for different time segments (peak load and peak reverse flows). The situation becomes even more complicated if storage is used to (partly) substitute grid investments. Three challenges arise:

1. multi-period variable grid charges
2. lumpiness of grid investment
3. strong location-dependency of grid-related storage value.

[6] Grids with high air-conditioning and cooling loads may be an exception to the rule, if the load peak coincides with periods of high solar radiation and hence high PV infeed. The so-called duck curve observed in California (cf. e.g. IEA 2019) yet suggests that even in such grids the coincidence is far from being perfect: in California, the peak in electricity consumption occurs in the early evening, when people return from work and turn on air-conditioning and other appliances at home.

The first point is the relatively straightforward generalisation of the point made previously on simultaneous scarcity pricing in different periods. The second one, by contrast, emphasises that the marginal calculus that may be used to underpin the first point is in reality hardly appropriate, given that grid operators invest in discrete pieces of hardware. If the investment is not necessary, the marginal value of substitutes is (close to) zero. If the investment is unavoidable, the marginal value becomes infinite. However, there may be a finite nonzero value for alternative lumpy investments (cf. Böcker et al. 2018). Finally, the third point emphasises that these considerations will be strongly dependent on the exact grid topology and congestion situation. These considerations are related to so-called nodal pricing which is discussed in Sect. 10.8, albeit nodal pricing usually focuses on short-run marginal cost.

6.1.4.5 Implications for Practical Grid Tariff Structures

The general principles and applications discussed in the preceding sections indicate elements of current and future grid tariffication, although no simple recipe for optimal tariff structures may be directly derived. Key practical implications to be retained include notably:

Consumers generally have a lower price elasticity for electricity than producers – this provides a rationale for charging mainly (or even exclusively) consumers with the markup over short-run marginal costs.

Producers should nevertheless be charged at least the marginal grid cost they are causing. This includes the direct connection costs (so-called shallow connection charges) but may also include indirect costs caused by new generation in other parts of the network, e.g. for grid expansion ("deep connection charges"). This leads to a locational component in grid tariffs that has been in place in the UK for more than a decade as a so-called G component. The calculation of such a component should be based on the long-run marginal costs for grid operation, renewal and expansion. However, such a grid tariff based on long-term costs provides only imperfect signals in terms of short-term congestion management (see Sect. 10.6.2).

For *electricity storage*, e.g. pumped hydropower stations or battery storage, similar considerations apply as for generators. They are rather price-sensitive and should therefore be charged mostly the direct and indirect connection cost.

For *prosumers* as grid users who not only consume electricity but also operate on-site generation facilities, notably PV or CHP systems, the design of grid tariffs is of particular relevance: high energy-related charges (see Sect. 6.1.4.2) lead to strong incentives for self-generation, which effective grid cost savings might not justify.

The increasing share of distributed generation and prosumers therefore challenges the existing grid tariff structures. This is particularly true in the ongoing transformation of the electricity system. If there is considerable uncertainty about the future generation mix and the corresponding costs, long-run marginal grid costs are also uncertain. Setting adequate incentives here without jeopardising innovation through frequent rule changes is undoubtedly a key challenge for regulation and grid management in future (see also Sect. 12.3).

6.2 Environmental Effects and Environmental Policy

Energy conversion leads to different environmental impacts, especially when transforming primary energy carriers like coal, gas, oil or uranium. The magnitude of these problems strongly depends on the technologies and energy carriers used. As long as the originator of the environmental impact does not have to bear the costs related to this impact, he will not take these costs into account. This constitutes a form of market failure that has to be addressed by governmental intervention, i.e. by the implementation of different policy instruments.

Section 6.2.1 addresses the problem of externalities. Different emissions caused by energy conversion processes based on fossil fuels, the corresponding environmental impacts and emission reduction technologies are discussed in detail (Sect. 6.2.2). Section 6.2.3 then reviews different policy instruments, with more specific considerations in Sect. 6.2.4 for climate change.

6.2.1 Externalities

In the market-oriented perspective of mainstream economics, the effects of economic activities on parties not involved in the underlying business transaction are called **externalities**. These externalities can be negative or positive for the third parties and are not compensated by the responsible entity.[7] Therefore, externalities lead to inefficient market outcomes (see Sect. 6.1.1) and induce market failures.

In the case of emissions from energy production, primarily negative externalities arise. As long as the producers of emissions do not have to bear the costs from these externalities, they will not consider them within their production process. This might lead to a situation where too many emissions are produced, a situation not being the social optimum. Theoretically, the emissions should be reduced to the economic optimum of emission reduction costs and damage costs, which will usually not be a reduction of the emissions to their minimum. From Fig. 6.5 it can be seen that the optimal emission level is e^* and the corresponding marginal abatement costs and marginal damage costs are t^* (for the interpretation of t^* see Sect. 6.2.3.1).

If negative externalities are monetarised, external costs are obtained. The costs are called "external", because they are not reflected in the market prices of the corresponding business transactions. Therefore, the marginal private production costs are below the marginal social production costs, which include private and external costs, leading to the dilemma (or market failure) that too many goods with negative externalities (x_0) are produced in the market equilibrium and the corresponding price (p_0) is too low (see Fig. 6.6).

Externalities can also be explained with the help of the theory of **public goods** (cf. e.g. Feess and Seeliger 2013, pp. 35–41). In general, goods can be differentiated with the help of the criteria rivalry and excludability into private goods, club goods,

[7] The "tragedy of the commons", i.e. the overgrazing of common land, is described in a seminal paper by Hardin (1968).

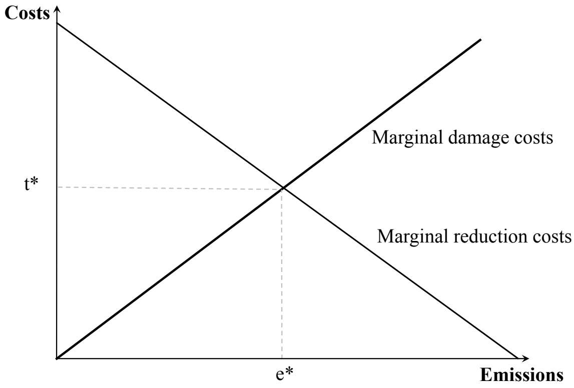


Fig. 6.5 Marginal damage costs versus marginal reduction costs. *Source* Own illustration based on Perman et al. (2003, p. 173)

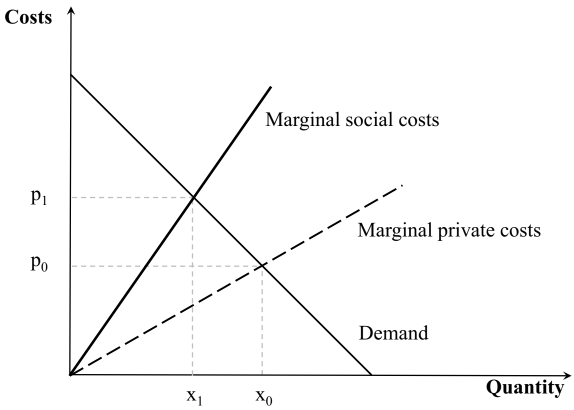


Fig. 6.6 Marginal private versus marginal social costs of production. *Source* Own illustration based on Fritsch (2018, p. 113)

common-pool resources and public goods (see Table 6.3). Whereas the owner of a private good enjoys private property rights, everyone can benefit from public goods. Many environmental goods (e.g. clean air) are seen as public goods, satisfying the corresponding constitutive criteria: non-rivalry and non-excludability. Sometimes, the extensive use of these goods already limits the benefit for one consumer by another. In such a case, the criteria of non-rivalry is no longer fulfilled leading to a classification of the corresponding environmental good as a common-pool resource. Public goods and common-pool resources lead to the so-called free-rider problem because one can benefit from this good even without paying for it. Therefore, there is hardly any incentive for firms or households to provide public goods, and the opposite is true in the case of so-called public bads

Table 6.3 Differentiation of goods according to the criteria rivalry and excludability

	Rivalrous	Non-rivalrous
Excludable	Private good, e.g. ice cream	Club good, e.g. subscription television
Non-excludable	Common-pool resource, e.g. fishery outside territorial waters	Public good, e.g. lighthouse

Source Own illustration based on Perman et al. (2003, p. 126)

(e.g. polluted air), where incentives are missing to avoid the negative externalities for other parties.

The quantification of external costs and their allocation to an economic activity gives rise to various empirical problems. First, the causal relationship between an economic activity and an externality has to be established. But even nowadays, many negative environmental externalities are not fully understood. In addition, sometimes the causal relationship is difficult to prove, e.g. because only the interaction of different economic activities induces the observed negative externality or because in particular situations additional emissions, caused by production processes, will even reduce some negative externalities. Furthermore, it can be extremely challenging to quantify the damage caused by an externality, e.g. in the case of a changed overall landscape appearance caused by the visual impact of (wind) power plants or overhead transmission lines. Therefore, besides the endeavour to directly estimate the **damage costs** (e.g. the economic losses caused by forest decline or casualties due to acid rain), so-called **avoidance cost** methods, which calculate the costs of alternative measures to avoid negative externalities, e.g. the costs of flue gas cleaning to prevent the precursor emissions of acid rain, are used. Another alternative is to determine the value of a good with the help of preference valuation methods (stated preferences and revealed preferences). The methodology of revealed preferences tries to identify people's preferences[8] by observing their purchasing behaviour, whereas in contingent valuation surveys people have to state their **willingness to pay** or their willingness to accept. Furthermore, the quantification of the external costs has to deal with the problem that frequently some damage will only become evident in the future, e.g. in the case of global warming. Then the question arises at which rate such damage should be discounted.

6.2.2 Emissions, Environmental Impacts and Emission Reduction Technologies

Converting energy carriers, e.g. into electricity and heat, leads to so-called **energy-induced emissions** in contrast to process emissions arising, e.g. in industrial production processes like cement production. By burning fossil fuels, greenhouse gas emissions (notably CO_2) and pollutants (e.g. NO_x) are produced

[8] These preferences will differ from person to person and may vary over time.

Table 6.4 Main emissions from different fossil fuels

	SO_2	NO_x	CO	VOC[a]	Particulates	CO_2
Coal	X	X	X	X	X	X
Oil	X	X	X	X	X	X
Natural gas		X	X	X		X

[a] Sometimes VOCs (Volatile Organic Compounds) are differentiated into methane (CH_4) and the remaining non-methane volatile organic compounds (NMVOC)

(so-called primary pollutants; see Table 6.4), which might lead to environmental impacts.

Whereas such air pollutants and greenhouse gases are not emitted when using nuclear energy, long-lived radioactive waste is produced in the normal operation mode of nuclear power plants. Radioactive contamination (e.g. by the spent nuclear fuels and parts of the reactor) constitutes a threat to ecosystems as radionuclides are carcinogenic. The produced nuclear waste has to be securely stored for thousands of years to avoid exposure to radionuclides.

The use of renewable energies also has some environmental effects, which seem to be of minor importance compared to those of fossil fuels and nuclear energy. Nevertheless, the installation of renewables, like hydropower plants, wind power plants, biomass power plants, solar thermal power plants and ground-mounted PV might lead to negative impacts on the natural landscapes (visual impact) and the use of land and water.[9] Besides the indirect emissions from the construction, hydropower plants may result in problems concerning fish migration and the ecosystems located on both sides of the dams (cf. Kaltschmitt and Jorde 2007, pp. 378–383). Wind power plants produce emissions of infrasonic noise and can be a threat to birds and bats (cf. Kaltschmitt et al. 2007, pp. 343–348).

In contrast, the following sections will focus on emissions and environmental impacts caused by burning fossil fuels and appropriate emission reduction technologies.

6.2.2.1 Emissions from Burning Fossil Fuels

The use of fossil fuels is inevitably linked to oxidation of carbon and accordingly to the emission of the greenhouse gas carbon dioxide (CO_2). However, the **CO_2 emission factors** of different fossil fuels differ. Even within the same fuel, substantial variations may arise depending on the fuel provenance and variety. In Fig. 6.7, some average CO_2 emission factors in kg CO_2 per kWh energy content are shown for different fossil fuels (cf. Juhrich 2016, pp. 45–47). The different emission factors already illustrate that switching to fuels with lower CO_2 emission factors or even without any CO_2 emissions (e.g. renewables) can be a promising CO_2 reduction strategy. Yet, in addition to the fuel-specific emission factor, the efficiency of the corresponding production process has to be considered when

[9] Besides, regional effects on climate are known (e.g. due to local reduction of wind speeds), while global impacts are not (yet) identified.

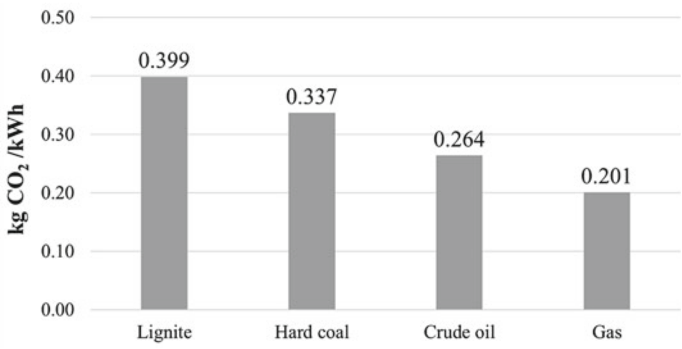


Fig. 6.7 CO_2 emission factors of different fossil fuels. *Source* Own illustration based on data from Juhrich (2016, pp. 45–47)

determining the CO_2 reduction of such a fuel switch (see Sect. 4.3). The average CO_2 emission factor of electricity produced in Germany in 2018 was about 0.474 kg CO_2 per kilowatt-hour electricity produced (kWh_{el}) compared to 0.764 kg CO_2/kWh_{el} in 1990 (cf. Icha et al. 2019). The emission factor of the electricity mix has been declining in Germany for many years due to measures like fuel switching and increasing efficiencies. In this context, it has to be mentioned that the lower the emission factor of an electricity mix is, the fewer emissions are reduced by saving one kilowatt-hour of this electricity, or in other words, the higher the specific CO_2 reduction costs of energy-saving measures (in €/t CO_2) are.

Besides the formation of CO_2 and hydrogen (see Sect. 4.1), also other chemical reactions take place during the combustion of fossil fuels, which result in emissions of air pollutants. The combustion of fossil fuels that contain sulphur leads to the formation of sulphur dioxide (SO_2), which might be oxidised to SO_3:

$$S + O_2 \rightarrow SO_2.$$

In addition, **oxides of nitrogen** (general formula NO_x) arise due to different sources of nitrogen (N) and different NO formation mechanisms. Besides fuel NO and thermal NO emissions, so-called prompt NO emissions can be distinguished. Thermal NO is produced from N_2 of the combustion air when reaction temperatures of about 1300 °C are reached; besides the temperature in the reaction zone also the air ratio and the residence time in the reaction zone have a strong influence on the NO formation (cf. Baumbach 1990, pp. 31–35). The fuel NO results from the nitrogen of the fuels, whereas the prompt NO emissions are produced through incomplete combustion (cf. Tan 2014, pp. 211–216). In a simplified way, the share of the different NO building mechanisms in total NO_2 emissions is shown in Fig. 6.8.

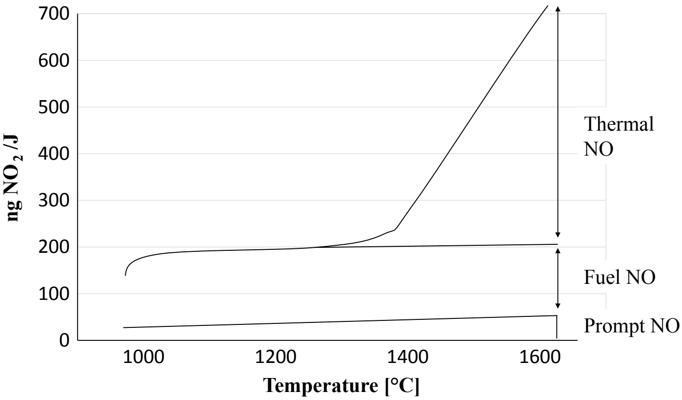


Fig. 6.8 Prompt NO, fuel NO and thermal NO depending on the temperature. *Source* Own illustration based on Hupa et al. (1989, p. 1497)

NO is oxidised to nitrogen oxide according to the following reactions:

$$NO + {}^{1}\!/_{2}O_2 \rightarrow NO_2.$$

Furthermore, trace elements like mercury may be emitted from combustion processes depending on the composition of the fuels via boiler ash and fly ash. Also, particulates [or particulate matter (PM)] are relevant pollutants. These are small solid or liquid particles, which can be considered as dust. Particulates are differentiated according to their diameters, e.g. PM_{10} comprise particles with a diameter of less than or equal to 10 μm, $PM_{2.5}$ accordingly with less than or equal to 2.5 μm.

6.2.2.2 Environmental Impacts

Emissions from burning fossil fuels can lead to different environmental problems and threats to human health. In the 1970s and 1980s, acid rain as a consequence of NO_x and SO_x emissions was the central ecological problem in many parts of the world. This has changed entirely, and nowadays, limiting climate change is on the top of the environmental agenda. Whereas the emissions of NO_x and SO_x have been reduced in many countries within the last 40 years due to the regulation put into place (see Sects. 2.3 and 6.2.3.1) and the emission reduction technologies needed for compliance (see Sect. 6.2.2.3), the worldwide emissions of the most important greenhouse gas carbon dioxide (CO_2) are still increasing (see Sect. 2.3).

Climate change is at least partly caused by human activities[10] (a comprehensive glossary presenting important terminology in the field of climate change can be found in Matthews 2018). Anthropogenic (i.e. human-made) emissions lead to a shift in the composition of gases in the atmosphere and thus form the so-called

[10] Natural activities causing changes of the climate are, e.g., variations of solar cycles.

anthropogenic greenhouse effect. Concerning emission quantities and impact, CO_2 is seen to be the most important greenhouse gas, yet there is a variety of greenhouse gases (GHG), including methane (CH_4), nitrous oxide (N_2O) and fluorinated gases. Greenhouse gases trap heat in the atmosphere. This leads to an intensification of the natural greenhouse effect and an increasing global mean surface temperature, so-called **global warming** – an essential part of climate change. The impacts of climate change may be different for different regions. Still, it is to be feared that in general climate change may lead to more extreme weather events (e.g. drought), rising sea levels, melting of the Arctic ice etc. (cf. e.g. Hoegh-Guldberg et al. 2018). According to the Intergovernmental Panel on Climate Change (IPCC), the anthropogenic temperature increase had already in 2017 reached about 1 °C compared to the pre-industrial level. However, it is essential to note that land regions suffer from even greater warming than the global average (cf. Allen et al. 2018, pp. 56–62). In this context, it has to be mentioned that the climate also changed in the past, but the current changes are much faster than what the earth has witnessed up to now, making it much more difficult for ecosystems to adapt to the new circumstances.

The effect of greenhouse gas emissions on climate change mainly depends on the following two properties of the corresponding gases (cf. Ardone 1999, pp. 85–90): the atmospheric lifetime, which is the residence time in the atmosphere, and the so-called radiative forcing, which is the ability of the gas to absorb infrared radiation. To compare the effects of different gases on the climate and to translate GHG emissions into carbon dioxide equivalents (CO_2-eq.), so-called **global warming potential** (**GWP**) factors of greenhouse gases over a given period are calculated relative to that of CO_2 (GWP factors of CO_2 are set to 1 for all considered periods) (cf. Forster et al. 2007, pp. 210–216):

$$\mathrm{GWP}_i = \frac{\int_0^T (\Delta F_i \cdot [\gamma_i(t)])\mathrm{d}t}{\int_0^T \left(\Delta F_{\mathrm{CO}_2} \cdot [\gamma_{\mathrm{CO}_2}(t)]\right)\mathrm{d}t} \tag{6.14}$$

with

i	Greenhouse gas
ΔF_i	Radiative forcing of the greenhouse gas i
$[\gamma_i(t)]$	Time-dependent abundance of the greenhouse gas i at time t based on a 1 kg initial emission impulse
T	Time period (e.g. 100 years) considered for the calculation of the GWP.

Using GWP factors makes it impossible to develop a greenhouse gas reduction strategy with the objective to reduce the concentration of these gases in the atmosphere during specific years, e.g. during the years with the maximum effect on the climate, within a given period. GWP factors cannot be clearly interpreted; a high GWP factor may result from a greenhouse gas having a low radiative forcing but a long residence time in the atmosphere or from a greenhouse gas having a high

Table 6.5 GWP factors of different greenhouse gases

	(Average) Lifetime (years)	Cumulative forcing over 20 years	Cumulative forcing over 100 years
CO_2	No single lifetime can be given	1	1
CH_4	12.4	84	28
N_2O	121.0	264	265

Source IPCC (2014, p. 87)

radiative forcing but a short residence time in the atmosphere (cf. Ardone 1999, pp. 86–87). Although the calculation of GWP factors inevitably results in some loss of information, these factors are often used to develop strategies to reduce greenhouse gas emissions as they are relatively easy to compute and handle. Table 6.5 shows the GWP factors of carbon dioxide, methane and nitrous oxide for periods of 20 (GWP_{20}) and of 100 years (GWP_{100}) (cf. IPCC 2014, p. 87).

Acid depositions are a consequence of emissions of NO_x and SO_x, which are converted in the atmosphere via nitrous acid (HNO_2) and sulphurous acid (H_2SO_3) into nitric acid (HNO_3) and sulfuric acid (H_2SO_4). Acid depositions are also called acid rains and have a pH below 5.0 on the pH scale for measuring the acidity, going from zero to 14. These depositions influence forests (forest dieback), waters and soils in many different ways due to acidification.

NO_x emissions (together with phosphorus emissions) furthermore contribute to so-called nutrient contamination. This eutrophication effect might, at first glance, seem to be somewhat positive. Still, it has many negative aspects as it can lead, e.g., to an extreme growth of algae in waters with undesired consequences like oxygen depletion and nitrate enrichment in soils and (drinking) groundwater. Furthermore, under sunlight, NO_x emissions and VOC emissions are starting substances for ozone formation (photochemical ozone or “summer smog”).

Similar to the procedure to compare the climate effects of different greenhouse gases with the help of GWP factors, also various pollutants can be integrated into the calculation of their potential for acidification (acidification potential, AP), for eutrophication (eutrophication potential, EP) and for photochemical ozone creation (photochemical ozone creation potential, POCP).

Emissions of trace elements like mercury (Hg) and particulate matter can directly impact human beings causing severe health problems. Particulate matter might get into organs or even the bloodstream of human beings, and mercury is toxic to the human nervous system.

Besides the mentioned air emissions, the combustion of fossil fuels leads to ashes and sludge, which have to be disposed of depending on their components. Furthermore, thermal power plants need cooling, leading to emissions of heat, e.g. to adjacent rivers. Finally, the various kinds of electricity production result in different visual and noise impacts.

Air pollutants can be transported over long distances (long-range transmission), resulting in so-called **immissions** and subsequent impacts on human beings and the environment far away from the point of origin of the emissions. During the transportation phase, the produced emissions might be degraded or converted to

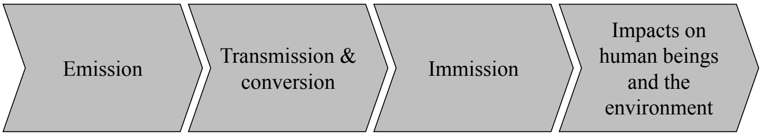


Fig. 6.9 Emission impact pathways

other substances (so-called secondary pollutants). These new substances will be deposited either with the help of atmospheric water (wet deposition) or as gases and particles (dry deposition).

To what extent emissions cause damages via immissions depends on different factors like the concentration rate. Calculations of the environmental and economic consequences are extremely challenging due to massive uncertainties as the whole pathway from the formation of the emissions to the resulting damages has to be considered (see Fig. 6.9 and e.g. ExternE 2018).

As emissions of air pollutants in one country can lead to immissions in another country, international cooperation is needed to identify where emissions should be reduced considering air dispersion (t_{ij}) from one region to other regions. Such an approach has been realised in Europe to find the cost-minimal strategy not to exceed the so-called critical loads, which can be interpreted as upper load limits. This has been implemented using the integrated assessment model Regional Air Pollution INformation and Simulation (RAINS), which can be expressed as a cost minimisation problem (6.15)–(6.17) (cf. e.g. Alcamo et al. 1990):

Objective function

$$\min \sum_c c_c \cdot v_c \cdot \xi_c \tag{6.15}$$

subject to the following restrictions

$$\sum_c v_c(1 - \xi_c)t_{cr} \leq \Gamma_r \quad \forall r \tag{6.16}$$

$$0 \leq \xi_c \leq 1 \quad \forall c \tag{6.17}$$

with

c country index
c_c per-unit emission reduction costs
v_c emissions in country c
ξ_c emission reduction rate
r region [e.g. 50 × 50 km]
t_{cr} transfer coefficient
Γ_r critical load in region r.

Already this simplified model description illustrates that a lot of input data is necessary – like critical loads, transfer coefficients, emission inventories and cost functions – to calculate the optimal reduction rates for air pollutants. Furthermore, appropriate incentive structures have to be put in place to realise this cost-minimal solution. Otherwise, it might lead to a financial burden for countries not partaking in the benefits of the realised emission reduction.

6.2.2.3 Emission Reduction Technologies

An efficient emission reduction strategy may not only make use of specific emission reduction technologies but also aim at avoiding emissions by a reduction of the consumption of the related energy services or the use of more sustainable production routes (see Sect. 2.4). Emission reduction technologies in the narrower sense can be differentiated according to the location where the pollution reduction takes place into primary measures (pre-combustion and in-combustion technologies) and secondary measures (post-combustion technologies) (cf. Tan 2014, p. 18). With the help of pre-combustion technologies, the input of a combustion process is cleaned from substances inducing pollutions even before the fuel is used in the combustion process. If the firing technology itself is adjusted to reduce the formation of pollutants, the related technologies are called in-combustion technologies. In contrast to these two types of reduction technologies, secondary measures, also called **post-combustion** or **end-of-pipe technologies**, are used after the pollutants have been produced and released into the exhaust gas, which then is cleaned with the help of these technologies. To reach the required emission reduction level, even combining some of these technologies may be necessary.

In electricity production, pre-combustion technologies are of minor relevance compared to the other reduction technologies. One example of a pre-combustion technology is the desulfurisation of the fuel. Since many gas fields produce sour gases, the so-called gas sweetening by scrubbers using amine solutions is applied to remove sulphur compounds of natural gas. Furthermore, the **pre-combustion carbon capture technology** could be used to mitigate CO_2 emissions in the future. This technology is based on the **IGCC – internal gasification combined cycle** – process. A gasification stage is thereby inserted upstream of the gas turbine to generate a synthesis gas (see Sect. 4.1). This synthesis gas mainly consists of hydrogen and carbon dioxide, so that in a following step CO_2 can be separated.

In conventional power plants, in-combustion and post-combustion technologies are dominating. Developments to increase the efficiency of energy conversion processes can be counted among **in-combustion technologies**. By increasing the efficiency of technologies using fossil fuels to produce electricity, less input is needed and accordingly, fewer emissions, e.g. CO_2 emissions, are generated for producing the same output. Furthermore, in-combustion technologies are mainly used for NO_x reduction because the NO production strongly depends on combustion temperatures, the air ratio, and the residence time in the reaction zone, which modifications of the combustion process can influence. This already shows that there can be conflicting effects regarding different emissions; lowering the combustion temperatures might help reduce NO_x emissions (notably thermal NO_x), but can lead to lower efficiency and accordingly to higher CO_2 emissions.

Technologies to reduce the formation of NO_x comprise, amongst others, air staging, fuel staging and flue gas recirculation (cf. e.g. Tan 2014, pp. 268–272; Baumbach 1990, pp. 341–347). As the NO production depends on the air ratio, a principle to reduce NO production is the limitation of the oxygen available in the central reaction zone. With the help of air staging technologies, the combustion zone is divided into different zones: a fuel-rich zone, where only a part of the needed air (and therefore oxygen) is supplied, and a fuel-lean zone, where the rest of the required air is provided. Besides the staging of the air, there is also the possibility that the fuel is staged. This fuel staging or reburning technology is also characterised by different zones. Here three different combustion zones exist: a primary zone with a primary fuel used under fuel-lean conditions, a secondary zone with a secondary fuel used under fuel-rich conditions, and a fuel-lean final combustion zone. In the secondary zone, also called reburn zone, parts of the NO_x emissions already produced in the primary zone are reduced again. Another primary reduction technology is the recirculation of the flue gas into the combustion area, which can help to reduce NO_x emissions by lowering the combustion temperature. These three technologies, staging of the air, staging of the fuel and recirculation of the flue gas, and their combinations are used in so-called low-NO_x burners and may lead to a reduction of NO_x up to 70% (cf. Baumbach 1990, p. 347).

The so-called **oxy-fuel process** may be considered an in-combustion solution to mitigate CO_2 emissions (cf. Tan 2014, pp. 358–360). An air separation unit is needed for this process because not air but oxygen is used in the firing process. The resulting flue gas mainly consists of the two products H_2O and CO_2, which can be separated in a final step.

Post-combustion technologies are widely used to remove pollution emissions – e.g. SO_2, NO_x, particulates (cf. Tan 2014, pp. 277–313). For capturing particulate matters besides cyclones, filters and electrostatic precipitation (ESP) are used. In electrostatic precipitation, the particles are charged electrostatically and then deposited on a collecting electrode, from where they have to be removed, e.g. with the help of mechanical vibration. The separation efficiency of ESP is above 95% (cf. Baumbach 1990, p. 336). An alternative to an ESP with a relatively similar separation efficiency (cf. Tan 2014, p. 281) is the use of filters, like bag-house filters.

Flue gas **desulfurisation** (FGD) technologies are widely applied to remove SO_x emissions ($DeSO_x$) from the flue gas of power plants. The dominating version is the wet FGD, where typically limestone, i.e. calcium carbonate ($CaCO_3$), is used to produce calcium sulphate dehydrate ($CaSO_4 \cdot 2H_2O$) and CO_2[11] by capturing SO_x according to the following reactions (cf. Khartchenko 1997, p. 120):

[11] Again a conflicting effect is observable here: SO_2 reduction via limestone scrubbing leads to an increase of CO_2 emissions.

$$SO_2 + {}^1/_2O_2 \rightarrow SO_3$$
$$CaCO_3 + H_2O \rightarrow Ca(OH)_2 + CO_2$$
$$Ca(OH)_2 + SO_3 + H_2O \rightarrow CaSO_4 \cdot 2H_2O.$$

Calcium sulphate dehydrates, better known as gypsum, is the final product of FGD with limestone. This process is characterised by a colossal absorption tower, in which the flue gas is fed in and sprayed with the limestone suspension. The resulting separation efficiencies lie above 95% (cf. Baumbach 1990, p. 376). It is also possible to use other inputs as, e.g., magnesium hydroxide instead of limestone.

Concerning the reduction of NO_x emissions (DeNO$_x$) by using end-of-pipe technologies, the selective catalytic reduction (SCR) process is widely spread, but also the selective non-catalytic reduction (SNCR) can often be found in industry. The main difference between these two technologies is the existence of a catalyst in the case of SCR making this technology more expensive than SNCR, but also leading to higher separation efficiencies of above 95% (cf. Tan 2014, p. 295). Typically, ammonia (NH_3) is used as input for this process leading to the following main reaction at the catalyst:

$$NH_3 + NO \rightarrow N_2 + H_2O.$$

These post-combustion technologies also help reduce the emissions of trace elements like mercury, for which often no specific reduction measures have been installed.

Figure 6.10 shows three possible arrangements of the different **end-of-pipe technologies** in a hard coal power plant: high-dust, low-dust and tail end. From a thermodynamic point of view, the high-dust arrangement is preferable, as the SCR needs rather high temperatures of more than 300 °C to be operated.

In future, these three post-combustion technologies could eventually be supplemented by a fourth end-of-pipe technology to remove CO_2 from the exhaust gas and store it underground to prevent its contribution to the greenhouse effect (**carbon capture and storage**, **CCS**). Up to now, this concept has been realised in some industrial large-scale demonstration projects. In this post-combustion process, CO_2 is separated from the exhaust gas by a solvent, e.g. an amine solution. Compared to the other processes to capture carbon, i.e. the IGCC process and the oxy-fuel process, one advantage of this technology is that existing power plants can be upgraded with this post-combustion technology. An essential prerequisite for such an upgrading is that enough space at the respective site is available. In the context of CCS technologies, it has to be mentioned that the CO_2 captured has to be transported and stored safely. This could be realised by pipelines from combustion installations to deep underground storage possibilities. In the case of injecting CO_2 into (partly depleted) oil and gas fields, this can even help increase the field's output (so-called enhanced hydrocarbon recovery[12]). One should note that there are

[12] Especially enhanced oil recovery (EOR) and enhanced gas recovery (EGR).

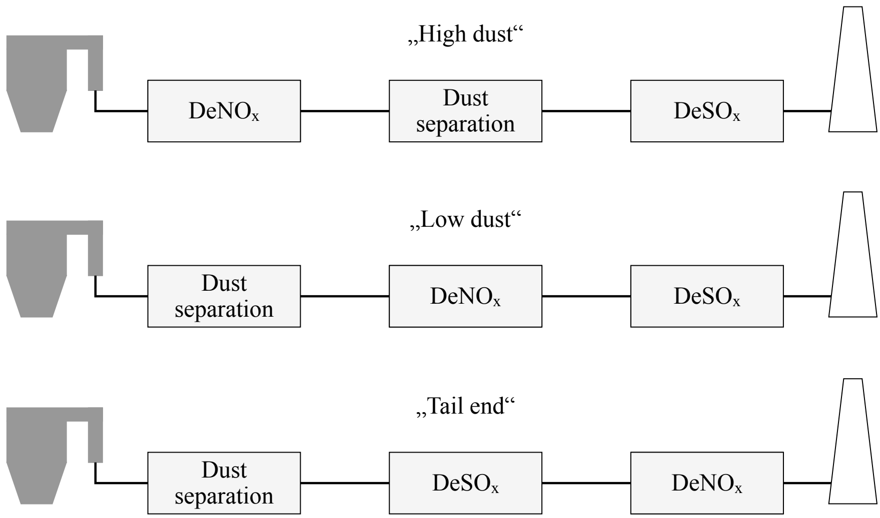


Fig. 6.10 Arrangements of the three end-of-pipe technologies $DeNO_x$, dust removal and $DeSO_x$. *Source* Own illustration based on Richers and Günther (2014, p. 38)

also limits to separation efficiency for CCS technologies, depending on the technology the attainable maximum is between 80 and 98% (cf. Mathieu 2010). Additionally, the energy conversion efficiency is reduced.

In the context of reducing the CO_2 concentration in the atmosphere, installations using bioenergy combined with CCS (BECCS) are an interesting option, as the plants take CO_2 out of the atmosphere during the period of growth and in the utilisation phase no CO_2 is released into the atmosphere (so-called negative emissions[13]). An alternative to storing the captured CO_2 in underground storage could be to use it as an input for the production of chemicals (**carbon capture and utilisation**, **CCU**).

6.2.2.4 Excursus: Life Cycle Assessment

When calculating the environmental impact of a product, a service, a technology or even an entire system (in the following just called "object under consideration"), the whole life cycle of the object under consideration should be considered. This comprehensive approach is often referred to as **life cycle assessment** (**LCA**), eco-balancing or cradle-to-grave[14] analysis. According to the standards of the International Organization for Standardization (ISO) (cf. ISO14040 2006; ISO14044 2006), an LCA consists of the four phases "Goal and scope definition", "Inventory analysis", "Impact assessment" and "Interpretation" (see Fig. 6.11). These phases do not have to be executed in a purely sequential way, rather it is

[13] Another possibility to produce negative emissions is the use of direct air capture (DAC).
[14] Only parts of the whole life cycle are considered in so-called cradle-to-gate or gate-to-gate analyses.

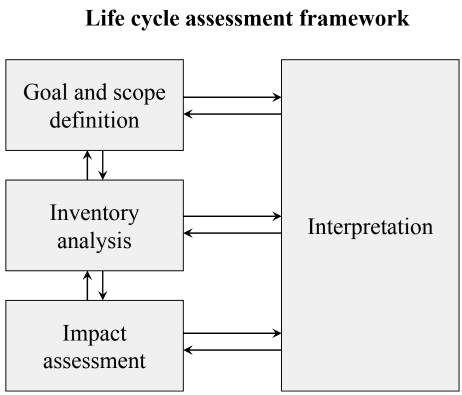


Fig. 6.11 Phases of an LCA according to ISO14040 (2006)

possible to jump back and forth between these stages to realise adjustments (cf. e.g. Matthews et al. 2014, p. 84).

In the first phase, framework conditions have to be defined, like the study's objective, the system boundaries and the so-called functional unit. The functional unit is needed to quantitatively describe the function of the object under consideration; so, if the environmental impacts of a technology for electricity generation are to be analysed, an appropriate functional unit (for the function electricity production) would be one kilowatt-hour of electricity produced (cf. Turconi 2014, pp. 5 and 11).

Based on these definitions, all the energy and material flows caused by the object under consideration, in other words all the inputs and outputs, are collected in the inventory phase; in the example of assessing an electricity generation technology, this would comprise data from the process of manufacturing the electricity generation technology via the emissions during the electricity production process up to the dismantling of the generation technology.

In the assessment process (third phase), the environmental impacts caused by the collected inputs and outputs of the object under consideration are analysed. This impact assessment phase consists of the three mandatory elements *selection*, *classification* and *characterisation* and further optional elements (cf. Matthews et al. 2014, pp. 366–396). First, the considered impact categories (e.g. global warming), indicators for these categories (e.g. radiative forcing over a given period) and characterisation models [e.g. concept of global warming potential (GWP)] have to be selected. Then the inputs and outputs connected to the object under consideration are linked to one or more of these impact categories, which is called classification. In the characterisation stage, characterisation factors (sometimes also called equivalence factors) resulting from the chosen characterisation model (e.g. GWP factors of the different greenhouse gases) are used to calculate the indicators. In addition, the ISO framework for LCA also allows for further optional elements, like the weighting to transfer the results for the different impact indicators into one value, showing the total impact of the object under consideration. The element of weighting is not mandatory as in many cases it might be extremely challenging (and subjective) to develop the needed weighting factors since the different impacts can hardly be compared; e.g. which weighting factors should be used to add up the impact categories of Global Warming and acidification? Finally, the results of the previous phases are discussed and recommendations are made.

6.2.3 Policy Instruments

6.2.3.1 First-Best and Second-Best Instruments

Negative externalities materialise if **property rights** are not applicable. Producing emissions and emitting them into the atmosphere leads to external costs if the costs caused by the damages resulting from these emissions are not reflected in the market prices. Then the producers of these emissions have no incentive to reduce them. However, there will be damages caused by environmental problems resulting from these emissions. Therefore, there is a need to address this market failure by implementing some policy instrument.

In an economic perspective, a straightforward solution is to establish property rights and create markets, an idea going back to Coase (1960). Coase showed that without the consideration of transaction costs and under further idealising assumptions, the allocation of property rights would lead to a bargaining process resulting in a solution, which is **Pareto efficient** (see Sect. 6.1). This so-called efficiency theorem implies that in the bargain outcome the marginal abatement costs of the polluters are equal to the marginal damage costs (see Sect. 6.2.1). The solution will be Pareto efficient, independently of the original allocation of property rights; however, the allocation of property rights will result in distributional effects. So an efficient outcome is possible, even if property rights are given to the polluters and not to the damaged third parties. And such a solution could even emerge without government intervention – through the willingness of the damaged parties to pay for pollution reduction. Yet this is a rather theoretical result since it is only valid in the absence of transaction costs.[15] In real-world problems, bargaining is likely to be difficult and costly if many polluters and damaged third parties are involved. Hence, a pure bargaining solution is, if at all, only practical for small-scale problems with only a few involved parties.

Yet in the same theoretical vein of mainstream environmental economics, two other welfare-optimal **first-best instruments** exist: the **Pigou tax** approach, also called Pigouvian tax and the first-best emissions trading approach.[16] Arthur Pigou developed the idea to shift the private cost curve up by increasing the costs with the help of a tax (Pigou 1920). This tax corresponds to t^* in Fig. 6.5. With the help of

[15] Note that transaction costs is used in the broad economic sense of costs related to a market transaction. These include here among others the cost for negotiating an agreement, for measuring pollution quantities and for enforcing the pollution limits.

[16] In principle such a trading approach can be used for all kinds of environmental goods, e.g. for land use (cf. Walz et al. 2009).

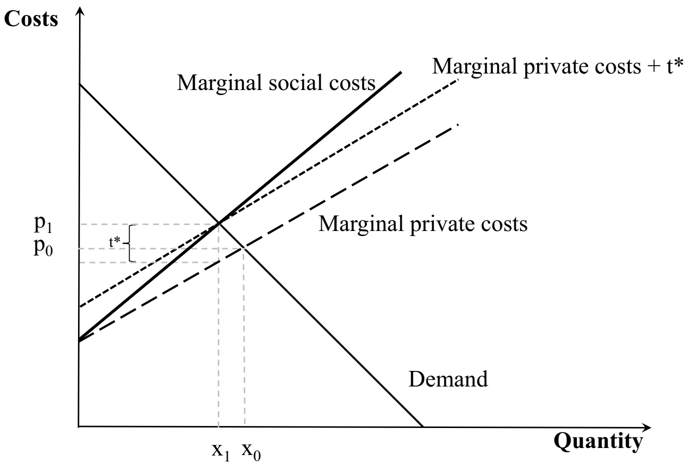


Fig. 6.12 Pigouvian tax shifting the marginal private cost curve. *Source* Own illustration based on Fritsch (2018, p. 113)

this tax, the marginal private cost curve (including the tax) intersects with the demand curve just in the point where the marginal social cost curve demand of the product intersects with the demand curve (see Fig. 6.12).

In the first-best **emissions trading** approach, a maximum limit for the total emissions being allowed is set – the so-called cap, which corresponds to *e** in Fig. 6.5. This cap is then broken down into emission allowances, with each allowance representing the right to produce the corresponding amount of emissions, e.g. one tonne of CO_2. Emission allowances can be traded. That is why this system is also called a "**cap and trade system**" (cf. Dales 1968).

The tax and the emissions trading approach are somewhat symmetric: Using a tax solution leaves the emission reduction to the market by setting a price, whereas in a trading solution, the emissions level is fixed and the price is left to the market (cf. Feess and Seeliger 2013, p. 119). But these two strategies for internalisation face challenges in practice related to information deficits: the government has to know the marginal emission reduction costs of all polluters as well as the marginal damage costs to determine the optimal control parameters of the respective instrument: the tax level (*t** in Fig. 6.5) in the case of the Pigou tax approach or the maximum of emissions allowed (*e** in Fig. 6.5) in the case of the first-best emissions trading approach.

Different policy instruments have been developed besides the non-economic idea to appeal to the emitters' sense of moral behaviour. These are frequently labelled as **second-best** solutions, since they are less efficient than the first-best solutions under idealised textbook conditions. For instance, the government might set an environmental target exogenously and identify a set of command and control measures to meet this target. The following major criteria may be used to assess second-best instruments for a given environmental problem:

- **target achievement/environmental effectiveness**: Will it be assured that the exogenously given environmental target is reached with the policy instrument at hand?
- **cost efficiency/static efficiency**: Will the exogenously given environmental target be reached at the lowest costs with the policy instrument at hand?
- **dynamic efficiency**: Will the policy instrument at hand set incentives to develop new technologies to minimise long-term costs?

Furthermore, criteria like political acceptability, practicability, distributional effects, social acceptance and adjustability (cf. e.g. Fais 2015, pp. 9–10) are relevant in the selection process.

Regarding the possible environmental policy instruments (cf. e.g. Perman et al. 2003, pp. 202–246; Cansier 1993, pp. 155–280), there is often a differentiation into four groups: Command and control instruments, economic incentive instruments, information instruments like information campaigns and voluntary instruments like voluntary agreements by the industry.

By using **command and control instruments**, the government directly intervenes in polluters' production processes. This can be realised using technology-based standards, with the help of which the permitted technologies are fixed; e.g. often only so-called Best Available Technologies (BAT) are postulated to be used. The other form of command and control instruments are performance-based standards, with the help of which emission limit values for a production process are fixed, but not the means by which these levels are to be reached. In Europe, the Industrial Emissions Directive (IED) sets different emission limit values for NO_x, SO_2 and dust emissions from large combustion plants. Upper limits for emissions from combustion units, which might be set in milligrammes per cubic metre of the flue gas (mg/m^3), can differ, e.g. depending on the size of the firing installation and the fuel used. Emission limit values have been one of the dominating policy instruments in the energy sector. They help to control emissions from individual installations and the corresponding emission reductions contribute to combat the related environmental problem, for instance, in the case of a regional ecological problem like acid rain. Yet, there is no guarantee that the total emissions, e.g. in a region, are capped by such an instrument. They might even increase if the number of used installations grows. But the main disadvantage of command and control instruments is that they will hardly lead to a cost-efficient solution, as the polluters do not have many options on how to comply with the given regulation. The company-specific situation is usually not considered, especially in technology-based standards. But also in the case of performance-based standards based on emission limit values, different marginal reduction costs across companies are not taken into account when all companies have to reach the same emission limit. In contrast, a cost-efficient policy instrument will lead to a situation where the marginal reduction costs of the companies involved are the same. This can be easily shown by using the Lagrangian method for the following optimisation problem (cf. e.g. Feess and Seeliger 2013, pp. 63–65):

Objective function

$$\min \sum_i C_i(\Xi_i) \tag{6.18}$$

subject to the following restriction

$$\sum_i \Xi_i \leq \Delta\Gamma \tag{6.19}$$

with

i company index
C_i absolute emission reduction costs
Ξ_i emission reduction
$\Delta\Gamma$ total emission reduction obligation.

For urgent environmental problems, like acid rain in the 1970s, emissions reduction obligations have been set at a very high level. In such a case, there are often very limited possibilities to fulfil the emission limit values, e.g. end-of-pipe technologies with separation efficiencies near to 100% have to be used to limit SO_2 emission values. So the economic drawback of command and control instruments is circumstantial.

Economic incentive instruments have as central idea to give incentives to polluters to change their behaviour. These instruments can be designed in-line with the **first-best instruments**, i.e. the Pigou tax, respectively, the emissions trading approach. To cope with the mentioned information deficits, the level of the tax – so-called price approach (cf. e.g. Baumol and Oates 1971) – or the allowed emissions – so-called quantity approach – is set administratively. Such a quantity approach clearly has to be distinguished from the theoretical first-best emissions trading approach based on the optimal emission cap. This may therefore be referred to as second-best emissions trading. The main advantage of economic incentive instruments is their cost efficiency. Each company is free to decide how to react: companies have to identify whether it is more favourable for them to reduce their emissions or to continue producing emissions and paying the tax or using emission allowances. All polluters will reduce their emissions until their marginal reduction costs are equal to the market price of the tradable emission allowances or the tax, so the marginal reduction costs of the different companies will be the same. As soon as the market price of the tradable emission allowances or the tax is below the individual marginal reduction costs, the polluter will choose to pay the market price of the emission allowances or the tax. One advantage of the second-best **emissions trading** approach compared to the **emission tax** is the environmental effectiveness: the exogenously given environmental target[17] in the form of the emission cap will

[17] It has to be mentioned here that in reality these targets are often the result of intense political negotiations.

be reached if efficient control mechanisms are established. In the case of a tax solution, this tax will probably have to be adjusted by the regulatory authority in a kind of trial-and-error procedure to converge to the envisaged environmental target.[18] Besides taxes and emission trading, there is also the possibility to incentivise polluters to change their behaviour by giving them subsidies. Subsidies can have the form of direct payments or investment grants, where the recipients get a fixed payment if they carry out a predefined action. Another form of incentivising market participants by subsidies is to put in place a price support mechanism, which enables the producers to get predefined prices for their goods (cf. Mechler et al. 2016).

Economic incentive instruments seem to be good solutions for the limitation of emissions causing global environmental problems. For such problems, the location where the emission reduction is realised is not decisive. On the other hand, these instruments could lead to regional hot spots with high immissions (cf. Feess and Seeliger 2013, p. 125), if all installations producing emissions that lead to immissions in this region decide not to reduce them – therefore, its application is not as straightforward if environmental damage is location-dependent. Emission trading schemes exist for different emissions: in 2005, an emission trading scheme was introduced to limit CO_2 emissions of European combustion installations, in the US emission trading schemes were realised even earlier, even for the reduction of emissions leading to regional environmental problems, as e.g. SO_2 [see, e.g., the Acid Rain Program (ARP)]. To avoid regional hot spots of environmental problems, the emission trading scheme was there supplemented by other regulations assuring local emission reduction.

6.2.3.2 The Implementation of Emissions Trading

To establish an emissions trading system, first, the system boundaries have to be set (e.g. the designation of the considered market players and emissions). As far as possible, limitations regarding participating sectors, countries, etc., should rather be avoided to have one comprehensive system. Another possibility to enlarge the system boundaries is linking existing emissions trading systems or integrating emission reductions realised in sectors not part of the trading scheme. Emission trading concepts can be designed for different target groups, the system might focus on either upstream players (e.g. entities placing emission-causing energy carriers on the market) or downstream players (e.g. producers of emissions). As soon as the system boundaries have been determined, a cap for the permitted total emissions has to be fixed administratively. In the next step, this cap has to be broken down into rights to produce a specified amount of emissions (so-called emission allowances). In SO_2 emissions trading, an emission allowance could, e.g., represent the right to produce one tonne of SO_2. These emission allowances are then allocated to the participating entities by using appropriate allocation mechanisms. This initial allocation can be realised by issuing the allowances free of charge according to the

[18] There is a broad discussion in environmental economics about the relative benefits of price- versus quantity-based instruments under uncertainty starting with Weitzman (1974).

emissions the company produced in a reference period in the past (so-called grandfathering), potentially multiplied by a particular reduction factor. This form of allocation might penalise companies that already invested in emission reduction measures resulting in lower emissions in the past and can – of course – not be applied for companies just entering the market. Another form of free of charge allocation is to use the emissions of a reference or benchmark process (e.g. of the BAT) and allocate the corresponding allowances to the used processes. Alternatively, emission allowances can be issued via auctions. The participants are subsequently free to trade the allocated emission allowances (e.g. via a secondary market). Furthermore, the involved companies have the responsibility to report their emissions. At the end of the compliance period, the participants finally have to deliver emission allowances equal to the emissions they produced during this period. Then the whole procedure starts again for the next compliance period. To control the compliance of the companies involved, a regulatory authority has to be established and an organisational and administrative effort is required. Within the emissions trading system, further flexibility mechanisms can be integrated: there might be the possibility to bank emission allowances to use them for compliance in later periods (so-called banking) or the option to use emission allowances that will be allocated in later compliance periods already in the current period (so-called borrowing) (cf. e.g. Flachsland et al. 2008, pp. 18–19).

Through emissions trading, a new factor of production arises in the participating companies,[19] which has to be integrated into production and investment planning processes. Depending on the cap level, this production factor might become scarce, leading to higher allowance prices on the market. An emission allowance represents a fundamental factor of production, which has at least one exceptional feature: as the participants are only obliged to deliver allowances at the end of the compliance period, this factor of production can be procured even after the production of the emissions for which it is used, in other words, the producer of emissions can go physically short (cf. Wallner et al. 2014, p. 18).

In line with the concept of opportunity costs, companies will price in a scarce production factor – independently of the chosen allocation mechanism,[20] as the company has the opportunity to use this factor of production in another way: the company could decide not to use it as an input for its own production process, but to sell it on the market. On the other hand, the allocation mechanism can lead to considerable distributional effects. Whereas free of charge allocation might help to create or sustain acceptance for the system, such an allocation may produce

[19] Factors of production are inputs needed to be able to produce the output of the company. In Economics usually the three factors of production land, capital and labor are differentiated, in Business Administration much more detailed classifications exist (cf. e.g. Dyckhoff and Spengler 2010, p. 16-19).

[20] This does not hold true for contingent allocation rules. A scarce production factor will not necessarily be fully priced in by a company if the allocation in future trading periods depends on the actions of the company still to be taken, e.g. if the reference period of a later allocation period is updated and incorporates the current year the production today might influence the allocation in future (cf. Weber and Vogel 2014).

additional profits (so-called windfall profits) as the involved companies might raise their product prices (in the power sector, the wholesale electricity prices) according to the economic value of this new factor of production. However, they do not have to pay for it. Alternatively, the auctioning of emission allowances will lead to revenue streams for the government.

6.2.4 Limiting Climate Change

One of the most important political achievements in combating climate change (see Sect. 6.2.2.2) is the **United Nations Framework Convention on Climate Change** (**UNFCCC**), which already entered into force in the year 1994. This convention has been operationalised by the so-called **Kyoto Protocol**, coming into force in 2005, and the so-called **Paris Agreement**, coming into force in 2016. Whereas the Kyoto Protocol set targets for the reduction of greenhouse gas emissions in industrialised countries for the commitment periods 2008–2012 and 2013–2020, according to the subsequent Paris Agreement all parties to this agreement have to present their contribution to the reduction of greenhouse gas emissions by the preparation of so-called nationally determined contributions (NDCs).

Countries have put in place different instruments to fulfil the objectives set by the Kyoto Protocol and the NDCs. This chapter will focus on two rather different ways to approach the greenhouse gas reduction targets, which both have been implemented: on the one hand, setting an emission reduction target, allocating the corresponding emission rights and allowing trading of emissions rights (Sect. 6.2.4.1), on the other hand, setting incentives for specific technologies, which do not or hardly lead to greenhouse gas emissions, by introducing support schemes only for them, e.g. feed-in tariffs for renewable energies (Sect. 6.2.4.2). Finally, in Sect. 6.2.4.3, possible interactions between these instruments, if they are used at the same time, will be discussed.

6.2.4.1 The EU Emissions Trading System (EU ETS)

In 2005, the **EU Emissions Trading System** (**EU ETS**) was launched in Europe[21] to limit CO_2 emissions of combustion installations with a thermal input exceeding 20 MW. Later, the system boundaries were expanded to integrate emissions of the greenhouse gases N_2O and perfluorocarbons (PFC) from specific industrial processes, which are converted into CO_2-equivalents by using GWP_{100} factors (see Sect. 6.2.2.2), as well as to emissions from the aviation sector.[22] About 45% of total EU greenhouse gas emissions, more than 11,000 installations in over 30 countries are covered by the EU ETS (cf. e.g. European Commission 2018).

[21] As the environmental problem that this emission trading system is aiming at is a global one, limitations regarding participating countries should rather be avoided.

[22] Temporarily the scope regarding the aviation sector was reduced to flights between airports in Europe.

As the European emissions trading scheme does not cover all the greenhouse gas emissions in Europe, emission reduction targets for the sectors not included in the EU ETS had to be put into place, which was done by setting national targets for each member state (Effort Sharing Regulation (ESR)[23]). The fact that European member states have, on the one hand, emission reduction obligations on a national level and that on the other hand combustion installations in these countries are participating in the EU ETS, leads to challenges concerning the breaking down of the national reduction targets to the different sectors, because under an emission trading scheme it is not clear in which installations the emission reduction will be realised.

The trading system started with a test period from 2005 to 2007, followed by the trading periods 2008–2012 (phase II) and 2013–2020 (phase III). The fourth trading period comprises the time horizon from 2021 to 2030. The emission allowances of the EU ETS are called EUAs (European Union Allowances) and represent the right to produce 1 tonne of CO_2-equivalents each. The given emission cap shrinks from year to year to reach the objectives to reduce the emissions from the participants (at the time of writing this book, the objective was 43% in 2030 compared to emission levels in 2005).

During the first two trading periods, the European countries had to develop so-called National Allocation Plans (NAPs), indicating how many emission allowances are issued in each country and according to which allocation mechanisms these allowances are distributed to the involved installations. According to these NAPs, most installations got the allowances free of charge, mainly based on benchmarks. This means that e.g. many power plants got the allowances according to a (fuel-specific) benchmark (kg CO_2/kWh), which had to be multiplied by a utilisation factor (full-load hours per year). This utilisation factor was calculated from historical data or had to be estimated by the plant operator or was set administratively by the government. In the third trading period, auctioning (using sealed bids and uniform pricing) has become the default allocation mechanism, but still industrial processes get (parts of) the allowances free of charge based on benchmarks. Especially companies that might relocate their production sites due to economic reasons into a country, where they do not have to undertake efforts to reduce their emissions, get their allowances free of charge, as such relocation could lead to even higher CO_2 emissions (so-called **carbon leakage**). Participants have to submit sufficient allowances by the end of April of the following year to cover their previous year’s emissions. The system allows banking of the allowances. Only between phase I and phase II, emission allowances could not be banked because 2008–2012 was the commitment period under the Kyoto Protocol and the member states did not want to jeopardise the fulfilment of their emission reduction targets through the banking of allowances into this period. On the other hand, the EU ETS does in principle not allow borrowing. Effectively, borrowing is at least partially possible because at least parts of the yearly allocation process take place before the

[23] To fulfil the national targets, the ESR provides different flexibility mechanisms, e.g. it is allowed that member states buy “surplus emission reductions” from other member states.

allowances have to be surrendered to demonstrate compliance in the previous year. Other flexibility options are the possibility of generating emission credits by reducing emissions through projects in countries (international offsets) or even sectors (domestic offsets) not involved in emissions trading and using these credits for compliance within the emission trading scheme. Before the fourth phase of the EU ETS, it was allowed to at least partly exchange different kinds of these credits for EUAs: Certified Emission Reduction (CER) credits from projects that reduce emissions in developing countries [Clean Development Mechanism (CDM)], and Emission Reduction Unit (ERU) credits from projects in industrialised countries [Joint Implementation (JI)]. The credits are calculated by comparing the emission level in the situation with the emission reduction project to a hypothetical emission level of a business as usual (BAU) scenario; the project has to prove the so-called additionality of the emissions reduction, meaning that it must be shown that without the project the emission reduction would not have occurred. Therefore, this form of emissions reduction is called a **baseline and credit program**. As long as a greenhouse gas emission trading scheme does not cover all sectors and emissions worldwide, these credits provide an incentive to identify and use cheap emission reduction measures, which otherwise would not be exploited.

As the EU ETS fixes the overall cap of emissions for the participating sectors, additional political requirements for these sectors, e.g. national (domestic) decisions to phase out coal-based power generation or to introduce a carbon floor price, do only lead to an additional emission reduction, if the cap can be adjusted. Otherwise, the emission reduction in one country will be compensated by additional emissions in other countries [so-called waterbed effect (cf. Perino 2018)].

Since the beginning of emissions trading in Europe in 2005, the prices of EUAs have shown relatively high volatility. Already in the first phase, allowance prices exceeded 25 €/tCO_2-eq. and then fell back drastically, a development rather similar to what was observed in the second trading period. Trading phase III was characterised by rather low prices till 2017, and since then, a considerable increase can be seen (see Fig. 6.13).

Fig. 6.13 Development of EUA prices. *Sources* Own illustration based on data from ICIS and EEX

The collapse of prices in the trading period 2005–2007 was a consequence of false expectations followed by the discovery that allowances issued by member states were abundant, which led to a surplus in the market. As soon as this became clear, the prices of EUA dropped, resulting in a price of zero because banking was not allowed between phase I and phase II. Also, in the trading phase 2008–2012, the prices crashed due to a surplus of about 2 billion emission allowances in the market. The reasons for this surplus are manifold:

- the economic crisis in 2008, leading to a reduction in industrial production,
- the intense use of relatively cheap international offsets and
- interrelating policy instruments, like the support for renewable energy sources, leading to reduced demand for emission allowances (see Sect. 6.2.4.3).

Owing to this surplus, many allowances have been set aside by the market players to be used in future periods as they expect a scarcer market and banking is possible. The European Commission reacted to the price decline and the accumulation of banked allowances by taking emission allowances out of the market in the years 2014–2016 through the so-called backloading, and eventually deciding to put them into the so-called **market stability reserve** (**MSR**).

For trading phase IV, the yearly emission cap has been tightened considerably. Depending on the amount of allowances that are banked, further allowances will be put into the MSR, or rather allowances in the MSR will be returned into the market. In addition, an upper bound for allowances within the MSR has been introduced and all allowances in the MSR above this threshold will be cancelled, which has different effects on the EU ETS, e.g. eventually leading to an extra emission reduction of additional domestic emission reduction strategies (cf. Perino 2018, p. 263). Furthermore, countries are now allowed to cancel allowances in the EU ETS if they perform additional measures like a national coal exit. Hence, the effects of additional measures may go beyond a drop in demand for emission allowances and the corresponding allowance prices – although the operation rules for the market stability reserve may partly offset these effects.

When analysing the future allowance price development in the European ETS, it has to be considered that besides the market fundamentals, the participants' trading behaviour may impact EUA prices. As power companies sell their production at least partly on long-term future markets, they face the risk of a rise in costs of the needed factors of production, which they may want to hedge (see Chap. 8). Therefore, it seems rational to assume that power companies will buy the factors of production or futures or forwards for them (including emission allowances) at the time when they sell their electricity (cf. Wallner et al. 2014, p. 49).

Depending on the prices of the EUAs, there may be considerable impacts on the planning processes and operation decisions of the companies involved in the emissions trading scheme. Figure 6.14 illustrates the effects of the production factor *emission allowance* on production decisions of power companies, the subsequent merit order (see Sect. 4.4.1) and the related costs (CO_2 costs). It should be stressed once again that these effects are independent of the chosen allocation mechanism as

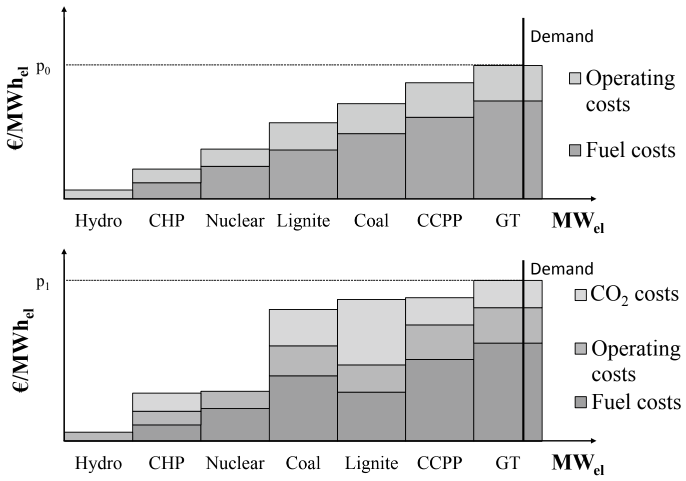


Fig. 6.14 Stylised merit-order curves and clearing prices with and without CO_2 emissions trading

the new production factor normally is priced in anyway. The wholesale electricity price increases by the CO_2 costs of the price-setting power plant from p_0 without emissions trading to p_1 with emissions trading. This illustration assumes that electricity demand is inelastic (see Sect. 3.1.4), which results in the vertical demand curve. Furthermore, in this illustration, the introduction of emissions trading leads to a change in the power plants' merit order (see also Sect. 7.1.1). Whereas the marginal costs of hard coal power plants are higher than the costs of lignite power plants without emissions trading, this changes under the assumed CO_2 costs: now the sum of all three variable cost items (fuel costs, operating costs and CO_2 costs) is higher for lignite power plants than for hard coal power plants (see lower part of Fig. 6.14).

The production factor *emission allowances* also influences power companies' investment planning (e.g. based on the net present value). With emissions trading, new cash flows occur: cash inflows change due to changed electricity prices, cash outflows change due to the purchase of emission allowances. Suppose emission allowances for new power plants are allocated free of charge (at least in some years of the installation's lifetime). In that case, this functions similar to an investment grant, stimulating new investments – but possibly also distorting the investment decisions (cf. Weber and Vogel 2014).

6.2.4.2 Renewable Support Schemes

A GHG emission trading system leads to incentives to invest in less greenhouse gas emitting technologies, such as renewable energies. Another possibility to increase the use of renewable energy sources for electricity production is to establish policy instruments that exclusively support these technologies. In addition to the incentives resulting directly from such a support scheme, renewable sources are often privileged by the priority connection of these installations and the priority purchase and transmission of the electricity produced in these units. Instruments to promote renewable energies can generally be differentiated into two basic clusters: those setting the remuneration for the technologies used (price-based instruments) and those setting the quantity of the technologies used (quantity-based instruments) (cf. e.g. Fais 2015; Held et al. 2014; Finon and Menanteau 2003). This is connected to some challenges in designing an appropriate support scheme for renewables: Should the instrument support the produced energy or the installed capacity? Should there be a different parameterization of the chosen instrument for different technologies or should the instrument be technology-neutral? To assess and compare different instruments to increase the use of renewable sources, evaluation criteria like efficiency and target achievement (see Sect. 6.2.3.1) may be used.

In the past, the price-based instrument of a **feed-in tariff** (**FIT**) was frequently used for increasing renewable electricity generation. Under this policy instrument, the producers of electricity from renewable energy sources are entitled to sell their green electricity to the (transmission) system operator and get a fixed payment, typically for each kilowatt-hour electricity produced or fed into the grid (e.g. in €ct/kWh$_{el}$). Typically, the level of the FITs depends on the technology and the year of installation, perhaps even on the weather conditions, like average wind speeds at the location concerned. The general idea of such specific FITs is that the remuneration payments are sufficient to cover the generation costs of the technology used. FITs are typically guaranteed for a fixed period of years. This instrument has been often used to accelerate the market introduction of a technology. On the one side, FITs lead to rather long-term price guarantees for the investor. On the other side, the instrument does not incentivise a real market integration. This is because the owner of a renewable energy installation does not have to care about the electricity market ("produce and forget") because the remuneration is fixed, totally independent of the market price. The electricity can be fed in whenever the unit is operating. The German Renewable Energy Sources Act (EEG) focussed for many years on this policy instrument, leading to a strong increase of renewable energy installations in Germany and a considerable reduction of the worldwide investment costs notably for PV systems at the expense of high additional costs for the German (non-privileged) electricity consumers due to the so-called EEG-levy.[24]

An extra incentive for increasing renewable electricity generation might exist if the electricity produced in decentralised units, e.g. rooftop PV, can be used to cover parts of the electricity demand of the so-called prosumer (self-consumption; see Sect. 10.7.4). Under a so-called **net metering** scheme the feed-in of electricity is subtracted from the amount of electricity obtained from the grid. In other words, the feed-in tariff has the same level as the respective electricity retail price. In contrast, there are systems that differentiate between the tariff a customer has to pay for electricity taken from the grid and the payment the customer gets for the feed-in of electricity produced in a decentralised production unit. A system with such a

[24] This renewable levy covers the gap between electricity wholesale market prices and the renumeration paid.

differentiation and the requirement that the whole electricity produced in the decentralised unit has to be fed in, in other words self-consumption is not allowed, is called **gross metering**.

A more market-oriented form of price-based instruments to foster renewable electricity production are **market premiums**, also called **feed-in premiums**. Here a premium is paid on top of the electricity wholesale price whenever operators of the renewable energy installation sell their electricity on the market. The operators must market their output. Therefore this instrument is also called direct marketing. So the renewable energy operator has different revenue streams: the wholesale electricity price and the premium. The premium can be determined in different ways: it can e.g. be fixed, variable (floating) or limited by a cap and floor (cf. Held et al. 2014, pp. 38–43). In the German market premium model the difference between the remuneration according to a fixed feed-in tariff and the monthly average electricity price at the exchange is offset with the help of a monthly market premium (see Fig. 6.15). This leads to an incentive to shift electricity production to hours with wholesale electricity prices above the monthly average and avoid production during hours with very negative electricity wholesale prices. If the electricity price at the market is higher than the fixed feed-in tariff, the operators of the renewable energy installations are allowed to keep this difference. This feature distinguishes this market premium mechanism from the instrument called "Contract for Difference (CFD)", where power generators have to pay back the positive difference between the market price and the feed-in tariff (also called the strike price; see Sect. 8.6).

An instrument that seems to have even higher compatibility with markets is a so-called quota obligation combined with a system to trade **green certificates**. Here a central institution sets a target concerning renewable energies, e.g. a minimum of MWh or a particular share of total electricity production that stems from a specific

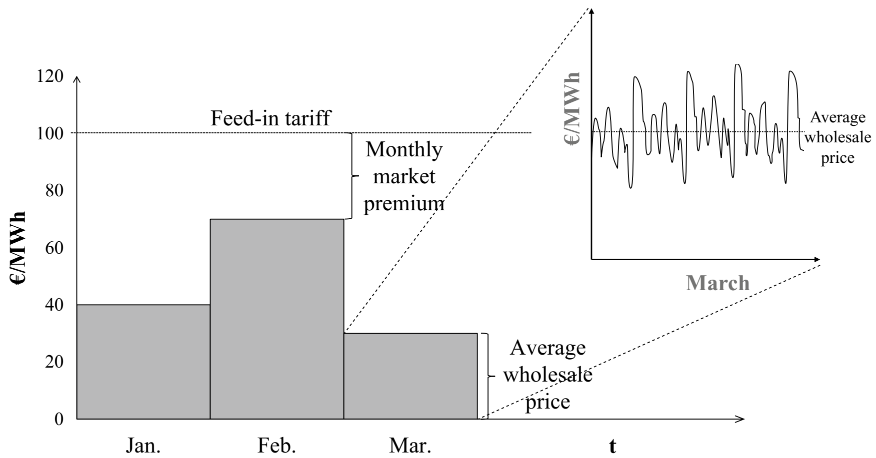


Fig. 6.15 Calculation of the monthly market premium in Germany

renewable energy technology or all renewable energy sources.[25] This quota then has to be fulfilled by each supplier of electricity. Therefore, the renewable electricity producers get certificates according to the number of MWh produced. These green certificates can be sold to suppliers (retailers), who use them to prove compliance with the required quota. Furthermore, certificates can be sold or bought via a secondary market, so the market determines the price for green certificates. This again leads to different revenue streams: besides the electricity market, a certificate market emerges, which might lead to new revenue streams, notably for companies having excess renewable certificates. In principle, it is also possible to install a quota obligation without the possibility for trading, yet this will typically lead to inefficient results since marginal procurement costs for renewables will not level out (see the argument on environmental command and control policies made in Sect. 6.2.3). The apparent advantage of this instrument of higher compatibility with markets might come at the expense of an additional risk premium producers of renewables are requiring due to the higher risk to recover their investments (cf. Haas et al. 2011).

A possibility to determine the financial support needed in a competitive way is the establishment of **procurement auctions** (for detailed information see IRENA and CEM 2015). To realise this, the government has to fix the additional electricity production in renewable energy installations or the capacity to be installed within a certain period and issue a call for tender. Depending on the governmental objectives, the auctions may be implemented as technology-neutral or technology-specific auctions. Pre-qualified market players are allowed to submit bids concerning the remuneration they need to realise their project. Finally, the auctioneer identifies the winning bids, normally the bids requiring the lowest financial support.

Often setting up one of these policy instruments is supplemented by additional instruments, such as tax exemptions, investment aids, information campaigns and low-interest loans (cf. Held et al. 2014, p. 82). Sometimes the market players want to avoid governmental interference and voluntarily agree to realise certain investment or production targets (so-called voluntary agreements).

Another strategy might be to take advantage of consumers' willingness to pay a surplus amount for electricity produced from renewable energy sources. This can be realised with the help of special tariffs (green tariffs), which ensure that the customers' electricity demand is (totally or at least to a certain percentage) covered by electricity from renewable energy sources. To prove that the consumed electricity stems from renewable energy sources, Guarantees of Origin (GoO) have been put in place.

For all financial support mechanisms, the financing of the difference between the remunerations paid to the producers of electricity from renewable energy sources and the electricity market prices has moreover to be decided. This could be done out of the general government budget or with the help of a levy, which electricity

[25] If the target with regard to renewable energies is given for all renewables together, the support scheme is called technology-neutral, which can lead to high profits for the producers of renewables if the renewable cost curve is rather steep (cf. e.g. Haas et al. 2011).

consumers have to pay via their retail price. In turn, such a levy leads to some distortions in competition, both between electricity and other energy carriers, and between domestic electricity users and international ones.

6.2.4.3 Interference Between Emission Trading and Renewable Support

In the energy sector, different political objectives exist, e.g. concerning environmental protection. Diverse policy instruments are sometimes put into place to reach these objectives, leading to a complex mix of instruments. Occasionally, different instruments are even deployed for one political objective. This seems to contradict the design rule, often referred to as the Tinbergen Rule: only one instrument should be used to reach one policy objective – in fact, Tinbergen stated that there have to be as many policy instruments as policy targets (cf. Tinbergen 1952).

With different instruments in place, interferences between them may occur. This is exemplarily shown subsequently for a (simplified) situation, where in a region (or a sector) first a CO_2 emissions trading system has been established and then a support scheme for renewable energies is introduced. In this setting, the support for renewable energies leads to more CO_2-free electricity produced in renewable energy units. But this does not necessarily lead to less CO_2 emissions in the region as the CO_2 emissions are limited by the fixed cap of the emissions trading system. This means that the reduction at one location within the system may induce increases at another site (see waterbed effect in Sect. 6.2.4.1). More electricity from renewable energy sources implies that less electricity has to be produced by other technologies, but the same number of emission allowances is still available (vertical line S in Fig. 6.16). In other words, the demand curve for allowances (line D) is shifted to the left by the emission avoidance E due to renewables and accordingly, the price for emission allowances decreases (see Fig. 6.16). At this point, it is to be

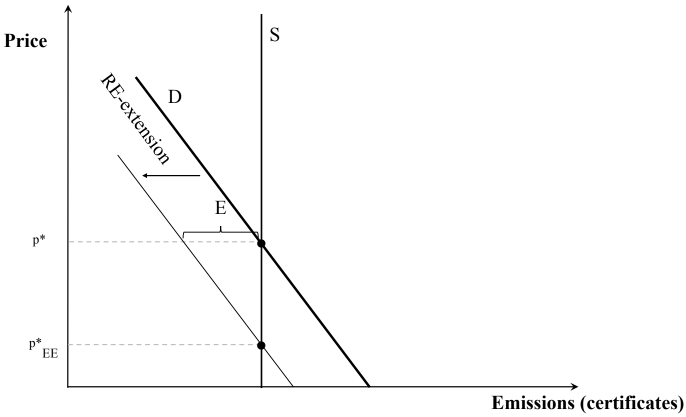


Fig. 6.16 Effects of renewable support schemes on an emissions trading scheme. *Source* Own illustration based on Marquardt (2016)

stressed that the same effect appears when a support scheme for other CO_2-free technologies (e.g. nuclear energy) or a phase-out for coal power plants is introduced in a region already having a CO_2 emissions trading system in place. Finally, it should be mentioned that to ensure that an additional CO_2 emission reduction is realised by introducing a support scheme for renewable energies in a system, which already has a CO_2 emissions trading scheme, the cap of the CO_2 emissions trading system has to be reduced as soon as the support scheme for renewable energies is put in place.

6.3 Further Reading

Varian, H. (2014). Intermediate Microeconomics. 9th edition. New York: W. W. Norton & Company, 2014.

This textbook gives an extensive overview of microeconomics, including case studies and examples.

Jamasb, T., & Pollitt, M. (2000). Benchmarking and regulation: international electricity experience. Utilities Policy, 9, 107–130.

This paper provides manifold information about incentive-based regulation and the used benchmarking methods.

Fritsch, M. (2018). Marktversagen und Wirtschaftspolitik – Mikroökonomische Grundlagen staatlichen Handelns. 10th edition. München: Vahlen.

The book Marktversagen und Wirtschaftspolitik provides a comprehensive presentation of different forms of market failure (e.g. due to external effects and market power) and possible countermeasures.

Perman, R., Ma, Y., McGilvray, J., & Common, M. (2003). Natural Resources and Environmental Economics. 3rd edition. Essex: Pearson Education Limited.

The book Natural Resources and Environmental Economics gives an extensive introduction into natural resources and environmental economics. In the context of power economics, especially the second part of the book dealing with environmental pollution is very relevant.

Tan, Z. (2014). Air Pollution and Greenhouse Gases – From Basic Concepts to Engineering Applications for Air Emission Control. Singapore: Springer.

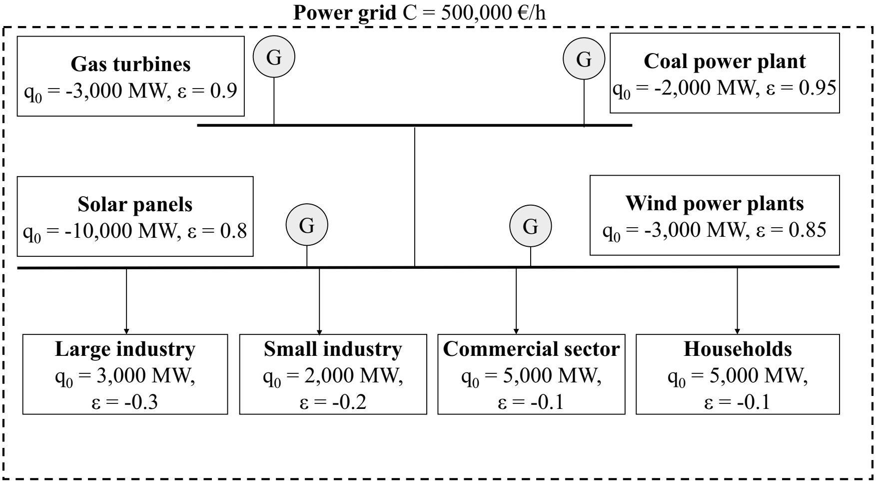
In contrast to the other books mentioned in this section, the book Air Pollution and Greenhouse Gases provides a much more technical perspective. The book presents insights into combustion processes, emissions, and emission control.

6.4 Self-check of Knowledge and Exercises

Self-check of Knowledge

1. What is meant by the technical term "subadditive costs"?
2. Distinguish between the different variants of unbundling.
3. Explain the differences between price cap regulation and revenue cap regulation.
4. Formulate the objective function and the restrictions of the RAINS-model.
5. The firing of which fossil fuels leads to which pollutants?
6. Name the CO_2 emission factors of the different fossil fuels.
7. Which NO formation mechanisms do you know?
8. Name two emission reduction technologies for CO_2, SO_x and NO_x.
9. Which criteria are used to assess environmental policy instruments?
10. Use these criteria to assess a CO_2 tax and a CO_2 emissions trading approach.
11. Compare feed-in tariffs for renewables with a quota obligation combined with a system to trade green certificates.
12. Why might it be difficult for a European country to fulfil its own CO_2-reduction target if energy-intensive companies from this country are included in the European ETS?

Exercise 6.1: Network Pricing
You are the owner of the illustrated power grid with total grid costs of 500,000 €/h on average (see the dimensioning capacities and the price elasticities in the figure). Which (uniform) grid fee would the market participants have to pay in the second-best solution (price equals average costs) if the price elasticities of all producers and consumers are not considered? How will this change if the given price elasticities are considered and the average wholesale price of 4 Cent/kWh is used as reference costs (please use the spreadsheet contained in the appendix to this book)?



Exercise 6.2: Emissions of Power Plants
Calculate the yearly emissions of CO_2, SO_2 and NO_x of a 750 MW hard coal power plant with 7500 full-load hours, an efficiency of 40% and the following emission factors: SO_2: 60 kg/TJ and NO_x: 50 kg/TJ.

Exercise 6.3: Effects of Emission Costs on Production and Investment Decisions
Your company plans to invest in a new CCGT. Calculate the yearly production costs (in €ct/kWh) using the techno-economic data regarding investment, O&M and fuel costs presented in Chap. 4, assuming 5,000 full-load hours, a CO_2 allowance price of 25 €/t and an interest rate of 10%. How would your bid in a competitive day-ahead market with a clearing price auction look like? How do these results change if the government introduces a free of charge allocation of emission allowances for the first 5 years of operation?

References

Alcamo, J., Shaw, R. W., & Hordjik, L. (1990). *The RAINS model of acidification. Science and strategies in Europe*. Kluwer Academic Publishers.

Allen, M., Dube, O., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S., Kainuma, M., Kala, J., Mahowald, N., Mulugetta, Y., Perez, R., Wairiu, M., & Zickfeld, K. (2018). Framing and context. In V. Masson-Delmotte, P. Zhai, H. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. Matthews, Y. Chen, X. Zhou, M. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.), *Global warming of 1.5°C*. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty (pp. 49–91). Available at: https://www.ipcc.ch/sr15/ [Accessed June 2, 2022].

Ardone, A. (1999). *Entwicklung einzelstaatlicher und multinationaler Treibhausminderungsstrategien für die Bundesrepublik Deutschland mit Hilfe von optimierenden Energie- und Stoffflußmodellen*. Peter Lang.

Averch, H., & Johnson, L. (1962). Behavior of the firm under regulatory constraint. *The American Economic Review, 52*, 1052–1069.

Baumbach, G. (1990). *Luftreinhaltung; Entstehung, Ausbreitung und Wirkung von Luftverunreinigungen - Meßtechnik, Emissionsminderung und Vorschriften*. Springer.

Baumol, W., & Oates, W. (1971). The use of standards and prices for protection of the environment. *The Swedish Journal of Economics, 73*, 42–54.

Baumol, W., Panzar, J., & Willig, R. (1982). *Contestable markets and the theory of industry structure*. Harcourt Brace Jovanovich.

Böcker, B., Kippelt, S., Weber, C., & Rehtanz, C. (2018). Storage valuation in congested grids. *IEEE Transactions on Smart Grid, 9*, 6742–6751. https://doi.org/10.1109/TSG.2017.2721982

Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. Springer.

Cansier, D. (1993). *Umweltökonomie*. G. Fischer.

CEER. (2015). *CEER benchmarking report 5.2 on the continuity of electricity supply*. Available at: https://www.ceer.eu/documents/104400/-/-/cbc48e6a-5d5e-a170-ae1d-7b7b298d46a4 [Accessed May 27, 2022].

Coase, R. (1960). The problem of social cost. *Journal of Law and Economics, 3*, 1–44.

Dales, J. (1968). *Pollution, property & prices: An essay in policy-making and economics*. University of Toronto Press.

Dyckhoff, H., & Spengler, T. (2010). *Produktionswirtschaft: Eine Einführung* (3rd ed.). Springer.

European Commission. (2018). *EU emissions trading system (EU ETS)*. Available at: https://ec.europa.eu/clima/policies/ets_en [Accessed December 19, 2018].

ExternE. (2018). *General methodology/The impact pathway approach*. Available at: http://www.externe.info/externe_d7/?q=node/46 [Accessed December 19, 2018].

Fais, B. (2015). *Modelling policy instruments in energy system models—The example of renewable electricity generation in Germany*. IER Forschungsbericht 121 and Dissertation: University of Stuttgart.

Feess, E., & Seeliger, A. (2013). *Umweltökonomie und Umweltpolitik* (4th ed.). Vahlen.

Finon, D., & Menanteau, P. (2003). The static and dynamic efficiency of instruments of promotion of renewables. *Energy Studies Review, 12*, 53–82.

Flachsland, C., Edenhofer, O., Jakob, M., & Steckel, J. (2008). *Developing the international carbon market. Linking options for the EU ETS*. Potsdam Institute for Climate Impact Research. Available at: https://www.pik-potsdam.de/en/news/latest-news/archive-news/files/pik_carbon-market-linking_2008 [Accessed May 27, 2022].

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., & Van Dorland, R. (2007). Changes in atmospheric constituents and in radiative forcing. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, & H. Miller (Eds.), *Climate change 2007: The physical science basis* (pp. 129–234). Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. Available at: https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter2-1.pdf [Accessed June 3, 2022].

Friedrichsen, N. (2015). Governing smart grids: The case for an independent system operator. *European Journal of Law and Economics, 39*, 553–572.

Fritsch, M. (2018). *Marktversagen und Wirtschaftspolitik - Mikroökonomische Grundlagen staatlichen Handelns* (10th ed.). Vahlen.

Haas, R., Resch, G., Panzer, C., Busch S., Ragwitz, M., & Held, A. (2011). Efficiency and effectiveness of promotion systems for electricity generation from renewable energy sources – Lessons from EU countries. *Energy, 36*, 2186–2193.

Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.

Held, A., Ragwitz, M., Gephart, M., de Visser, E., & Klessmann C. (2014). *Design features of support schemes for renewable electricity*. ECOFYS. Available at: https://ec.europa.eu/energy/sites/ener/files/documents/2014_design_features_of_support_schemes.pdf [Accessed May 27, 2022].

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K., Engelbrecht, F., Guiot, J., Hijioka, Y., Mehrotra, S., Payne, A., Seneviratne, S., Thomas, A., Warren, R., & Zhou, G. (2018). Impacts of 1.5°C global warming on natural and human systems. In V. Masson-Delmotte, P. Zhai, H. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. Matthews, Y. Chen, X. Zhou, M. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.), *Global warming of 1.5°C* (pp. 175–311). An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Available at: https://www.ipcc.ch/sr15/ [Accessed June 2, 2022].

Horowitz, J., & McConnell, K. (2003). Willingness to accept, willingness to pay and the income effect. *Journal of Economic Behavior & Organization, 51*, 537–545.

Hupa, M., Backman, H., & Boström, S. (1989). Nitrogen oxide emissions of boilers in Finland. *JAPCA, 39*(11), 1496–1501. https://doi.org/10.1080/08940630.1989.10466644

Icha, P., AGEE-Stat am UBA, & Kuhs, G. (2019). *Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990–2018*. Climate Change 10/2019. Umweltbundesamt.

IEA. (2019). *More of a good thing—Is surplus renewable electricity an opportunity for early decarbonisation*? Available at: https://www.iea.org/commentaries/more-of-a-good-thing-is-surplus-renewable-electricity-an-opportunity-for-earlydecarbonisation [Accessed May 28, 2022].

IPCC. (2014). *Climate change 2014: synthesis report*. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R. K. Pachauri, & L. A. Meyer (Eds.)]. IPCC. Available at: https://www.ipcc.ch/report/ar5/syr/ [Accessed June 03, 2022].

IRENA, & CEM. (2015). *Renewable energy auctions—A guide to design*. Available at: https://www.irena.org/publications/2015/Jun/Renewable-Energy-Auctions-A-Guide-to-Design [Accessed May 27, 2022].

ISO14040. (2006). *Environmental management—Life cycle assessment—Principles and framework*. International Organization for Standardization. Available at: https://www.iso.org/standard/37456.html [Accessed May 27, 2022].

ISO14044. (2006). *Environmental management—Life cycle assessment—Requirements and guidelines*. International Organization for Standardization. Available at: https://www.iso.org/standard/38498.html [Accessed May 27, 2022].

Joskow, P. (2007). Regulation of natural monopoly. In A. Polinsky & S. Shavell (Eds.), *Handbook of law and economics* (Vol. 2, pp. 1227–1348). North Holland.

Juhrich, K. (2016). *CO_2 emission factors for fossil fuels*. German Environment Agency (UBA). Climate Change 28/16. Available at: https://www.umweltbundesamt.de/publikationen/co2-emission-factors-for-fossil-fuels [Accessed May 27, 2022].

Kaltschmitt, M., & Jorde, K. (2007). Hydroelectric power generation—Economic and environmental analysis. In M. Kaltschmitt, W. Streicher, & A. Wiese (Eds.), *Renewable energy—Technology, economics and environment* (pp. 373–383). Springer.

Kaltschmitt, M., Schröder, G., & Schneider, S. (2007). Wind power generation—Economic and environmental analysis. In M. Kaltschmitt, W. Streicher, & A. Wiese (Eds.), *Renewable energy—Technology, economics and environment* (pp. 339–348). Springer.

Khartchenko, N. (1997). *Umweltschonende Energietechnik*. Vogel.

Kjolle, G. H., Samdal, K., Singh, B., & Kvitastein, O. (2008). Customer costs related to interruptions and voltage problems: Methodology and results. *IEEE Transactions on Power Systems, 23*, 1030–1038.

Kuosmanen, T., Saastamonien, A., & Sipilainen, T. (2013). What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED methods. *Energy Policy, 61*, 740–750.

Laffont, J.-J., & Tirole, J. (1993). *A theory of incentives in procurement and regulation*. MIT Press Books.

Littlechild, S. (1983). *Regulation of British telecommunications' profitability*. Report to the Secretary of State. Department of Industry.

Lyon, T. P. (1996). A model of sliding-scale regulation. *Journal of Regulatory Economics, 9*, 227–247.

Marquardt, R. (2016). Hebt das Emissionshandelssystem die Einsparungen des Ökostromausbaus auf? *Wirtschaftsdienst, 7*, 515–523.

Mathieu, P. (2010). Oxyfuel combustion systems and technology for carbon dioxide (CO_2) capture in power plants. In M. Maroto-Valer (Ed.), *Developments and Innovation in carbon dioxide (CO_2) capture and storage technology* (pp. 283–319). Woodhead Publishing.

Matthews, H., Hendrickson, C., & Matthews, D. (2014). *Life cycle assessment: Quantitative approaches for decisions that matter*. Open access textbook. Available at: https://www.lcatextbook.com/ [Accessed May 28, 2022].

Matthews, J. (2018). Annex I: Glossary. In V. Masson-Delmotte, P. Zhai, H. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. Matthews, Y. Chen, X. Zhou, M. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.), *Global warming of 1.5°C* (pp. 541–562). An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Available at: https://www.ipcc.ch/sr15/ [Accessed June 2, 2022].

Mechler, R., Pulido-Velazquez, M., Koehler, M., Jenkins, K., Surminski, S., Williges, K., Botzen, W., Cremades, R., Dransfeld, B., Hudson, P., Lopez-Nicolas, A., Lorant, A., Manez, M., & Michaelowa, A. (2016). Economic instruments. In J. Aerts & J. Mysiak (Eds.), *Novel multi-sector partnerships in disaster risk management* (pp. 69–97). Results of the ENHANCE project. EU FP7 project ENHANCE.

Perino, G. (2018). New EU ETS phase 4 rules temporarily puncture waterbed. *Nature Climate Change, 8*, 262–264.

Perman, R., Ma, Y., McGilvray, J., & Common, M. (2003). *Natural resources and environmental economics* (3rd ed.). Pearson Education Limited.

Pigou, A. C. (1920). *The economics of welfare*. Macmillan & Co.

Richers, U., & Günther, V. (2014). *Erweiterte Überlegungen zur integrativen Vermeidung von Umweltauswirkungen am Beispiel der Stickoxidemissionen aus Abfallverbrennungsanlagen*. KIT Scientific Reports 7680. KIT Scientific Publishing.

Schmalensee, R. (1989). Good regulatory regimes. *The RAND Journal of Economics, 20*, 417–436.

Schober, D., Schaeffler, S., & Weber, C. (2014). The impact of idiosyncratic risk on the cost of capital—The case of firm size for regulated electricity networks. *Journal of Regulatory Economics, 46*, 123–151.

Shivakumar, A., Welsch, M., Taliotis, C., Jakšić, D., Baričević, T., Howells, M., Gupta, S., & Rogner, H. (2017). Valuing blackouts and lost leisure: Estimating electricity interruption costs. *Energy Research and Social Science, 34*, 39–48.

Shleifer, A. (1985). A theory of yardstick competition. *The RAND Journal of Economics, 16*, 319–327.

Tan, Z. (2014). *Air pollution and greenhouse gases. From basic concepts to engineering applications for air emission control*. Springer.

Tinbergen, J. (1952). *On the theory of economic policy*. North Holland Publishing Company.

Turconi, R. (2014). *Life cycle assessment of electricity systems*. DTU Environment and Dissertation: Technical University of Denmark.

Varian, H. (2014). *Intermediate microeconomics* (9th ed.). W. W. Norton & Company.

Viscusi, W., Harrington, J., & Vernon, J. (2005). *Economics of regulation and antitrust* (4th ed.). The MIT Press.

Wallner, K., Glock, D., Runge, P., Tschach, I., & Ruf, P. (2014). *Analysis and assessment of market structure, trading activities and further developments in the EU ETS*. Project No. (FKZ) 3713 41 504. German Emissions Trading Authority (DEHSt).

Walz, R., Toussaint, D., Küpfer, C., & Sanden, J. (2009). *Gestaltung eines Modells handelbarer Flächenausweisungskontingente unter Berücksichtigung ökologischer, ökonomischer, rechtlicher und sozialer Aspekte*. Umweltbundesamt. Available at: https://www.umweltbundesamt.de/sites/default/files/medien/publikation/long/3839.pdf [Accessed May 28, 2022].

Weber, C., & Vogel, P. (2014). Contingent certificate allocation rules and incentives. *Journal of Regulatory Economics, 46*, 292–317.

Weitzman, M. (1974). Prices vs. quantities. *The Review of Economic Studies, 41*, 477–491. https://doi.org/10.2307/2296698

7 Simple Electricity Market Equilibrium Models

Competition has been introduced in electricity systems only at the end of the 1980s given the peculiarities of the sector. To better understand these markets, specific models are needed to depict market equilibria in electricity systems in the short and long run. In this chapter, several graphical and formal models are introduced to investigate the functioning of electricity markets. Against this background, the following chapter aims at answering the following key questions:

- What are peculiarities of the sector?
- How can the standard model of supply and demand be adapted to the electricity market?
- How can markets and the resulting prices be represented in technoeconomic models?
- How can the model be expanded to include multiple regions and load flows?
- What is the difference between short-term and long-term market equilibrium?

Subsequently, we address the basics of market equilibria in electricity systems in the short and long run. In the following, several graphical and formal models describing electricity markets are introduced, which we believe are helpful in understanding the functioning of electricity markets. These models especially describe the functioning of the wholesale electricity markets. The retail market or final consumer market is in general not addressed by this kind of models. The first model addresses the short-term market equilibrium without any transmission constraints—it is also known as the "merit-order model" (Sect. 7.1). A model with a very stylised, single transmission constraint follows in Sect. 7.2. In Sect. 7.3, the inclusion of the linearised transmission constraints from Sect. 5.1 in the short-term market equilibrium model is discussed. The fourth model finally describes the long-term equilibrium in electricity markets, including investments in generation capacities (Sect. 7.4). This model is also sometimes referred as the peak load pricing model.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_7

Key Learning Objectives
After having gone through this chapter, you will be able to

- Describe graphically as well as mathematically the standard merit-order model of electric power systems.
- Explain the merit-order effect.
- Formulate a fundamental electricity market model and explain the difference between long- and short-term market equilibrium.
- Describe how power flows are modelled in bottom-up electricity market models.
- Formulate the nodal pricing model.
- Know how to calculate power transmission distribution factors (PTDFs) in a given electricity network.
- Explain how capacities are determined in the long-term market equilibrium model.

7.1 Short-Term Market Equilibrium Without Transmission Constraints

In the standard microeconomic textbook model of perfect competition, market equilibrium corresponds to the point at which quantity demanded and quantity supplied are equal. The market price results from market equilibrium. This price is also called the market clearing price. The basic textbook model can also be extended to markets with imperfect competition, where players act strategically. Then supply curves do not purely reflect marginal cost (and possibly demand curves do not reflect marginal utility). These types of markets are addressed in Chap. 9, while in this chapter the focus is laid on modelling perfect competition. Subsequently, we start by first describing the short-term equilibrium verbally and graphically using the assumption that markets are perfectly competitive (Sect. 7.1.1). We then discuss the applicability of the conditions for perfect competition in the electricity markets (Sect. 7.1.2), followed by a mathematical description of the market equilibrium resulting from an optimisation model (Sect. 7.1.3). Finally, a small application to the case of Germany is discussed in Sect. 7.1.4.

7.1.1 Simple, Graphical Approach: Merit-Order Model

Electricity has several **specificities** compared to other goods and other energy carriers (see Sect. 2.5), which must be considered in the standard short-term **market equilibrium model**:

- **Non-storability**: electricity cannot be stored, at least not in large amounts (see Sect. 5.2). This comes with the challenge of ensuring a permanent equilibrium of electricity supply and demand. Hence, the market for final physical delivery has to take this fact into account. Accordingly, there are different market prices for different periods of time.
- **Grid-bound**: electricity can only be transported in electricity networks (see Sect. 5.1). This necessitates the consideration of physics in the economic models, at least when congestions occur. In a very stylised form, this is discussed in Sect. 7.2. The more elaborated version is contained in Sect. 7.3.
- **Non-elasticity of demand in the short-term**: in the short term, many consumers do not see real-time wholesale market prices (see Sect. 3.1.6). As a result, no incentives exist to react to market prices in the short-term. The lack of incentives means that any (potential) elasticity is not exploited, resulting in a short-term inelastic demand of households. Industrial and commercial customers seeing real-time prices have only a very limited elasticity due to an extremely limited substitutability of electricity. In consequence, electricity demand is very price inelastic in the short-term.

Moreover, the supply function in the electricity market may be generally described as a stepwise function, since marginal costs of single plants can be assumed to be nearly constant and since different technologies have different levels of marginal costs depending on the fuel used. As capacities, fuel costs and efficiencies of plants are generally well-known, the empirical shape of the supply curve may usually be estimated rather accurately (see Fig. 7.1). This supply curve is also called the **supply stack** or the merit order because the ordering of the power plants is done by increasing marginal cost. The approach is similar to the simple scheduling approach a single operator would use in the absence of operational restrictions (see Sect. 4.4.1.1).

Market equilibrium then occurs at the point at which the demanded and the supplied quantity are equal.[1] Due to the inflexibility of demand and the lack of storability, different prices result from varying demand levels over time – if the supply stack is time-invariant (cf. below for the extension to time-variable supply).

This simple model illustrates the short-term market equilibrium without transmission restrictions and obviously can represent power markets with one centrally operated (day-ahead) spot market.[2] Thereby, traders (representing demand and supply) submit their bids resulting in the aggregated demand and supply curves and the market operator determines the market prices based on the offered and demanded quantities. With inelastic demand, the main characteristic of the model is

[1] That both the approach and the obtained market equilibrium are similar to the optimal dispatch obtained by a portfolio manager for a power plant portfolio as discussed in Sect. 4.4.1.1 is not a coincidence. It rather illustrates the general microeconomic result that market outcomes under perfect competition are equal to the optimal planning made by a perfectly informed central planner – i.e. a system portfolio manager in our case.

[2] In Sect. 8.5.1, we introduce the law of one price which justifies the use of this model even for the case of decentralised (e.g. bilateral) trading.

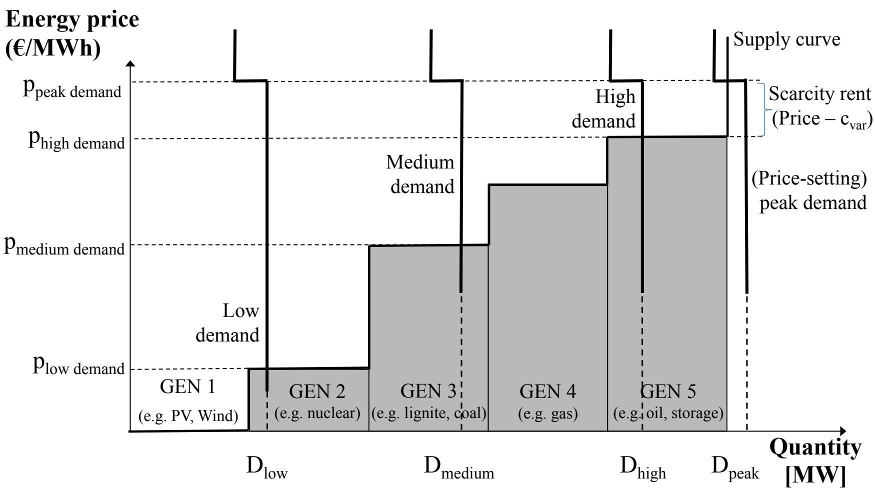


Fig. 7.1 Stylised supply curve and intersection with demand curves at different demands

the stepwise supply curve; therefore, the model is often called the merit-order model. Under the assumptions of perfect competition (cf. below), generators submit their bids based on their marginal costs. In general, marginal costs of generators are derived from their variable costs, mainly fuel costs and possibly emission certificate costs (cf. Sect. 4.3.2). Hence, technologies with no variable costs, such as wind or PV plants, are on the left side of the supply curve, while generators with relatively high marginal costs are at the end of the supply curve, notably oil and gas turbines. In cases where capacity is getting scarce, price-sensitive customers at high prices may be price-setting and markups above marginal costs may occur, resulting in scarcity rents. This phenomenon may be the consequence of a small price elastic demand as depicted in Fig. 7.1 in combination with scarce capacity (see Sect. 7.4), or it may result from the exercise of market power (cf. Chap. 9). So these price markups require a more detailed reflection both from a theoretical viewpoint and in view of practical market design issues.

For more descriptive and forecasting purposes, the assumption of an entirely inflexible demand may be less problematic. Then it is possible to derive the price curve (of a day or year) graphically from the simple merit-order model as illustrated in Fig. 7.2. On the left side of the figure, a typical demand curve of a whole day is depicted in the fourth quadrant of the diagram. Each hourly demand results in a price of a day. In this illustrative model, it is assumed that the merit-order curve is time-invariant over the day. The first quadrant shows the merit-order curve and in the second quadrant, the resulting price curve is illustrated. The third quadrant helps to transform the time axis from the fourth to the second quadrant. On the right side of the figure, the link between demand and prices is illustrated based on probability density functions. Hence, the figure shows nearly the same, but now yearly demand

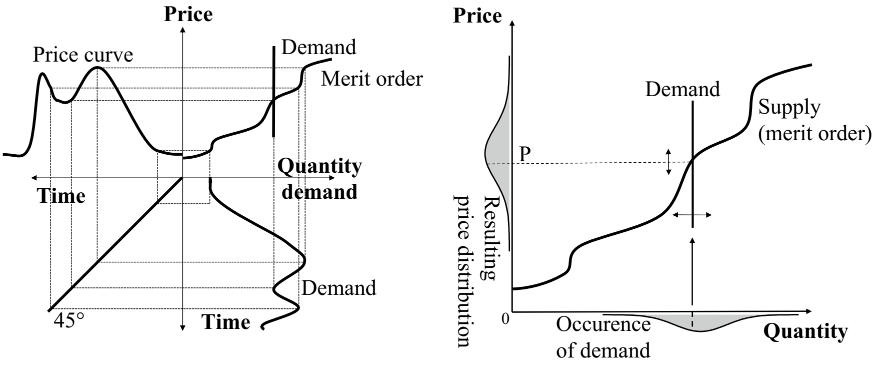


Fig. 7.2 From demand curve with the stylised merit-order model to the price curve

is considered and how it translates into a distribution of prices when a simple graphical merit-order model is used.

The merit-order model may be slightly modified to cope with the time-variable feed-in from renewables like wind and solar. Since these renewables enter the market in general with zero marginal costs, their production will always be used.[3] In the graphical approach, two possibilities for consideration exist: they can be drawn on the left side of the merit order (as depicted in Fig. 7.1) or subtracted from the demand side instead of including them on the supply side. The supply curve then encompasses only the conventional units (and possibly the controllable renewables like biomass and hydropower), while the demand reflects the so-called **residual load**, also called **net demand**.

In this simple merit-order model, operating constraints of the generators are not considered. Besides the typical variable costs (operating and fuel costs), the operating restrictions discussed in Sect. 4.4, as well as the following cost components and constraints, can affect the individual bids of generators:

- **Availability**: in reality, capacities may be unavailable due to either planned maintenance or unplanned outages.
- **Load changes**: frequent generation changes of power plants may influence maintenance intervals of plants due to increased wear and tear and may hence be included in the supply bids.
- **Start-up/ramp-up**: starting a fossil power plant, especially from a cold start-up, requires additional fuel for heating the plant. If the plant is run only for a few hours, these costs for heating the plant (before generating power) are generally considered in the generators' bids as start-up costs. This argument is also valid for power plants that are already generating electricity but would be switched off due to low prices in one of the following hours. Hence, a generator may bid below marginal costs as long as the potential loss is lower than shut-down costs.

[3] Except when the renewable production exceeds demand.

- **Minimum operation and downtime**: minimum operation times and (planned and unplanned) downtimes have to be considered.
- **Heat production** (from combined heat and power plants) or restrictions from other markets, such as ancillary services, may force generators to bid below marginal costs. This is also one explanation for negative prices in European electricity markets.
- While capacities of **storage power plants** may be included (cf. Sect. 7.1.4), they are often not considered in this simple merit-order model. Moreover, operational restrictions of storage power plants resulting notably from limited storage volume cannot be considered.

These aspects show the limit of this simple graphical model and have to be considered when discussing model results.

7.1.2 Assumptions Underlying the Concept of Perfect Competition

Before pushing further, the **assumptions** underlying the **perfect competition** model deserve a bit more attention.[4] First, it is worth noting that if the electricity market is perfectly competitive, the market equilibrium is **Pareto efficient**, meaning that none of the market players can be made better off by government interventions or other means without making someone else worse off.

Yet several conditions have to hold for the model of perfect competition to be applicable. We review them in light of the current European electricity markets:

- **A large number of buyers and sellers**: a lot of utilities act on the European wholesale electricity markets, which are quite well interconnected. Market concentration in some countries is high at first sight (e.g. France), but with European competition in a meshed electricity system, single utilities' impact is generally limited. Moreover, European regulation aims to lower barriers to market entry, and new market players (notably based on renewable energies) are entering the electricity market. The number of market players is continuously increasing.
- **Homogenous products**: products on the electricity exchange are well defined and product quality is specified by technical norms (e.g. frequency, voltage level, etc.) and hence does not vary. Consequently, electricity at the wholesale electricity market is a homogenous product.
- **Perfect information**: as the product variety is very limited and data availability for electricity markets is relatively high, all market players have detailed market information. However, due to changes in market design (e.g. introduction of flow-based market coupling), some additional data needs may arise, and thus a certain delay in relevant market data may sometimes occur.

[4] Cf. also the discussion of the more realistic concept of "workable competition" in Sect. 9.5.

- **No barriers to entry or exit**: with the unbundling of electricity markets, regulatory barriers to enter or exit the market do not exist anymore as every investor can enter the market with new generation capacities and exit the market by mothballing or decommissioning capacity. However, due to the time lag of up to 2–8 years between investment decision and the entry into service of a plant, there may be scarce capacities for a limited time period explaining markups on competitive prices due to scarce capacities. Yet this can be seen as a market signal to provide additional capacities.
- **Every participant** is a **price taker** or no participant has market power to set prices: in non-peak time segments, many buyers and sellers are active in the market and all market participants are (in general) price takers. When capacity might get scarce in a peak load situation, prominent players can become pivotal and may not act as price takers but instead set prices with a markup over marginal cost. However, significant price markups are also necessary for long-term equilibrium (cf. below). They are also observed in the simple merit-order model of Fig. 7.1, when a slightly price elastic demand curve hits the supply curve at the capacity boundary. Hence, a simple observation of markups of prices over (short-run) marginal cost does not necessarily prove the exercise of market power, especially if market entry barriers are low. In some electricity markets, e.g. the control reserve markets (see Sect. 10.3), this assumption may not always be valid as some prominent players may have a dominant role. But with further growth of small and decentralised generators and storage capacities (and their further integration into the markets), the markets are becoming increasingly competitive.
- **Profit maximisation of sellers and rational buyers**: after unbundling and as private investors own utilities, they (have to) act as profit maximisers, which is sometimes also criticised by society and politicians.[5]
- **No externalities**: unfortunately numerous externalities exist in energy markets (see also Sect. 6.2). Some of these externalities are (partially) internalised, such as CO_2 emissions with the help of the EU emission trading system and resulting carbon prices. Subsequently, it is yet assumed that all externalities are internalised; hence, there are no remaining externalities.
- **Zero transaction costs**: transaction costs on energy exchanges are very low or even close to zero compared to the financial volume of transactions.

In practice, rarely all of these conditions are met, but the depicted models help to understand the functioning of power markets. The impact of not satisfied conditions – in case there are some – on model results can then be discussed on the basis of the model implementation.

[5] Comparing the situation of power companies to companies from other industries, society and politicians seem to assess financial results differently: financial statements with enormous profits of (for example) large mobile or car manufacturing companies tend to be more tolerated, or are sometimes even hailed in society. Rather seldom is this the case for electric utilities and energy companies, although they are acting as stock-quoted company with the same profit maximisation objective. This (at least perceived) discrepancy merits to be further investigated.

7.1.3 Formal Model

The simple merit-order model can be formulated as a **welfare maximisation problem**. Welfare is calculated as the difference between the consumer utility, derived from electricity consumption (the first integral), and the costs incurred in the system (the second sum):

$$\max_{y_{ut},\overline{D}_t} \sum_t \left\{ \int_0^{D_t} p_t^D(q) \cdot \mathrm{d}q \cdot \Delta t - \sum_u y_{ut} \cdot \Delta t \cdot c_u^{\mathrm{var}} \right\} \tag{7.1}$$

This optimisation is subject to the capacity restrictions of the generation plants

$$0 \leq y_{ut} \leq K_u \qquad \forall u, t \tag{7.2}$$

and the market clearing conditions

$$\sum_u y_{ut} \Delta t \geq D_t \Delta t \quad \Leftrightarrow \quad \sum_u y_{ut} \geq D_t \qquad \forall t \tag{7.3}$$

Herein $p_t^D(q)$ corresponds to an inverse demand function in time period t. For each demand level q, the price p_t describes the marginal **willingness to pay** and hence the marginal utility (in monetary expenditure equivalents) for the last demand unit. Its integral thus describes the consumer utility of electricity consumption D_t in economic terms. Note that demand D_t is expressed as average power over the time interval of length Δt to obtain consistency with generation capacities K_u expressed in power units, e.g. megawatt. Correspondingly also the generation y_{ut} of unit u produced at time t is the average production power, c_u^{var} the variable generation cost of plant u per energy unit. With this simple optimisation model, welfare is maximised using a monotonously decreasing, in most cases linear demand function and a piecewise constant supply function. Graphically, this corresponds to maximising the area obtained when subtracting the cost of generation (area under the supply curve) from the area below the demand function.

Under the assumption that demand is completely inelastic, the above welfare maximisation is equivalent to a cost minimisation problem with exogenously given demand D_t on the right-hand side.[6]

[6] Formulating the welfare maximisation problem as a cost minimisation problem with price inelastic demand significantly reduces computation time, as the cost minimisation may then be formulated as a linear optimisation problem. In contrast, the welfare maximisation with linear demand curves is in general a quadratic constrained optimisation problem (QCP).

The objective function then reads

$$\min_{y_{ut}} \sum_{u,t} y_{ut} \cdot \Delta t \cdot c_u^{\mathrm{var}} \tag{7.4}$$

subject to the capacity and demand restrictions:

$$0 \leq y_{ut} \leq K_u \qquad \forall u, t \tag{7.5}$$

$$\sum_u y_{ut} \Delta t \geq D_t \Delta t \quad \Leftrightarrow \quad \sum_u y_{ut} \geq D_t \qquad \forall t \tag{7.6}$$

In this simple representation of the short-term equilibrium without transmission constraints, variable costs of the last dispatched generator in the merit-order set the price. From a mathematical point of view, the price is obtained as the value of the **dual variable** of the demand restriction. This is often called the shadow price of demand and describes the influence of an infinitesimal demand increase on the objective function. A key outcome of this short-term equilibrium is that the price equals marginal costs as expected from microeconomic theory.

This simple **linear optimisation problem** (**LP**) can be extended by the above-mentioned cost components relevant for power plant operators, such as start-up and shut-down costs, heat restrictions, etc., which leads to more precise representations of electricity markets.[7] Also, the stepwise demand curve as depicted in Fig. 7.1 may be included by considering **demand response** (cf. Sect. 3.1.4) through additional decision variables with associated costs[8] and maximum capacities. These extensions generally increase the computing time for solving the optimisation model, primarily because some extensions result in a mixed-integer optimisation problem (MILP, cf. Sect. 4.4.1.3).

In the basic formulation of the problem, negative marginal costs cannot occur. However, negative prices are common in European power markets. The extensions mentioned in Sect. 7.1.1, e.g. heat or ancillary service market restrictions, combined with a surplus of renewable feed-in, also result in negative prices in this extended model.

7.1.4 Application

In the following, the short-term equilibrium model is applied to a case study for Germany. Figure 7.3 shows a stylised merit-order curve for Germany with the respective capacities and the range of demand. Renewable capacities, especially

[7] A scientific analysis of factors, which are relevant for electricity market modelling, can be found in Martínez-Díaz (2008).

[8] Instead of actual cost, these may also represent the maximum willingness to pay in certain customer segments.

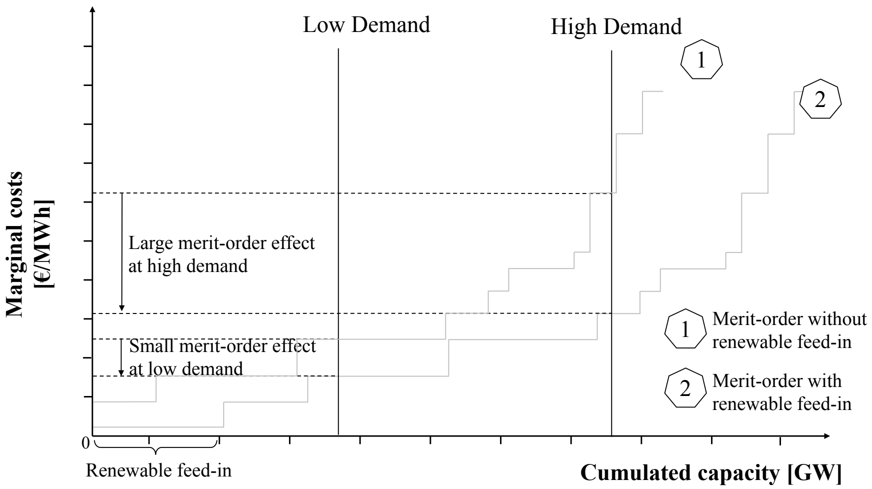


Fig. 7.3 Stylised merit-order curve in Germany and price influencing factors (negative prices are not illustrated but may result from introducing further restrictions in the above scrutinised model)

wind, photovoltaic and run-of-river, are depicted on the left side of the merit-order curve with marginal costs of zero. Even if the total generation costs of renewable electricity (LCOE, see Sect. 4.3.3) are higher than for conventional technologies, their marginal costs are zero when wind or solar energy is available. Correspondingly, the amount of production is essentially determined by the available resources. Wind and solar capacities are steadily increasing in Germany due to the Renewable Energy Sources Act (cf. BMWI 2017). In 2016, nearly 90 GW wind, solar and run-of-river capacities existed, but maximum injection from renewables in 2016 was approximately 42 GW.[9] In selected hours of a year with low demand and a high renewable injection, demand could be fully covered by renewable capacities (despite the fact that still some conventional capacities are needed for reasons of ancillary services and grid stability). Conventional capacities then follow next on the piecewise merit-order curve. The height of the steps in the merit-order curve depends on the fuel type, fuel and CO_2 prices as well as the efficiency of the plants, while the installed capacity determines the width of the step.

In 2016, this stepwise function was continued by

- Nuclear capacities with approximately 11 GW at marginal costs between 5 and 10 €/MWh.
- Lignite with around 20 GW and marginal costs between 13 and 18 €/MWh.
- Coal with approximately 27 GW and marginal costs between 20 and 30 €/MWh.

[9] Source for this value is the data in the Entso-E transparency platform (https://transparency.entsoe.eu/). As wind and solar energy is not available to its maximum all over Germany at the same point in time, the maximum injection from all wind and PV capacities is significantly lower than the installed capacities.

- Gas with around 22 GW and marginal costs between 30 and 60 €/MWh.
- Oil with approximately 3 GW and marginal costs above 50 €/MWh.
- Furthermore, biomass plants are installed with approximately 7 GW and about 6 GW of pump storage capacities operating based on opportunity costs, in general in the peak load segment.

Marginal costs of the conventional power plants mainly depend on the type of fuel and the vintage, as the efficiency of new plants has generally increased over the years. The level of the curve for each fuel type is mainly influenced by variations in fuel and CO_2 prices. In contrast, the horizontal width of the segments is influenced by the availability of plants and in the long-term by decommissioning and the installation of new capacities (cf. Sect. 4.4.1.1, Fig. 4.31). Hence, this simple short-term equilibrium model helps explain and understand the electricity market's functioning and the influencing factors on the power prices.

One influencing factor on power prices is also the feed-in from generators using renewable resources. A higher amount of generation from renewable resources shifts the supply curve to the right, leading to a reduction in power prices. This phenomenon is called the **merit-order-effect of renewables**. In most countries, the merit-order curve is typically flat at the left side and in the middle, while it is very steep on the right side of the curve. The shape of the curve implies that the merit-order effect of renewables depends on the demand level. Consequently, the price decrease from renewable resources feed-in is higher at a high demand than at low demand, as illustrated in Fig. 7.3. Additionally, the price decrease is also higher if a shift from one technology to another is necessary, corresponding to a step in the merit-order curve. From a theoretical point of view, the price decrease due to an infinitesimal increase of feed-in from renewable resources corresponds to the mathematical gradient of the merit-order curve. Especially storage power plants, which profit from the price spread between peak and off-peak, are strongly influenced in their profitability by the merit-order effect of renewables, as it is obvious that for constant conventional capacities, the price spread decreases with a higher feed-in from renewable resources. The phenomenon of decreasing price spreads has been observed in the European energy markets for several years. It is contra-intuitive to weather-dependent and thus volatile feed-in from renewables requiring a higher amount of storage capacities. The relationship between renewables and storage is further discussed in Sect. 7.4.3.

The short-term equilibrium model is widely applied to electricity markets in the form of the above-described optimisation model. Thereby, the basic model is often extended by additional restrictions and price components. In general, the power plant dispatch from the model conforms reasonably well to the dispatch in the real world and also power prices can be rather well explained.

To measure model quality, the model output $\hat{x}_t$ (here the marginal costs of power production) is generally compared with the market observation x_t (e.g. power prices on the spot market) by using the **mean absolute error** (**MAE**), the **root mean square error** (**RMSE**) or the **mean absolute percentage error** (**MAPE**):

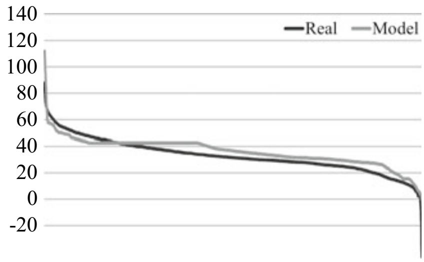


Fig. 7.4 Exemplary price duration curves in €/MWh from market data and model output for Germany in 2015

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T}|x_t - \hat{x}_t|$$

$$\mathrm{RSME} = \frac{1}{T}\sqrt{\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}$$

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{|x_t - \hat{x}_t|}{x_t} \tag{7.7}$$

As data availability in European energy markets is very good, a high model quality can often be achieved with such a type of model. Depending on the country, the data availability and the details included in the model, a mean absolute error of 3–10% of the mean price can be achieved, which is very accurate for so-called fundamental models.[10] Figure 7.4 exemplarily depicts power prices in sorted order for one year compared to an exemplary model output. Besides the above-shown error measures, this allows a graphical illustration of model accuracy. In general, bottom-up models can well explain the **price duration curve**, as depicted in Fig. 7.4. Deviations mainly occur at peak and off-peak times. Negative prices (as shown in the figure) may not be well represented due to a lacking implementation of technical restrictions and an overestimation of flexibility.

The European Power Exchange (EPEX SPOT, cf. Sect. 10.1) provides hourly supply and demand bid curves for each hour of the year. In Fig. 7.5, a supply and demand curve for the market area of Germany and Austria is depicted. However, these supply and demand curves do not correspond precisely to the model studied above. But why? The differences between the above model and the bid curves on the EPEX SPOT can be explained by the fact that trading is not only possible on the day-ahead market. Trading already takes place in the future markets with longer

[10] These models are often called fundamental models or bottom-up models as they just make use of (fundamental) technoeconomic data to describe the supply and demand equilibrium.

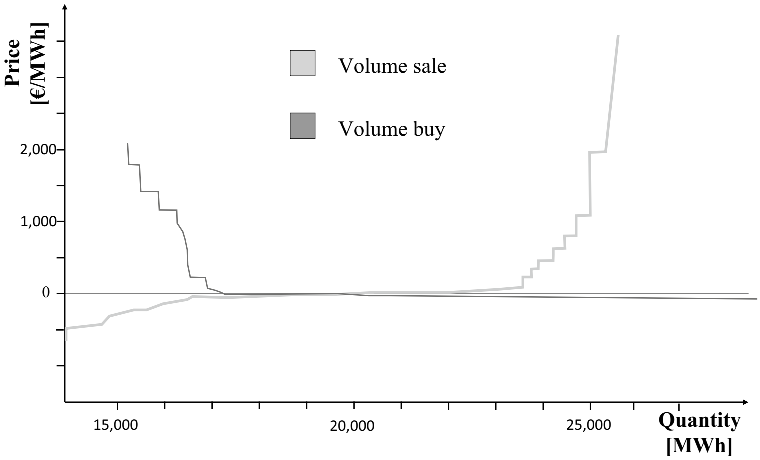


Fig. 7.5 Supply and demand curve for the market area Germany/Austria from EPEX SPOT at 27.10.2015 10:00–11:00

lead-times (cf. Sects. 8.5 and 10.2). Correspondingly, generators do not necessarily appear on the supply side as they may already have sold their electricity months ahead on the future markets. On the spot market, they may now bid on the demand side to reoptimise their portfolio. The same is true for bidders on the demand side, just vice versa. Furthermore, not all quantities are traded on the spot market. These are reasons, why the supply and demand curves resulting from the bids in non-mandatory markets with future trading deviate from the theoretical model (cf. also Sect. 8.2). Further reasons might be deviations from the assumptions about perfect competition listed above. Notwithstanding, the model still has a very high explanatory power, as it can help to understand market outcomes and the functioning of electricity markets.

7.2 Short-Term Market Equilibrium with Two Grid Nodes

The short-term equilibrium model discussed so far only considers supply and demand curves of one country. However, electricity markets from different European countries are interconnected with substantial yet limited transmission capacities and hence an extension of the basic short-term market equilibrium model to a case with transmission constraints is necessary. The short-term **market equilibrium model with two grid nodes** is therefore presented in the following, starting again with the graphical illustration followed by a mathematical description.

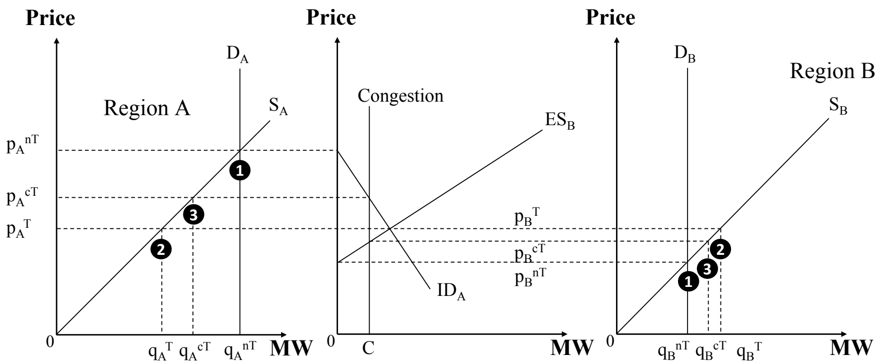
7.2.1 Graphical Model

The short-term equilibrium market model can easily be extended to a model considering several countries or regions taking interconnections between regions and thereby imports and exports into account. This model is often called a "transshipment model" as Kirchhoff's mesh law (cf. Sect. 5.1.2) is neglected. In this model type, electricity can be exported or imported between the considered regions as in a typical transportation model in operations research.

If only two regions are considered, the model may be illustrated as depicted in Fig. 7.6. In principle, three situations may be distinguished: (1) no trade between the two regions, (2) full trade or trade without any congestions, and (3) congested trade between the two regions due to limited transmission capacities.

1. **No trade**: if regions A and B are not interconnected, the two supply and demand curves can be seen as two individual diagrams with no interaction. In this case, market prices are determined (as in the presented short-term equilibrium market model) at the intersection of demand (D_A as demand in region A and correspondingly in B) and supply (S_A as supply in region A and correspondingly in B), resulting in the no-trade (nT) market prices (p_A^{nt} as market prices with no trade in region A and correspondingly in B).

2. **Full trade**: if electricity can be exchanged, region B has an incentive to sell some of its (cheaper) supply to region A, as the price in A is at the outset higher than the cost of un-dispatched supply in B. Correspondingly, region A has an incentive to import from region B, as prices in B are lower than marginal costs of dispatched supply in A. Hence, the resulting export supply of B (ES_B) can be derived from the un-dispatched part of the supply curve in B, starting at the



D_r	Demand curve in region r	ES	Export surplus
S_r	Supply curve in region r	ID	Import deficit
p_r^{nT}	Price in region r with no trade	q_r^{nT}	Production in region r with no trade
p_r^{cT}	Price in region r with congested trade	q_r^{cT}	Production in region r with congested trade
p_r^T	Price in region r with trade and no congestion	q_r^T	Production in region r with trade
C	Congestion		

Fig. 7.6 Graphical illustration of supply and demand curves as well as prices in the short-term equilibrium model with two regions (left diagram: supply and demand in region A, right diagram: supply and demand in region B, middle diagram: export surplus from B and import demand from A)

intersection point in the case of no trade (1). This export supply of region B ($\mathrm{ES_B}$) is illustrated in the middle diagram of Fig. 7.6. Accordingly, the import demand of region B ($\mathrm{ID_A}$) can be derived: the import demand of region A ($\mathrm{ID_A}$) is depicted in the middle of the diagram starting from the intersection point in the case of no trade (1). However, the supply curve in region A has to be vertically flipped. The intersection between export surplus ($\mathrm{ES_B}$) and import demand ($\mathrm{ID_A}$) in the middle diagram shows the market solution with uncongested trade resulting in the same market price in both regions ($p_\mathrm{A}^T = p_\mathrm{B}^T$). In this simple model, transmission losses are not considered. If transmission losses were considered, prices in the two regions would not be identical and the transmission losses could explain the price difference.

3. **Congested trade**: if the trade is congested due to a physical line constraint, the intersection point in the middle of the diagram (Fig. 7.6) is not feasible. The power flow between the two countries is shown on the x-axis of the diagram in the middle. A maximum transfer capacity can hence be depicted as a vertical line in the graph. This line intersects the import demand curve ($\mathrm{ID_A}$) and the export supply curve ($\mathrm{ES_B}$) and restricts the maximum import and export. As a consequence, different prices occur in the two countries, respectively, the two price zones.

Consequently, it can be summarised that prices will converge if there are no congestions (or, in other words, the power grid is strong enough to cover all exchanges between regions). If transfer capacities are not sufficient, congestions will occur, resulting in a split of prices between the two zones. Trade also results in a change of consumer and producer surplus in both countries. While producers in the exporting country benefit from exporting at higher prices, consumer loose in this country (due to higher prices) and vice versa in the importing country.

7.2.2 Formal Model

The mathematical model introduced in the previous section (see Eqs. (7.4)–(7.6)) can be easily extended to cover the transport of electricity between countries or regions. From a mathematical point of view, there is no difference between introducing a second or several other regions. Hence, the model below can be formulated for multiple regions, e.g. for the whole of Europe, by introducing a further index r for regions. It is necessary to differentiate between the generation and demand for the different regions and introduce transfer capacities between the regions.

The short-term **market equilibrium model with transmission constraints** can then be formulated as follows:

The objective function is again a cost minimisation[11]

$$\min_{y_{urt}} \sum_{u,r,t} y_{urt} \cdot \Delta t \cdot c_{ur}^{\mathrm{var}} \tag{7.8}$$

[11] In the case of an assumed inelastic demand curve.

subject to the capacity restrictions for generators u and an extended demand restriction by introducing the sum of all (net) power flows P from region r to all other connected regions $\bar{r}$:

$$0 \leq y_{urt} \leq K_{ur} \qquad \forall u, r, t \tag{7.9}$$

$$\sum_{u} y_{urt} \geq D_{rt} - \sum_{\bar{r}} P^{\mathrm{imp}}_{r\bar{r}t} + \sum_{\bar{r}} P^{\mathrm{exp}}_{r\bar{r}t} \qquad \forall t, r \tag{7.10}$$

Thereby, the flow P is either an import P^{imp} or export flow P^{exp}. The flow P between two regions is restricted by the available transport capacity Ψ. Thereby, the available transport capacity can be dependent on the direction (import or export), as described by the following equations:

$$P^{\mathrm{imp}}_{r\bar{r}t} \leq \Psi^{\mathrm{imp}}_{r\bar{r}} \qquad \forall r, \bar{r}, t$$

$$P^{\mathrm{exp}}_{r\bar{r}t} \leq \Psi^{\mathrm{exp}}_{r\bar{r}} \qquad \forall r, \bar{r}, t \tag{7.11}$$

This simple merit-order model taking transport between regions into account can be further extended by introducing transportation fees and transportation losses.

Since the liberalisation of European power markets, this model concept has been the basis for many European electricity market models. The concept allows to model European electricity markets, including the commercial trade between countries based on so-called **net transfer capacities** (**NTCs**, cf. Sect. 10.6.1). With upcoming flow-based market coupling in several regions in Europe, the concept can still serve as a basis, but has to incorporate further restrictions resulting from flow-based market coupling. The interested reader is referred to Schönheit et al. (2020), where a flow-based market coupling model is described in detail. Furthermore, a detailed introduction towards a fundamental understanding of flow-based market coupling is given in Schönheit et al. (2021).

7.3 Optimal Power Flow Model and Nodal Pricing

The idea of the economic **short-term market equilibrium** model may be easily linked to the **optimal power flow** model, well-known in electrical engineering. In contrast to the economic short-term market equilibrium models considered so far, which are throughout linear models, the optimal power flow model in its general form is nonlinear, as constraints from power flows are nonlinear (cf. Sect. 5.1.2). Optimal power flow models seek to optimise the operation of an electric power system subject to the physical restrictions imposed by electrical laws and engineering limits. Optimal power flow models have been studied in depth within the electric power systems community and constitute at the same time one of the practically most important and well-researched subfields of constrained nonlinear optimisation. They are also getting more and more attention in economics since **nodal pricing models** (cf. Sect. 10.8) are closely related to optimal power flow models and even coincide if power flows are modelled identically.

In the economic vein of building power system models, optimal power flow models combine an objective function (e.g. minimising system costs) with additional power flow equations to form an optimisation problem. The inclusion of the power flow equations is the feature that distinguishes optimal power flow models from other classes of power system problems, such as the above described, more standard economic models. Put differently, the pure economic dispatch derived from models without explicit power flow constraints may result in unacceptable flows or voltages in the network. This type of inconsistencies can be avoided by integrating power flow restrictions into power systems and electricity market models.

Optimal power flow models notably consider equality constraints for the power balance at each node, the so-called power flow equations (see Sect. 5.1.2). Furthermore, inequality constraints are included for network operating limits (line flows, voltages) and limits on control variables. Decision variables specifically include the active and reactive power output of the generating units, the power flows on the lines as well as the voltage and voltage angles at the buses. Depending on the purpose of the model, further technical details can be incorporated.[12]

Yet the size of the problem may get quite large with already thousands of lines and hundreds of control variables even in small applications, especially when time-coupling constraints such as those related to unit commitment or storage operation (see Sect. 4.4) are considered. Moreover, the AC power flow constraints are both nonlinear and non-convex (see Sect. 5.1.2.2), and the trigonometric functions sine and cosine complicate the construction of approximations. This is why many practical optimal power flow models are solved using a linear direct current (DC) power flow approximation, as discussed in Sect. 5.1.2.3.

Taking the formal model in Sect. 7.2 as the starting point, the power transmission equations from Sect. 5.1.2.3 have to be included (see Eq. (7.13)). Moreover, the energy balances are now formulated at the level of each node (Eq. (7.14)). Additionally, operating limits both for generators (Eq. (7.15)) and lines (Eq. (7.16)) are included:

$$\min_{y_{ut},\theta_{nt},P_{nmt}} \sum_{u,t} y_{ut} \cdot c_u^{\text{var}} \tag{7.12}$$

$$P_{nmt} = V^2 b_{nm}(\theta_{nt} - \theta_{mt}) \qquad \forall n, m, t \tag{7.13}$$

[12] The more technical details are considered, the more difficult is a representation as optimisation model with decision variables. For example, a more detailed representation of the grid may include modern, controllable devices, the position of transformer taps, the position of phase shifter (quad booster) taps, the status of switched capacitors and reactors, controls of power electronics (HVDC, FACTS) and curtailed load. Instead of a representation as optimisation model (optimal power flow), simulation approaches are very often used. This allows to represent more technical details but does not allow for an optimisation of decision variables.

$$\sum_{u \in U_n} y_{ut} - D_{nt} = -\sum\nolimits_{m \in \Lambda_n} P_{nmt} \qquad \forall n, t \tag{7.14}$$

$$y_{ut} \leq K_u \qquad \forall u, t \tag{7.15}$$

$$|P_{nmt}| \leq \Psi_{nm} \qquad \forall i, j, t \tag{7.16}$$

Thereby, the set U_n contains all generators connected to node n and set Λ_n encompasses all lines related to this node.

The optimal power flow model corresponds to the so-called nodal pricing model in the DC approximated form, which forms the basis of day-ahead market clearing in US markets (cf. Sect. 10.8).[13] In the simple (one-node) model in Sect. 7.1, the argument was made that the so-called dual variable of the demand constraint (also called the shadow price of the demand constraint) corresponds to the marginal cost of an additional unit of demand. This should be the price in a competitive market equilibrium. The same reasoning applies in the optimal power flow model: the dual variable corresponding to the nodal balance of node n (Eq. (7.14)) describes the objective function value increase resulting from an additional unit of demand in that node. This marginal cost corresponds to the price at that network node n in a competitive equilibrium. It can be shown that if none of the transmission constraints (7.16) is binding, the resulting price in all nodes will be the same. Yet as soon as one transmission constraint is binding, prices across the network deviate.

An alternative, equivalent representation of the network constraints can be obtained by using the so-called **power transmission distribution factors (PTDF)** discussed in Sect. 5.1.2.3. Thereby, the nodal energy balance (Eq. (7.14)) is rewritten, introducing the net nodal generation surplus Γ_{nt} as new variable:

$$\sum_{u \in U_n} y_{ut} - D_{nt} = \Gamma_{nt} \tag{7.17}$$

And the power flows are then computed by multiplying these net nodal positions with the PTDFs $A_{nmm'}$, which hence represent the marginal line loading on line mm' induced by an additional unit of generation in node n and consumed in some arbitrarily chosen reference node n_0:

$$P_{mm't} = \sum\nolimits_i A_{nmm'} \Gamma_{nt} \tag{7.18}$$

This approximation avoids the introduction of the nodal angles θ_{nt} as new variables, which are not easily interpreted. Also, the PTDFs have a relatively straightforward non-electrical engineering interpretation compared to the susceptances b_{jk}. On the

[13] For the real-time market, some US markets even integrate the AC power flow equations including N-1 security constraints (cf. Sect. 5.1.4.1) into their clearing algorithms to ensure feasible operation schedules.

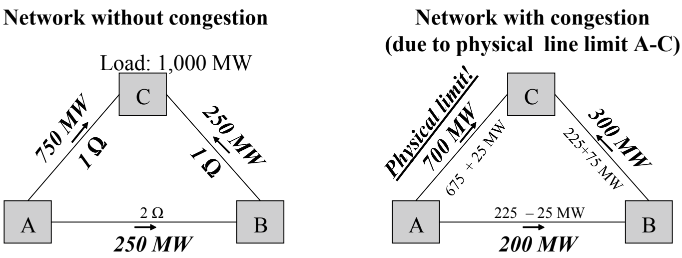
other hand, the resulting coefficient matrix for the restrictions has far more nonzero elements, which may slow down the computation time.

The difference between commercial and physical flows is demonstrated in the following example to illustrate optimal power flows. Furthermore, the application of PTDFs is illustrated using the same example. Figure 7.7 shows a simple power network with the three nodes A, B and C, each connected by a power line. Line impedances (cf. Sect. 5.1.2.1) are assumed to be 1 Ω, while the connection between A and B has an impedance of 2 Ω. Assume that a generator in node A has the lowest generation costs and thus supplies a load (a consumer) in node C with 1000 MW of electrical power (see the left side of the figure). This delivery contract can be concluded bilaterally and directly between A and B. The physical power flow does not correspond to the commercial flow as it depends on Kirchhoff's laws. Accordingly, part of the power flows via node C. In the simple example, the power flows are inversely proportional to the impedances, as depicted in Fig. 7.7. If a **nodal pricing system**[14] is assumed, prices at all nodes are identical (left side of Fig. 7.7), as no congestions occur in this example, and they correspond to the marginal costs of the generation capacity in node A. On the right side of Fig. 7.7, the line between A and C is assumed to be limited at 700 MW. Consequently, capacity A cannot fully provide the electricity for the load in C. Optimal power flows and nodal prices now result as depicted in the right side of Fig. 7.7. The nodal price in node C is even higher as the marginal costs of both generators in A and B. A further increase of the load in C would require a further reduction of the production in node A and a higher production in B (due to the physical line limit).

The same example is used to demonstrate the application of PTDFs. PTDFs express how the power flow is distributed in the network. The PTDFs translate a commercial flow (economically) into power flows (physically). In other words, PTDFs describe how a commercial transaction between two nodes affects all network elements. The PTDF can be obtained from the line impedances with the help of Kirchhoff's laws. Alternatively, a power flow simulation can be used, which in a first step determines the power flow on each line resulting, e.g. from the transaction from node A to node B. The corresponding PTDF values of the described example are shown in Table 7.1. In this simple example, PTDF values are easily obtained because power flows are inversely proportional to the impedances of the network.[15]

[14] Nodal pricing models are closely related to optimal power flow models.

[15] Note that for the application of the optimal power flow problem consisting of Eqs. (7.12) and (7.15)–(7.18) only the columns of the PTDF matrix in Table 7.1 have to be retained that share the same sink. E.g. if node C is selected as sink (referenc node), only the elements of the two columns A → C and B → C are used as coefficients $A_{n,mm'}$ in Eq. (7.18). The superposition principle valid for linear equations then implies that the impact of a commercial transaction between nodes A and B is represented by the combination of a bilateral transaction A → C with a transaction of opposite sign on B → C.



Assumption:
Load C: 1,000 MW,
Generation capacity A: 1,500 MW, B: 1,500 MW,
Costs A: 50 €/MWh, Costs B: 55 €/MWh

Result:

Optimal production

$y_A = 1{,}000\ MW;\ y_B = 0\ MW$

Nodal prices

$p_A = 50$ €/MWh

$p_B = 50$ €/MWh

$p_C = 50$ €/MWh

Optimal production

$y_A = 900\ MW;\ y_B = 100\ MW$

Nodal prices

$p_A = 50$ €/MWh

$p_B = 55$ €/MWh

$p_C = 57.5$ €MWh

Fig. 7.7 Simple power grid with three nodes and optimal power flows with/without physical line limit (left side: no restrictions on power lines assumed and impedances of lines are shown, right side: assumption of a physical limit on power line A to C)

Table 7.1 PTDF matrix of the example shown in Fig. 7.7, units dimensionless

		Commercial transactions		
		A → B	A → C	B → C
Lines	AB	1/2	1/4	−1/4
	AC	1/2	3/4	1/4
	BC	−1/2	1/4	3/4

Both versions of the DC approximation mentioned earlier are appropriate for most applications. However, inaccuracies may occur for heavily loaded transmission lines.[16] Furthermore, the DC approximation does not provide results for reactive power dispatch.

Generalisations of the optimal power flow model may be applied to decision making at various planning horizons, from short-term scheduling to long-term transmission network capacity planning. For electrical engineers, the optimal power flow model is a short-term model with fixed capacities, while capacities can also be adjustable as in a long-term market equilibrium, as described in the next section.

[16] There are also AC models that try to approximate nonlinearities. A real-world application of an AC model for Germany is for example presented in Hinz (2017).

7.4 Long-Term Market Equilibrium

The above-described short-term market equilibrium models treat power plant capacities as exogenously given. In the long run, power capacities might change, primarily due to the installation of new plants, extensions, decommissioning, etc. In contrast to the short-term market equilibrium, these power plant capacities are endogenously determined in the long-term market equilibrium. Hence, a model endogenous capacity adaptation is possible in **long-term market equilibrium**, while this is not the case in short-term market equilibrium. In terms of a system optimisation model, this means that capacities are not a right-hand side constraint, but a decision variable, which has to be introduced and handled in the model.

In the following, the short-term equilibrium model will be extended to a long-term equilibrium model by introducing the adaptation of capacities with the help of a decision variable (Sect. 7.4.1). A graphical interpretation and solution approach follow in Sect. 7.4.2 and an application in Sect. 7.4.3.

7.4.1 Formal Model

In long-term equilibrium, all input factors are variable, including power generation capacities. Consequently, the optimal amount of capacities in the system is determined by the model. The above-described short-term equilibrium model can be easily extended by introducing a new decision variable describing the amount of power plant capacities. Thereby, one may differentiate whether some pre-existing capacities are taken into account as exogenously given in the model or whether all capacities are decided upon endogenously. The latter one is often referred to as a greenfield approach as all capacities are built from scratch. If existing capacities are given and the model does not have to choose from scratch, the model is labelled as a brownfield approach. In the **theoretical economic concept of "long-term equilibrium"**, all production factors are variable. Hence no production capacities should be fixed in advance. In consequence, only greenfield approaches fulfil this condition from the outset, while for brownfield approaches, it depends on whether and when all capacity is adaptable, which might be the case for future modelled years.[17]

The long-term market equilibrium model with transmission constraints can be formulated by extending the former short-term equilibrium model by an additional variable for the newly installed capacity per technology K_{ur}^{new} multiplied with annual fix costs c_{ur}^{fix} (including annnualised investments, cf. Sect. 7.4.2) in the objective function:

$$\min_{y_{urt}, K_{ur}^{\text{new}}} \sum_{u,r,t} y_{urt} \cdot \Delta t \cdot c_{ur}^{\text{var}} + K_{ur}^{\text{new}} \cdot c_{ur}^{\text{fix}} \tag{7.19}$$

[17] Long-term therefore does not refer to a particular number of years ahead in the future but is rather defined through this theoretical assumption of all production factors being variable.

The capacity restriction has to take the expansion of capacity K_{ur}^{new} into account. Besides, already existing generation capacity may be considered in the model through the exogenously given parameters K_{ur} in a brownfield approach, whereas in a greenfield approach, all capacity has to be installed from scratch:

$$0 \leq y_{urt} \leq K_{ur} + K_{ur}^{\text{new}} \qquad \forall u, r, t \tag{7.20}$$

All other restrictions are identical to the short-term equilibrium model:

$$\sum_{u} y_{urt} \geq D_{rt} + \sum_{\bar{r}} P_{r\bar{r}} \qquad \forall t, r \tag{7.21}$$

$$P_{r\bar{r}} \leq \Psi_{r\bar{r}} \qquad \forall r, \bar{r} \tag{7.22}$$

In terms of solving the model, there is a significant difference compared to the previously presented short-term model: the same decision variable K_{ur}^{new} appears in the constraints for different time steps. Consequently, the solution of the optimisation problem has to be realised for the entire planning period in a joint optimisation run. In the previously defined short-term model, the solution for each time step may be computed separately (as done in the graphical solution approaches of Sects. 7.1.1 and 7.2.1).[18]

Analogously to the extension of power plant capacities, an extension of transmission capacities can be formulated by introducing a transmission capacity expansion variable in Eq. (7.22). Then, the model can calculate the optimal amount of transmission capacities and the interaction between optimal transmission and generation capacities.

Also in this model, prices are derived based on marginal costs. However, as investments in new capacities are taken into account, marginal costs may include, besides variable costs, also costs for installing new technologies. Thereby, the investment is described in our approach by the annualised expenditures of the investment. Marginal costs above variable costs only occur in hours when capacity is scarce and new generation must be installed to satisfy demand. There is usually only one hour of the year with maximum (residual) demand in a deterministic model. In this period, marginal costs will skyrocket since the marginal generation unit has to be installed to meet the demand in just this one hour. As the model is deterministic, meaning that no uncertainty is considered in the formulation, the price in this single hour includes the annualised investments. In all other hours, the marginal capacity is available (resulting in a surplus of capacities in these hours) and perfect competition leads to prices at the short-run marginal cost level. Consequently, the model yet provides long-run marginal costs instead of short-run

[18] Note that the introduction of storage or the inclusion of start-up costs or operation constraints like in the unit commitment and dispatch model of Sect. 4.4 also leads to time-coupling constraints, which require a simultaneous solution for all time steps. These intertemporal restrictions significantly increase calculation times and in general imply that the problem cannot be parallelised on several computers and has to be solved in a closed approach.

marginal costs with this “capacity markup” being attributed to one hour. This model is often also referred to as peak load pricing model.

Suppose the model is extended by taking uncertainty with regard to the peak load hour into account. In that case, the capacity markup will not occur for the one deterministic peak hour but spread among several probabilistic peak hours according to the assumed distribution. Compared with the real world, the model can, in principle, be interpreted in the sense that price markups can occur in times of scarce capacity.

An issue that has been raised repeatedly is whether these price markups may occur in real-world markets. Notably, administrative or technical price caps (as present also on the European power exchanges) may prevent prices from attaining the levels which would allow the recovery of investments. Also, political intervention might induce prices not to reach the occasional peaks needed for a long-term equilibrium. This problem is also known as the “missing-money problem”. If the price peaks are suppressed and prices only reflect short-term variable costs (as suggested by the models in Sect. 7.1), money to repay the investment is missing. And if this “missing money” is anticipated beforehand by the generation companies, they may decide not to invest at all, which may endanger the security of supply. Yet the very first model in Sect. 7.1 already points at one solution to that problem: if demand is (slightly) price-sensitive, prices may exceed short-run generation costs in hours with scarce supply, yet still, supply and demand are matched. With the resulting markups – possibly distributed over several hours – the generators may refinance their investments. Hence, proponents of purely competitive markets emphasise the necessity to increase **demand-side price-responsiveness**. Another possible way out of the missing-money problem is the introduction of so-called **capacity mechanisms**. These will be discussed in Sect. 10.5.

A further difference between the described modelling approach and the real world is that capacity adjustments in the model happen immediately so that the market is instantly reaching equilibrium capacities.[19] In reality, such an instantaneous capacity adjustment does not occur. It takes several years to install or phase out capacities due to the lead-time for planning, approval processes and implementation. However, the model illustrates the direction of adjustments of prices and capacities as a reaction to (unforeseen) demand changes[20]: in the short run, decreases (increases) in residual demand, e.g. by an additional feed-in of renewables (additional consumers like battery vehicles), will cause prices to decrease (increase) in a competitive market. In the longer run, decreases (increases) in demand in a competitive market will cause decreases (increases) in the installed capacity. In real market situations, markets with a decrease (increase) in demand

[19] The long-term equilibrium is only achieved, if all capacities are completely adjustable. This is the case for the greenfield approach; however, in the brownfield approach, not all capacities may be immediately adjustable. This is particularly relevant when the long-term market equilibrium requires the capacity to be phased out more quickly than capacity restrictions from the brown field approach allows.

[20] The issue of “unforeseen” or more precisely “non-anticipated” changes is at the heart of the discussion of “information efficiency” in Sect. 8.5. There it is related to the link between spot and future markets.

will have generators experiencing economic losses (profits). Over time, markets with generators experiencing economic losses (profits) will have generators exiting (entering) the market, and prices will increase (decrease) towards adequate levels. Hence, the long-term market equilibrium reflects the total electricity generation costs (for all load segments, cf. below). In contrast, the short-term market equilibrium is derived from short-run marginal costs of production.

7.4.2 Graphical Model

The (true) long-term equilibrium problem for a single region in a greenfield approach can be easily illustrated graphically (see Fig. 7.8). In the lower part of the figure, the so-called **screening curves** of technologies are depicted. These curves describe the "overall costs" of a specific electricity supply. These costs are not evident at first sight, given that generation capacity is measured in MW and electricity generation is measured in MWh. The graphical answer is quite simple as energy costs c^{var} are summed over the hours of operation to obtain the annual operation costs in EUR/MW/a as a function of the utilisation hours. Similarly, the investment expenditures are converted using an annuity factor (depending on the relevant interest rate and the lifetime) into an annual equivalent cost, known as the annuity of the plant. Additionally, the quasi-fixed cost for insurance, permanently employed staff, etc. may be included to obtain the annual per-unit fixed cost c^{fix}, also given in EUR/MW/a. Total costs are then a linear function of the (yearly) full-load hours fh of the power plant.[21]

$$c^{\text{tot}}(fh) = c^{\text{fix}} + c^{\text{var}} \cdot fh \tag{7.23}$$

Total costs, sometimes also referred to as annual revenue requirement per MW, are illustrated in the lower part of Fig. 7.8. Thereby, three generic technologies are depicted: one with low fixed costs and high variable costs, corresponding to a peaking technology such as, e.g. a gas turbine; a second technology with medium fixed and variable costs, i.e. a mid-merit technology such as a coal-fired or a gas combined cycle power plant; and finally a third technology with high fixed costs and low variable costs, such as a nuclear or lignite-fired power plant.[22] The cheapest technology can be derived for all possible annual operation hours graphically and mathematically by considering for all **full-load hours** fh the minimum of the three screening curves. Or put differently: the intersections of the screening curves delineate different choices for the cost-efficient technology. For a low utilisation (operation range t_1), technology 1 is the cheapest, followed by technology 2, which is the cheapest for the operation range t_2 and technology 3, respectively for t_3.

[21] Alternatively, the cost might also be expressed as a function of the capacity factor (more widely used in the USA), cf. Sects. 2.1.1 and 4.3.1.

[22] Total costs are considered from an investment and operation perspective. External costs, which are not or only partially included here, are discussed in Sect. 6.2.1.

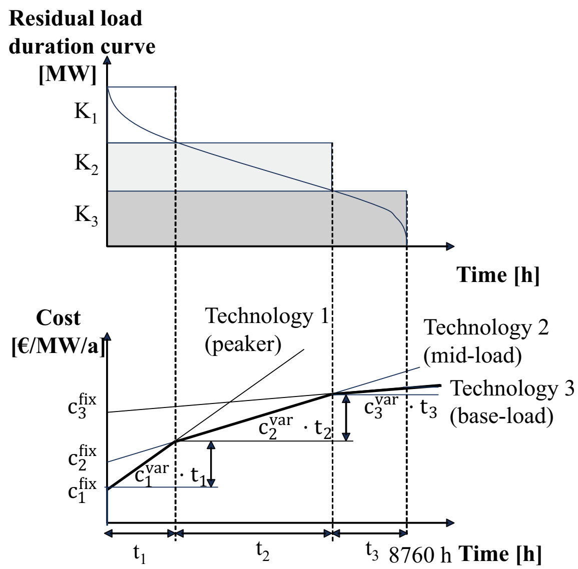


Fig. 7.8 Graphical illustration of optimal capacities in long-term equilibrium

But then the question arises of how the efficient portfolio of these three technologies can be determined. Therefore, the ordered residual load curve or the so-called residual load duration curve has to be introduced. The **residual load duration curve** is defined as the residual load (cf. Sect. 7.1.1) sorted by decreasing values. In the upper part of Fig. 7.8, the residual load duration curve is depicted, starting with the highest residual loads on the left side of the figure and the lowest residual loads on the right side. The optimal capacity of technology 3 is determined at the load level that exceeds an utilisation of $t_1 + t_2$ hours per year. It is guaranteed that this capacity is running $t_1 + t_2$ or more full-load hours per year, i.e. technology 3 is the cheapest option and thus the optimal choice. Technology 2 can be operated cheapest within the operation range between t_1 hours and $8760 - t_3$ hours. Analogously, the optimal capacity results from the diagram as K_2. Finally, the optimal capacity of technology 1 can be derived from the operation range below t_1 resulting in the optimal capacity K_1 for technology 1. For this range of full-load hours, it is guaranteed that technology 1 is the cheapest choice. Consequently, the optimal portfolio of technologies is obtained from this diagram.

A change of any cost parameter would alter the points of intersections and hence the optimal portfolio. Yet, a change in the residual load duration curve does not affect the intersection points. Therefore the cost-efficient operation ranges of the technologies do not change, although the capacities will adjust to the modified residual load. In this long-term equilibrium model, capacities fully adjust to the optimal portfolio, which is in reality not possible as adaptations of capacities take in general several years. Hence, the model must be understood as a theoretical long-term equilibrium or an equilibrium where capacities would adapt immediately.

Finally, the question arises of how marginal costs can be derived from this long-term equilibrium model. Like the short-term equilibrium model, prices are determined based on the shadow prices for satisfying the demand or, in other words, from the marginal costs. If capacities are not scarce in the model, marginal costs are only based on variable costs, in general fuel costs. If capacity is scarce and the capacity constraint is binding, then shadow prices will include the peaker technology's annualised investments besides variable costs. For this single hour, marginal costs would thus be extremely high. The **price duration curve** can generally be derived from the derivative of the lower envelope of the screening curves, i.e. the efficient cost curves. For the different load segments, this results in a piecewise function as variable costs are different for the three, respectively, four time segments. For period t_3 (cf. Fig. 7.8), marginal costs correspond to the variable costs of the baseload technology, for period t_2 of the mid-load technology and for period t_1 of the peak technology. The fourth time segment is hour 0 (at the coordinate point of the x-axis in Fig. 7.8). The derivative of the screening curve of technology 1 at hour 0 is not defined in continuous time, the jump in the cost curve corresponds to the marginal long-run cost of one additional MW of peak load capacity.

Important to note that this simplified example with three technologies could easily be extended to include many more technologies. Furthermore, this approach may also be applied to other sectors, such as, e.g. the heating sector.

7.4.3 Application

As an example, the scrutinised graphical long-term equilibrium model can be applied to answer the question what capacity is needed if further renewable power plants with an intermittent feed-in characteristic are integrated into the market. Thereby, it is also of interest, how the need for storage capacities is impacted by this development.

To answer these questions, **storage** power plants are first introduced to the graphical version of the long-term equilibrium model. Today, the state-of-the-art technology for storing large amounts of electricity are pump storage power plants. To integrate storage power plants, capital as well as variable costs have to be defined. Capital costs of pump storage power plants strongly depend on the regional conditions, resulting in a large range of capital costs. In our example, capital costs of the storage power plant are assumed to be lower than those of the mid-load technology but more expensive than that of the peak load technology. Variable costs of storage power plants depend on the operation of the pump, which necessitates electricity. In consequence, variable costs of pump storage plants mainly depend on electricity prices. In the scrutinised long-term equilibrium model, electricity prices are based on marginal costs of the three price-setting technologies, which are depicted in Fig. 7.9 in the lower diagram on the left side. The storage power plant is assumed to benefit from the low baseload prices (marginal costs of the baseload technology), resulting in lower marginal costs as the peak load technology. These low prices are available for t_3 hours of the year. However the relation between the capacity of the pump and the turbine as well as the round-trip efficiency of the pump storage power plant have to be taken into account. If pump and turbine capacities are identical, the pump storage plants can provide cheap energy using baseload electricity for $\eta^{\text{cyc}} \cdot t_3$ hours of the year, with η^{cyc} as round-trip efficiency of the plant. With differing pump and turbine capacities, available hours of the turbine at the lowest prices have to be adapted with the following calculation:

$$t_{\text{turb}} = \frac{K_{\text{pump}}}{K_{\text{turb}}} \cdot \eta^{\text{cyc}} \cdot t_3 \tag{7.24}$$

with t_{turb} available hours for turbine generation at lowest opportunity costs and K as installed capacity of the pump respectively the turbine. Correspondingly, marginal costs of the pump storage plant are based on the electricity prices used for pumping. However, these have to be corrected for the round-trip efficiency:

$$c_{\text{var}} = \frac{1}{\eta^{\text{cyc}}} p \tag{7.25}$$

with p the prices for pumping electricity consumption, respectively here the marginal costs of the baseload technology.[23] Only three levels of marginal costs can be observed (base, mid and peak load) in the model without storage. The marginal cost of the storage power plant depends on the prices used for charging and may add additional marginal cost levels in the duration curve.[24] If this example is expanded to include all installed power plants with their main characteristics, pumped-storage power plants also play a role in this model. They use low (or possibly negative) prices to provide their capacity during peak load times.

With the increase of renewable generation capacities, what happens in the long-run market equilibrium with conventional capacities is of interest. The effects of higher shares of renewable capacity on the optimal conventional capacities can be easily understood with the help of a graphical illustration of the long-run optimal capacity equilibrium model. First, it is necessary to know how the residual load duration curve changes with higher shares of renewable capacity – the residual load curve results from the difference between the load and the renewable feed-in. As marginal generation costs of intermittent renewables are generally zero, it is implicitly assumed that the produced electricity reduces the load. The residual load duration curve is just the sorted residual load curve; hence, the highest residual loads are on the left side. In contrast, the lowest residual loads are sorted to the right side of the diagram (see upper left graph in Fig. 7.9).

How is the residual load duration curve changing with low to high shares of renewable energies? On the left side of the diagram, which corresponds to times of peak demand, nearly no change of the residual load duration curve can be observed

[23] More detailed discussions of the role and properties of storage in the long-term market equilibrium may be found in Steffen and Weber (2013).
[24] This increasing level of complexity of price patterns is discussed in Böcker and Weber (2020).

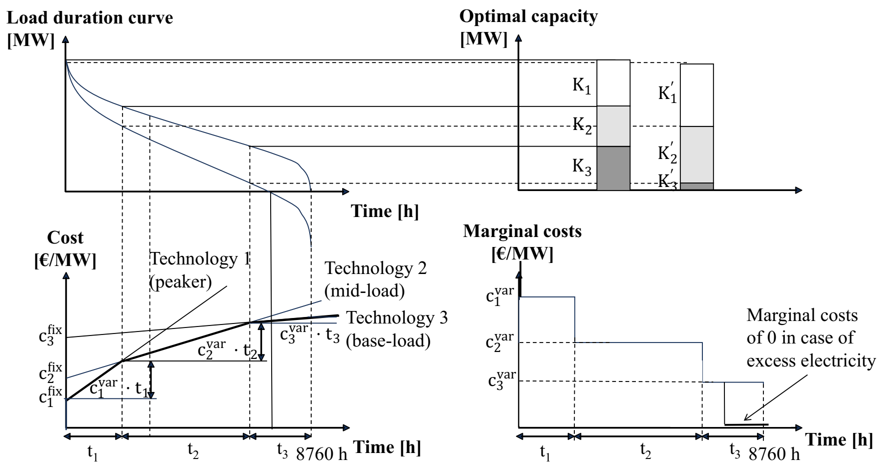


Fig. 7.9 Graphical illustration of optimal capacities in long-term equilibrium. Lower left diagram: screening curves of three technologies and shape of efficient cost curve, upper left diagram: two residual load curves, upper right diagram: installed capacities in long-term equilibrium for the two different residual load curves, lower right diagram: marginal costs in long-term equilibrium, almost identical for both residual load curves

with higher renewable shares. The reason is quite simple: high loads (in Europe) often occur on winter working days at noon or in the evening; however, renewables may not produce electricity at all at these times, e.g. as no wind is available and the sun has set or PV generation plants are covered with snow or not producing due to foggy weather. This situation is often called a "**dark calm**". On the other hand, it is also likely that relatively low loads can coincide with high renewable generation resulting even in (very) negative residual loads. Situations with high renewable feed-in are sometimes also referred to as a "**bright storm**". Consequently, the residual load duration curves are quite similar on the left side of the diagram, while a sharp decrease with even negative loads can be observed on the right side of the graph. As shown in the section before, the optimal capacities in market equilibrium can be derived with the help of the curve of cost-effective technologies and the points of intersection. It is obvious that the amount of baseload technologies is significantly lower at higher shares of renewable energies, while the amount of peak load (and storage) technologies is higher. To summarise, the long-run optimal market equilibrium with higher shares of renewables results in higher peak and lower baseload capacities compared to a situation with lower percentages of renewables.

In addition to how much capacity is needed at the market equilibrium, it is also of interest, what the impact of this change on electricity prices is. As depicted in Fig. 7.9, prices derived from marginal costs of the technologies remain unchanged with and without renewable technologies as long as no renewable surplus occurs, meaning that total renewable feed-in remains smaller than demand. This result may be surprising at first sight but can easily be explained. The long-run optimal equilibrium model assumes that capacity adapts immediately and optimally to the new market situation. Hence, overcapacities, as well as scarce capacities, are not considered. The shape of the efficient cost curve and therefore marginal costs and intersection points remain the same. In consequence, it can be concluded that the merit-order effect of renewables does not occur in the long run (as opposed to the short-term consideration). It has to be kept in mind that in reality, the capacity adjustment is not realised immediately, resulting (among others) in the observed short-term merit-order effect of renewables. Planning horizons for installing new capacity take several years, but also, phasing-out of capacity is a decision process, which is not implemented immediately.

Two further aspects have to be mentioned concerning this long-term market equilibrium:

1. **A renewable surplus decreases prices**: if a renewable surplus occurs (renewable feed-in larger than demand), the above-mentioned statement that no merit-order effect occurs is no longer valid. The baseload technology sets the price as long as the residual demand is larger than zero in the baseload segment. However, if a renewable surplus occurs, renewables will most likely be curtailed and prices are zero as depicted in Fig. 7.9 (or even negative depending on the support scheme for renewables). Consequently, longer periods with renewable surplus result in a higher number of hours with zero prices. In that sense, there is a merit-order effect in long-term market equilibrium when renewable capacities are exogenously given.
2. **Long-term equilibrium with endogenous determined renewable capacities**: in the scrutinised model, renewable capacities are assumed to be exogenously determined (e.g. as a consequence of renewable support schemes), while conventional capacities are endogenously (and immediately) adapting to the new situation in the model. In a more stringent definition of long-term market equilibrium, all capacities, including renewables, should be determined endogenously in the model. A specified CO_2 cap is then decisive for the expansion of renewables. In such an equilibrium, the feed-in profiles of renewables can be expected to have higher importance to avoid self-cannibalisation (see Eising et al. 2020).

7.5 Further Reading

Stoft, S. (2002). Power System Economics – Designing Markets for Electricity. New York: Wiley.

This book provides an introduction to power system economics and introduces relevant key design elements of modern electricity wholesale markets and puts them in their economic context.

Weber, C. (2005). Uncertainties in the Power Industry – Methods and Models for Decision Support. New York: Springer.

The book aims at an integrative view of power companies' decision problems in liberalised markets. It systematically investigates the uncertainties power companies are facing and develops mathematical models to describe them – with a focus on combining fundamental and finance models.

Schönheit, D., Kenis, M., Lorenz, L., Möst, D., & Delarue, E. B. (2021). Toward a fundamental understanding of flow-based market coupling. Advances of Applied Energy, Volume 2, art. 100027.

This article easily explains the fundamentals of flow-based market coupling. It provides an exemplary open-access model based on a test network and the data and code for the model. The functioning and effects of the most influential parameters are demonstrated by providing a guide to the theory and conducting several case studies.

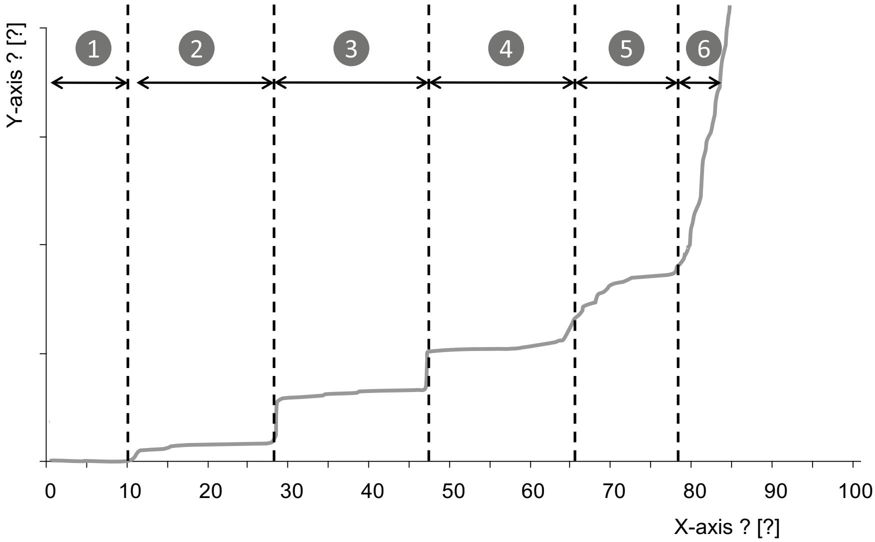
7.6 Self-check of Knowledge and Exercises

Self-check of Knowledge

1. Explain the specificities of electricity markets.
2. Illustrate a stylised supply curve and intersection with demand curves at different consumption levels.
3. Explain the term market equilibrium in the context of the stylised diagram of question 2.
4. Discuss the assumptions underlying the concept of perfect competition concerning electricity markets.
5. What is the difference between a welfare maximisation approach and a cost minimisation approach? In which context can a cost minimisation approach be applied and what is the advantage of using this approach?
6. What is the merit-order effect of renewables?
7. How can bottom-up models be benchmarked, and what are typically used indicators?
8. What is the difference between a single zone and a two (or multiple) zone model?
9. What does optimal power flow mean?
10. What are PTDFs and for what purpose are they needed?
11. What is the difference between long-term and short-term market equilibrium?
12. Explain the shape of the efficient cost curve with the help of screening curves of three different technologies (base, mid and peak technologies).
13. Illustrate the optimal capacities in a long-term equilibrium for all three technologies for two different residual load curves (today's situation and with an extremely high share of intermittent renewables).

Exercise 7.1: Merit-Order Curve
The diagram below shows a simplified representation of the German merit-order curve before the introduction of emissions trading.

1. Label the axes (incl. units) and assign the energy sources (hard coal, lignite, nuclear, hydro + wind, storage + oil, gas) to the six areas.
2. Define the term peak load power plant. Which of the power plants in part (1) would be labelled peak load plants? Discuss their profitability in the context of an increasing carbon emissions allowance price.



Exercise 7.2: Investments in Power Markets
The annual load curve of a country is characterised by the following functions:

$$P_1(t) = 100\ \text{GW} - 0.02 * t\frac{\text{GW}}{\text{h}}\ [0\,\text{h} < t \leq 2000\,\text{h}]$$

$$P_2(t) = 60\,\text{GW}\ [2000\,\text{h} < t]$$

There are three types of power plants available:

	Lifetime (years)	Investment (€/kW_{el})	Fuel costs (€/MWh_{th})	Efficiency (η, %)
Lignite	40	1500	4	42
Coal	40	1100	8	46
Natural gas	30	700	21	59

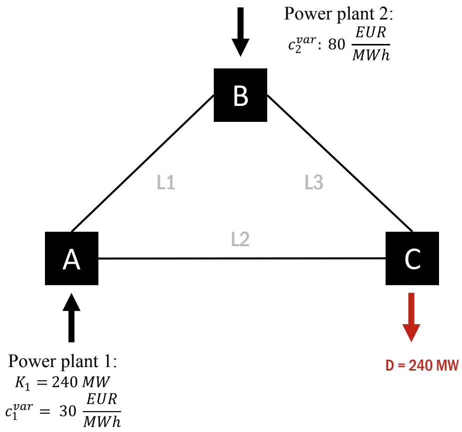
The interest rate r applied for all types of power plants is *10%*.

1. Sketch the annual load duration curve of this country in a suitable diagram. Be sure to label the axes, including the units, correctly.
2. Determine the cost-optimal utilisation (full-load hours) for each power plant technology (*mathematical solution*) and illustrate your result qualitatively using the so-called screening curves. (Note: Perform calculations using annual annuities of the investments.)
3. A *maximum of 30 GW* of lignite power plants can be installed due to the availability of the resource. Determine the cost-effective power plant fleet. How much capacity (in GW) needs to be installed for each power plant technology?
4. Now assume that renewable energy sources partly cover the stated demand. First, define the term residual load. Qualitatively characterise the shape of the residual load curve with a moderate feed-in of renewable energies in your sketch in part (a) of this problem. What are the implications of the new curve for the optimal capacity level? Qualitatively present your arguments. Be sure to discuss the effects of the highest and lowest system loads.
5. Formulate the problem as an optimisation model. How does introducing a maximum capacity of 30 GW of lignite and the change of the residual load curve change the formulated problem?

Exercise 7.3: Nodal Prices

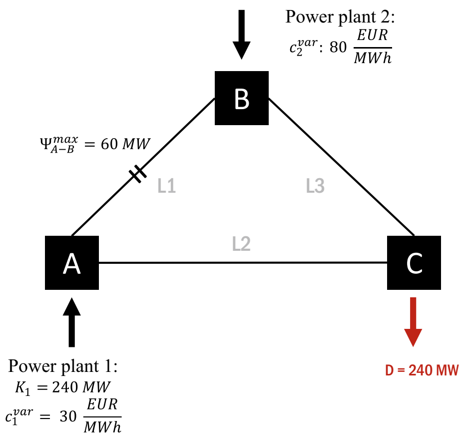
1. In the context of electricity market designs discussed in the chapter, describe the basic features of a nodal pricing scheme. Name *two advantages and disadvantages* of a nodal pricing scheme.
2. A three-node network is illustrated in the diagram below. All lines have identical impedances and are unlimited in terms of transport capacities. Power plant 1 at node A is subject to a generation restriction $G_1^{\max}$. The generation of power plant 2 at node B is unrestricted. The power plants have variable costs G_1^{var} and G_2^{var}. Demand L at node C is price inelastic.

   (a) Which generation levels result in the cost-optimal provision of electricity to meet the demand?
   (b) Which power flows occur?
   (c) Which nodal prices are obtained at nodes A, B, C?



3. In contrast to the example in the previous task, the capacity of line L1 between nodes A and B is now limited to 60 MW ($P_{A-B}^{\max}$).

   (a) Which generation levels result in the cost-optimal provision of electricity to meet the demand?
   (b) Which power flows occur?
   (c) Which nodal prices are obtained at nodes A, B, C?

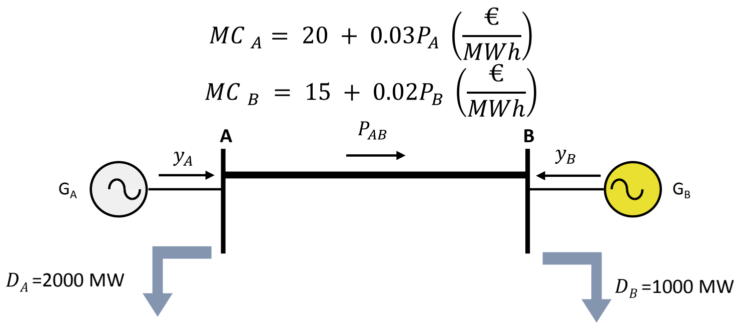


Exercise 7.4: Congestion Rents

Consider a very simplified two-node power system as below. The marginal cost of production of the generators connected to buses A and B are given, respectively, by the following expressions:

$$\mathrm{MC}_A = 20 + 0.03P_A\left(\frac{€}{\mathrm{MWh}}\right)$$

$$\mathrm{MC}_B = 15 + 0.02P_B\left(\frac{€}{\mathrm{MWh}}\right)$$



Assume the demand is constant and price inelastic, that electricity 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 node, the production of each generator and the flow on the line for the following cases:

(i) The line between the nodes is in service and has an unlimited capacity.
(ii) The line between the nodes is in service, but its capacity is limited to 600 MW. The output of the generators is unlimited.

(a) Calculate the congestion rent for case ii.
(b) Assume a zonal pricing scheme. How must the dispatch be adjusted to correct the network congestion in case ii. Calculate the corresponding cost increase (redispatch cost).

References

Böcker, B., & Weber, C. (2020). A primer about storage in bottom-up models of future energy systems. *Fundamentals of storage operation and investment in competitive long-term equilibria* (HEMF Working Paper No. 02/2020). University of Duisburg- Essen. House of Energy Markets & Finance (HEMF). Available at: http://hdl.handle.net/10419/234067 [Accessed May 22, 2022].

BMWI. (2017). *Renewable energy sources act—Translation of the German EEG 2017*. Federal Government of Economic Affairs and Energy. Available at: https://www.bmwk.de/Redaktion/EN/Downloads/renewable-energy-sources-act-2017.pdf%3F__blob=publicationFile&v= [Accessed May 22, 2022].

Eising, M., Hobbie, H., & Möst, D. (2020). Future wind and solar power market values in Germany—Evidence of spatial and technological dependencies? *Energy Economics, 86, art. 104638*.

Hinz, F. (2017). *Voltage stability and reactive power provision in a decentralizing energy system*. Series of the Chair of Energy Economics and Dissertation: Technische Universität Dresden (Vol. 12). Available at: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-229585 [Accessed May 22, 2022].

Martínez-Díaz, D. (2008). Production cost models with regard to liberalised electricity markets. Univ.-Verl. Karlsruhe and Dissertation: Universität Karlsruhe (TH). Available at: https://publikationen.bibliothek.kit.edu/1000009066 [Accessed August 18, 2022].

Schönheit, D., Hladik, D., Hobbie, H., & Möst, D. (2020). *ELMOD documentation: Modeling of flow-based market coupling and congestion management*. EconStor Preprints. ZBW - Leibniz Information Centre for Economics. Available at: https://EconPapers.repec.org/RePEc:zbw:esprep:217278 [Accessed May 22, 2022].

Schönheit, D., Kenis, M., Lorenz, L., Möst, D., Delarue, E., & Bruninx, K. (2021). Toward a fundamental understanding of flow-based market coupling for cross-border electricity trading. *Advances in Applied Energy, 2*, art. 100027.

Steffen, B., & Weber, C. (2013). Efficient storage capacity in power systems with thermal and renewable generation. *Energy Economics, 36*, 556–567. Available at: https://doi.org/10.1016/j.eneco.2012.11.007.

Stoft, S. (2002). *Power system economics—Designing markets for electricity*. Wiley.

Weber, C. (2005). *Uncertainties in the power industry—Methods and models for decision support*. Springer.

8 Markets: Organisation, Trading and Efficiency

Historically, electricity systems were developed by single companies (either privately, municipally or state owned) operating an integrated system of generation and networks. Competition was introduced in electricity systems only at the end of the 1980s. Against this background, this chapter aims at answering the following key questions:

- How can electricity markets be organised?
- Which forms of trading and auctions do exist?
- How are schedules used for the coordination between trading and grid operation?
- Which instruments can be used for the reduction of price risks?

Electricity markets are an essential element of a deregulated electricity sector. Section 8.1 thus discusses the basic organisational structures of the electricity sector, whereas Sect. 8.2 is devoted to the basics of electricity trading. In Sect. 8.3, key market design choices are discussed, whereas the coordination between trading and grid operation through balancing groups is discussed in Sect. 8.4. Section 8.5 then addresses the link between markets with different delivery horizons, namely spot and futures markets. Section 8.6 explores the role and functioning of futures markets, considering also the extension to options.

Key Learning Objectives

After having gone through this chapter, you will be able to

- Describe the electricity market structure after the introduction of competition.
- Differentiate between electricity spot and derivatives markets and describe the link between these two markets.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_8

- Define the key market design elements.
- Understand how electricity trading and the physical electricity flows are connected.

8.1 Organisation of the Electricity Sector

As discussed in Chap. 6, the electricity sector has been organised through regional or national monopolies for most of its existence. Whereas some legislations established nationwide **integrated utilities** such as EDF in France or the former Central Electricity Generation Board (CEGB) in England & Wales, others with more federal traditions like Germany, Switzerland or Norway, had more decentralised structures with one or several large-scale integrated utilities and tens or hundreds of smaller, usually **municipally-owned utilities**. The components of the conventional electricity system discussed in Chaps. 4 and 5 were then allocated among the stakeholders as shown in Fig. 8.1. Large integrated utilities were responsible for large-scale generation – along with (hydro) storage where relevant – and the transmission grid. By contrast, the distribution grid was frequently managed by regional and municipal utilities, although also large-scale utilities covered part of the electricity distribution, notably in rural areas. Moreover, several municipal utilities also had stakes in generation, mainly in CHP units providing district heating. Additionally, these utilities sometimes had (and still have) stakes in the gas and water distribution.

The deregulation of the electricity sector implies that competitive and monopolistic parts of the electricity value chain have to be separated – the so-called **unbundling** (see Sect. 6.1.2). Moreover, new entities may emerge, notably trading houses and energy exchanges. In the case of full unbundling, the resulting interrelations may be schematically represented as in Fig. 8.2.

Markets thereby emerge at two stages: on the one hand, generators, traders and retailers (also called suppliers) trade among each other. This is the so-called wholesale market where the produced good (electricity) is traded between parties without being consumed. On the other hand, the retail market covers trades involving the final customers of the good electricity and others – notably suppliers.

8.2 Basics of Electricity Trading

Trade describes the transfer of goods or services from one person or entity to another, in general in exchange for money. Also, electricity as a commodity can be traded, even if it has some unique characteristics as the non-storability and the

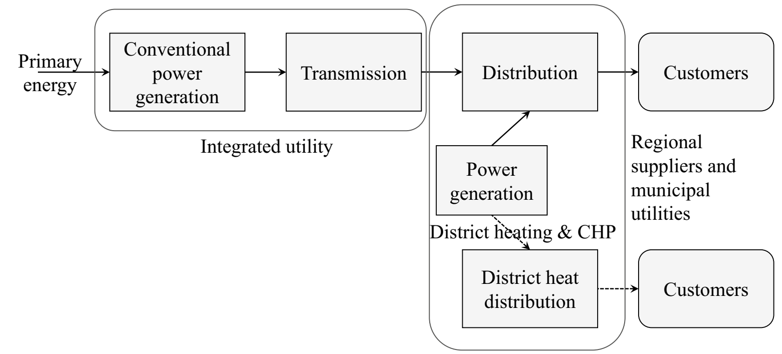


Fig. 8.1 Traditional market structure before liberalisation

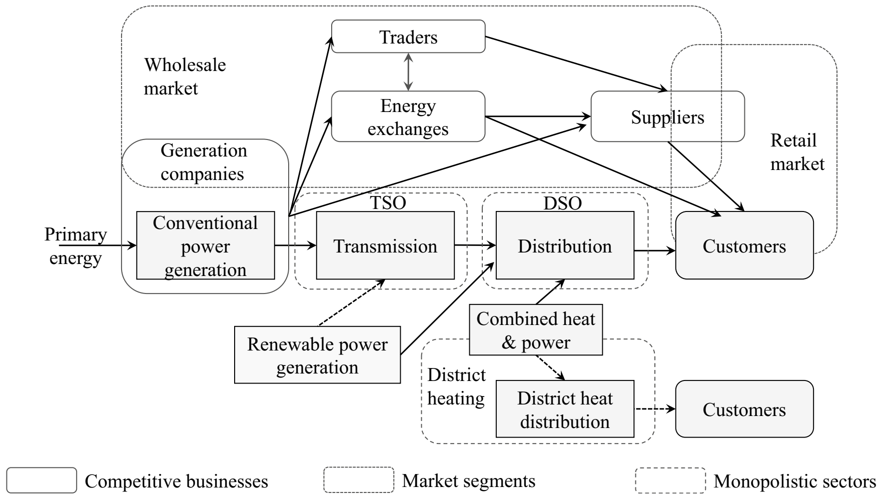


Fig. 8.2 Market structure after the introduction of competition and full unbundling

necessity to balance production and consumption in real-time. In the following, some basic concepts to describe (electricity) trading are introduced. Details of the organisation of markets in Europe are discussed in Chap. 10.

A basic distinction for trades is the number of participants: trade between two traders is called bilateral trade, while trade between more than two traders is called multilateral trade and is often organised as mediated trade (cf. e.g. Stoft 2002, p. 86).

- In a **bilateral trade**, buyers and sellers trade directly. These markets need little design and are less organised. The advantage of such bilateral trade is the flexibility as the involved parties can specify any contract terms they desire. However, bilateral trade has often the disadvantage of high-transaction costs, e.g. for writing and negotiating contracts, even if standardised contracts can be used. In general, bilateral trade is only utilised for exchanges of larger quantities so that flexibility can be exploited. At the same time, the disadvantage of high-transaction costs plays a minor role. A typical example might be a full electricity delivery service provided by a larger utility to a "Stadtwerk" (municipal utility).

- **Mediated trade** is more centralised and standardised than bilateral trade. In general, mediated trade can be organised by brokers, platforms and finally by energy exchanges. They provide marketplaces where standardised products can be traded. Despite the standardisation, a transaction is only realised if offers by sellers and bids by buyers are matched. Besides trading of standardised products, exchanges provide additional services, such as, e.g. market clearing. The clearing is necessary because the speed of trades is faster than the execution time for validating the underlying transaction. It ensures that trades are settled following the market rules, even if a buyer or seller becomes insolvent before settlement. With the liberalisation of the European energy market, several energy exchanges have been founded in Europe, such as, e.g. Nordpool (Scandinavia), APX Power NL (The Netherlands), Powernext (France), APX Power UK (Great Britain), OMEL/OMIE (Spain) and European Energy Exchange (Germany). As exchanges continuously adapt their products to market needs, several new products and market platforms (e.g. intraday-trading) have emerged, but also mergers and consolidations of exchanges (e.g. EPEX SPOT) have occurred since liberalisation.

Additionally, trading may be organised either on a voluntary or on a mandatory basis.

- In most European countries, participation in energy exchanges is voluntary. Consequently, buyers and sellers decide what exchanges and products they want to choose and whether they participate in the future, day ahead, intraday or reserve energy markets.
- In mandatory or compulsory markets, often organised as **compulsory pools**, all participants are required to sell their output to the pool at the pool's price. The utilities agree that the dispatch is controlled by a dispatch office or a pool administrator in power pools. All the tasks regarding the exchange of power and the settlement of disputes are assigned to the pool administrator. Power pools (may) provide potential advantages resulting from synergies, such as saving in reserve capacity requirements, more reliable operation and decreased operating costs. However, power pools have also some shortfalls, namely that costs associated with establishing a central dispatch office may be quite high, the pool agreement may be very complex, and pool members may have to give up their rights to engage in independent transactions outside the pool (see Sect. 10.8).

Trading requires an agreement about the product characteristics: the most important ones are the time and place of delivery. As electricity is not directly storable, a fine granularity is required for planned physical deliveries. Therefore, the spot markets usually trade products for delivery periods of one hour or even less (e.g. 15 min). A certain grid location is specified as delivery place, e.g. the entire area of a transmission grid operator or a specific grid node.

The term **spot market** thereby designates markets for immediate delivery of the traded product. This definition is not specific to the electricity or energy markets but rather applies to commodity and financial markets in general. In the case of electricity markets, immediate delivery usually means that trades occur one day ahead of delivery (**day-ahead markets**) or on the same day (**intraday markets**, in the US **real-time markets**). Details on spot markets in Europe are discussed in Sect. 10.1.

Besides spot markets, **derivative markets** exist. As the name indicates, these are derived markets, which refer to another market or object. In financial markets, a broad range of derivative markets exists. The assets traded there are then simply labelled **derivatives**, and the reference object to which a product refers is labelled the **underlying**. E.g. many derivatives refer to stocks traded on exchanges like the New York Stock Exchange (NYSE) or the London Stock Exchange (LSE).

In electricity markets, the underlying of derivatives is generally the electricity traded at the electricity spot markets. The most essential derivative markets are then the **futures markets**, which allow trading for more distant delivery periods, e.g. months, quarters or years to come. If the trades occur on a registered power exchange like EEX or Nordpool, the products are named **futures**. If the products are traded bilaterally or on other trading platforms, they are labelled **forwards**. Typically, forwards include the possibility of physical delivery of the product, whereas futures are settled purely financially. Other derivative products include so-called **options**. Whereas forwards and futures describe contracts for a firm delivery of a product, options give a right to the holders without putting an obligation on them. This may be the right to purchase the underlying at a later stage at a price agreed today (call option) or the right to sell the underlying (put option). Derivatives are mainly used to guard their owners against volatile prices of short-term markets, in other words, for hedging reasons. More about the role of futures and options and some key characteristics will be presented in Sect. 8.6.[1]

Furthermore, specific markets and clearing mechanisms are needed to ensure the balance of electricity supply and demand in real-time, supporting grid stability. Since market mechanisms are not fast enough, the responsibility for the operation of the electricity system in the very short-term remains in the hands of grid operators. The markets in Europe operate until the so-called gate closure (usually less than one hour before delivery) and afterwards, the system operation responsibility is put into the hands of the TSOs. In order to fulfil their task, they first need information from

[1] More details on options may be found in Hull (2018), yet with a more general perspective on financial markets. Options on electricity have so far not been traded very actively (see Sect. 10.2), yet the concept is important to describe flexibilities (see Chap. 11). Also other derivatives discussed in Hull (2018), such as swaps, are sometimes traded on energy and specifically electricity markets. But they are also of minor importance compared to forwards and futures.

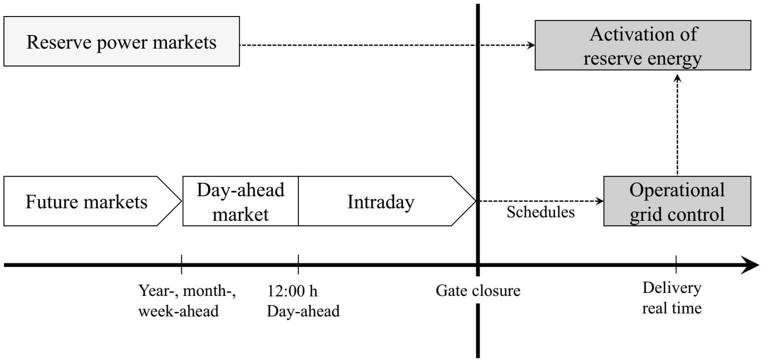


Fig. 8.3 Sequence of market and grid control operations for a specific delivery time segment

the market participants on their planned operations, and the trades they have concluded. Here, the concept of balancing groups plays a key role (see Sect. 8.4). Second, they need means to handle unexpected changes in the system. As unbundling implies that grid operators do not own generation assets, they must procure flexible capacities as so-called reserves. In Europe, this is usually done on specific reserve power markets (see Sect. 10.3). Third, these reserves have to be used to maintain grid stability – this is the reserve activation. Finally, the costs related to the reserve use have to be attributed to the responsible parties – here again, the balancing groups play an essential role. The reserve activation and corresponding cost attribution are also summarised under the term of **balancing mechanism.** The sometimes employed term "balancing market" is instead a misnomer, as there is no real matching of demand and supply on a marketplace at this stage.

The sequence of the different market segments in Europe is also summarised in Fig. 8.3. The key design choices for the market segments are further discussed in the following subsection.

8.3 Key Market Design Choices

For the market segments mentioned in the previous section, several market design[2] choices have to be made. According to Ockenfels (2018), "**market design** is the art of designing institutions in such a way that the behavioural incentives for individual market participants are in line with the overarching goals of the market architect". Designing electricity markets is different from designing markets for other commodities due to the peculiarities of the good electricity, like securing a permanent equilibrium between supply and demand without having the possibility to store electricity by itself and the necessity of an electric network. Furthermore, as the technical and economic characteristics of electricity systems change, the electricity

[2] A deeper discussion of market design can be found in Roth (2002).

market design has also to be seen dynamically; the market design might have to be modified to adapt to the changes of the system.

Subsequently, the focus is on crucial market design elements that may apply to the categories mentioned above of spot, derivative and reserve power markets. The details of the actual implementation in Europe are discussed in Chap. 10. Because of the ongoing transformation of the electricity system towards a low-carbon system, the interplay of these market design elements with carbon certificate markets and renewable support mechanisms (see Sect. 6.2.4) has also to be considered carefully. Challenges in that field are discussed in Chap. 12.

A first fundamental choice is between continuous trading and auction-based market clearing. **Continuous trading** is the standard approach in financial markets and allows market participants to adjust their positions at any time during trading hours – in this setting, new information may lead immediately to changes in positions and prices. The **information efficiency** in such an approach is hence high (see also Sect. 8.5). Continuous trading is usually based on an "open order book". The **open order book** collects so-called **limit orders**, i.e. quantity-price pairs and stacks buy orders and sell orders separately. Buy and sell orders are only matched if quantities and prices fit together. E.g. the buy orders are sorted in descending order concerning the limit **(bid) prices** of the participants. In the open order book, sellers can now see whether they are willing to sell at the highest bid price or not. If they are willing, they may directly submit a so-called market order, unconditional on price that will be matched with the available buy orders for execution. Alternatively, they may also place a limit order (with an **ask price**), which will only be matched if the ask price does not exceed the highest bid price.

By contrast, **auction-based trading** in the electricity markets collects bids until one point in time and performs a market clearing after that. These auctions are typically held as sealed-bid auctions.[3] Trading results reflect the information available until that point in time and later updates cannot be considered. On the other hand, the collection of bids increases the liquidity in the market. Complex matching and settlement mechanisms may be implemented in auction-based trading, e.g. to consider grid capacity constraints. This tends to improve **allocative efficiency**, notably when scarce grid resources are to be used.

In contrast to many other auctions (e.g. for fine artworks), power market auctions are **multi-unit auctions** since multiple units of the same product are contracted.[4] Within auction-based markets, a further key distinction is between two-sided and single-sided auctions. In **single-sided auctions,** all market participants submit sell orders.[5] Only one single buyer (or a group of buyers who act collectively) procures a good or service through this auction. Such single-sided auctions are typically held to procure reserve power in the electricity markets (see Sect. 10.3). **Two-sided auctions**

[3] Unsealed bid auctions using, e.g. an ascending or descending clock approach (cf. e.g. Krishna 2010) are rarely found in power markets.

[4] For a general introduction to auctions with focus on single-object auctions, we refer the interested reader to Krishna (2010).

[5] Alternatively, single-sided auctions may also be run on the basis of a collection of purchase orders. Yet this case is not relevant for the power markets and therefore not dealt with subsequently.

by contrast allow the submission of both purchase and sales orders. This is a typical setting for spot markets, notably the day-ahead markets. There, electricity suppliers will submit purchase bids and generators sales bids, whereas pure traders may position themselves on either side of the market. The market clearing will then determine the market price that allows the execution of the maximum trading volume.

A specific issue that arises in multi-unit single-sided auctions is the selection of the pricing approach. **Uniform pricing** implies that all selected bids receive the same price – typically the price of the last accepted bid in the case of procurement auctions. Uniform pricing – also known as a **clearing price auction** or **pay-as-cleared** – is the standard approach for two-sided auctions since it corresponds to the economic textbook approach of determining the market clearing price at the intersection of supply and demand curves. For single-sided auctions, **discriminatory pricing** seems at first sight more attractive from the viewpoint of the single buyer. The buyer only pays the bidders the price they have bid – therefore, such auctions are also known as **pay-as-bid auctions** – and thus saves compared to a remuneration based on the marginal price. Yet under this auction scheme, bidders have a clear incentive to align their prices with the expected marginal price. This "guess-the-price" bidding behaviour leads to inefficiencies as market participants may align their behaviour with their peers instead of revealing true scarcity.

Another peculiarity observed, notably in reserve power markets, is that of **multi-part bids**. Thereby, bidders submit not only one price per bid but a bid with multiple prices – one price for reserve capacity and one for the corresponding energy. A related concept is so-called **complex bids**, which are frequently used in U.S. electricity markets (see Sect. 10.8). Thereby, detailed characteristics of a power plant, such as minimum stable operation limits or reserve provision capabilities, are transmitted as part of the bid. At the other extreme, bids in continuous trading usually only include a bid price and a bidding quantity. This allows quick and easy matching and thus helps to establish markets with high liquidity. On the other hand, multi-part or complex bids generally require complicated matching algorithms and thus are hardly implemented in continuous trading.

8.4 Balancing Groups: Coordination Between Electricity Trading and Grid Operation

As the transport of electricity is grid-bound and thus depends on the infrastructure, trade cannot neglect the physics of electricity transport. Consequently, a link between trading and physical delivery and hence with grid operation is necessary. Physical delivery of electricity requires a permanent balance of generation and consumption (taking also grid losses into account). Permanent refers to the time scale, hence this balance has to be guaranteed at any time. However, on day-ahead markets, trading is usually organised on an hourly basis; consequently, 24 single hours a day are differentiated. With regard to physical delivery, these single hours are average values of the physical delivery. On some intraday markets, trading is already possible for quarter hour products, so that a further differentiation for the four quarter hours of an hour is possible. Finally, in real-time, balancing is necessary in continuous time, which is organised with the help of reserves, discussed in Sect. 10.3.

But how is this balance organised in the electricity system? For this purpose, so-called **balancing groups** are installed. A balancing group is a virtual energy account for any market participant in the wholesale electricity (and also gas) market. With the help of a balancing group, the virtual world of electricity (and gas) trading and the physical world of energy flows and grid stability are brought together. The size of a balancing group can be very different, e.g. a city can be covered by different balancing groups. Balancing groups are not only established for utilities, but also for larger industrial facilities, which purchase their electricity on their own. Suppliers and generators are obliged to assign the consumers they supply and their feed-in points (e.g. their own power plants) to a balancing group. The balancing group managers (also called balancing responsible parties) have to guarantee that their power balance is balanced in every quarter hour. Therefore, the balancing group managers have to provide a forecast of their balancing group and deliver the forecast to the grid operator. This forecast is called a "schedule" and has to be provided for each quarter hour of the following day. These schedules have to be submitted to the system operators, who perform so-called **day-ahead congestion forecasts** and – in the case of congestions – will take counteractive measures (see Sect. 10.6).

Deviations from this schedule might result from power plant failures or inaccurate forecasts for load and renewable feed-in, which can lead to a shortfall of power or a surplus of power in a balancing group. As the control area of the transmission system operator (TSO) typically consists of a multitude of balancing groups, the positive and negative deviations of the different balancing groups might offset each other at least partially. The remaining deviation of the whole control area has to be compensated by the TSO using control reserves (so-called active balancing),[6] which the system operator procured on markets (see Sect. 10.3). In case of a deviation in their balancing group, the responsible balancing group managers will have to pay or be compensated by the so-called **imbalance price** for their deviations from their schedules. This imbalance price is calculated for every quarter hour (settlement period of the imbalance price).

There are different ways how this imbalance price (IP) might be determined. In principle, the imbalance price should represent the costs the TSO had or the compensation[7] the TSO received when procuring the control reserve energy[8]

[6] Some countries also permit passive balancing. In that case, TSOs send a timely price signal to balancing groups which are then allowed to be intentionally unbalanced to compensate the current imbalance (see Hirth et al. 2015).

[7] A compensation is possible, e.g. because the provision of negative control reserve might lead to reduced fuel costs for power operators.

[8] For some forms of control reserves, the providers are also paid for the reserve capacity provided. This capacity is normally procured for a much longer time period than only a single settlement period (e.g. of 15 min). On this account, these costs are typically not attributed to the balancing groups deviating from their schedules but to all users of the power grid via the use-of-system charges.

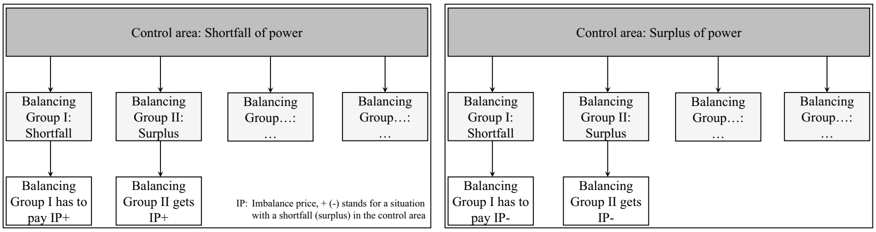


Fig. 8.4 Settlement of balancing energy in a one-price system by the TSO

needed to compensate for the deviation in his system. Typically, the imbalance price is rather low in a situation with a surplus of power in the whole control area (IP–), as in such a situation negative control reserve is needed (right-hand side in Fig. 8.4), and rather high in a situation with a shortfall of power in the whole control area (IP+), as in such a situation positive control reserve is needed (left-hand side in Fig. 8.4). Furthermore, the imbalance price can be used to set incentives to avoid a deviation from the schedule provided. In a **one-price system**, the imbalance price to be paid by some balancing groups and the imbalance price with which other balancing groups are compensated is the same. In contrast, the imbalance prices differ in a **two-price system** (cf. e.g. Vandezande 2011, pp. 37–46). The charge or compensation of the balancing group manager depends on whether the whole control area of the TSO has a shortfall or surplus of power and whether the deviation in the balancing group is in the same or opposite direction (see Fig. 8.4 for the example of a one-price system).

Overall, the balancing group management is responsible for the following activities:

- The provision of the load forecast of consumers, the operational schedule of power plants and storages, etc. These activities are daily business and are carried out for the following day (day ahead) and for the same day (intraday). The schedule is submitted to the system operator, who performs day-ahead congestion forecasts.
- Determination of the actual (real-time) consumption and delivery.
- Billing of the required balancing energy (used to balance actual consumption and feed-in).
- Responsibility of contracts between the balancing group manager, transmission system operators, distribution system operators as well as the generators and consumers (not single small customers, but aggregated via energy retail companies).

8.5 Information Efficiency: Links Between Spot and Future Markets

The basic models of electricity markets sketched in Sect. 7.1 describe equilibria in a market with physical production and consumption of electricity. Actual markets for electricity are much more differentiated than these models. Besides the design aspects discussed in the previous subsections, the duality of spot and future markets has to be examined in detail.

As noted before, **spot markets** are commonly defined as markets for immediate delivery of the traded product. The spot market price is then frequently denoted $S(t)$ for delivery at time t.

On the contrary, **futures markets** are used for trading products at a certain point in time t for delivery at a future point in time T or during a future period. Hence, future market prices are always written using at least two indices, e.g. $F(t, T)$ for trading at time t and delivery at time T. As the distinction between **forwards markets** and futures markets is not essential for the following considerations,[9] we follow the common practice to use the term future markets in a broader sense, also encompassing the forward markets.

Before taking a closer look at the link between spot and future markets in Sect. 8.5.2, a fundamental general relationship between different marketplaces has to be highlighted in Sect. 8.5.1, the so-called law of one price. Furthermore, for the price formation on futures markets, the "efficient market hypothesis" is an important theoretical reference point. Therefore, it is discussed in Sect. 8.5.3 together with the implications for the price link between future and spot markets. This is then deepened for storable commodities in Sect. 8.5.4, and the case of limited storability is discussed finally in Sect. 8.5.5.

8.5.1 Law of One Price

The law of one price, sometimes also called Jevons' law, stipulates: "In the same open market, at any moment, there cannot be two prices for the same kind of article" (cf. Jevons 1871, p. 92).

The implications of that law can be seen in the currency markets. Differences in the Euro-Dollar exchange rate in the fourth decimal place between, e.g. Frankfurt and New York immediately induce massive computer-based arbitrage trading activities so that the differences in prices are almost immediately reduced to zero.

What Jevon labels "open markets" includes notably two salient features: low barriers to market entry and low-transaction costs. Trading in major energy and specifically electricity markets is typically subject to bid-ask spreads below 1% and even lower platform transaction costs. The costs for market entry are not negligible. They include, e.g. personnel costs for traders, costs for market access, training and

[9] For a deeper discussion of the relevance of the distinction we refer to Hull (2018).

certification, computer system and licencing costs. However, electricity exchanges for most West-European countries now have more than one hundred participants. Hence, deterrence of market entry is not a major issue.

We can therefore emphasise two implications of the law of one price for electricity trading: first, the existence of different trading platforms in one market will not lead to multiple, divergent prices for the same product (with the same place of delivery and same delivery period) at the same time. The transaction costs set the upper bound to the simultaneous price difference. Second, the law of one price does not preclude price changes between two different trading times. Yet the possibility of storing a good, non-physical arbitrage trades and efficient information processing will shape the relation between future and spot markets (see Sect. 8.5.3).

8.5.2 Link Between Spot and Futures Markets

In an efficient market design, the link between spot and futures markets is well defined and asymmetric:

The **spot market** as the last market before delivery reflects the actual supply and demand situation. Therefore, the spot market price reflects the effective scarcity situation at delivery. The spot market is the fundamental market for the physical matching of demand and supply and the spot price is the fundamental reference price.

As the **futures markets** are derivative markets (see Sect. 8.2), the prices there refer to the corresponding spot price(s) for the delivery time or period. An obvious question then is: if the spot market is the "true" physical market, why are futures markets needed—or more precisely: what are the economic benefits of having a futures market on top of a spot market? The key answer is that future markets enable market participants to hedge part of their risk in the physical market. Without future markets, market participants would be obliged to sell or buy energy at a potentially volatile spot market rate. This may lead to important losses on the balance sheet of producers or consumers and consequently, they may run into a financial distress situation or even bankruptcy. Physical players may limit or even eliminate their revenue or cost risk with transactions on the futures markets. This was also historically a major motivation when futures markets started to develop for agricultural products in Chicago in the 19th century.

Trading on futures markets has not necessarily to be done based on physical products. Instead, a financially settled futures market has clear advantages for all market participants: for market participants with physical positions, the main benefit of reducing financial risk is achieved as well by a financial futures market as by a physical forwards market. And for financial (i.e. non-physical) players, market entry is much easier if any position taken can be settled purely financially. Also the transaction costs tend to be lower since no physical delivery needs to be organised. Hence, a financial market tends to attract more participants and thus better serves the hedging needs of participants with physical positions. There is only one crucial caveat: the link to the underlying spot market must be well defined, and the spot market must be sufficiently liquid to allow settling of physical positions.

Futures markets usually operate on a **mark-to-market principle**, i.e. all open positions are settled daily against the current settlement price. This minimises the financial exposition of the trading platform and the clearing house (cf. Hull 2018, pp. 29–32). Nevertheless, it has also implications for the financial reporting of companies. Notably smaller, municipally-owned utilities are fearing balance sheet volatility when they use financial products for hedging purposes—as these have to be valued on the accounts using volatile market prices. In contrast, the physical portfolio counterpart (generation assets or retail contracts) is valued at standard book values. Together with increased regulations for financial products introduced in the aftermath of the global financial crisis of 2008, this may raise non-monetary barriers for entry into these markets both for small and large players (cf. ECA 2015).

8.5.3 Efficient Market Hypothesis and Link Between Spot and Future Prices

The efficient market hypothesis is a cornerstone for linking future prices to spot prices. It states for financial assets in general that current prices are reflective of all currently available information (cf. Fama 1970, 1991; Malkiel 1973). Since futures are a financial asset class, this claim may also apply to futures—and even forwards.[10] In a risk-neutral world, the efficient market hypothesis implies:

$$F(t,T) = \mathbb{E}[S(T)|\Omega_t] \tag{8.1}$$

i.e. the futures price at time t for delivery in time T corresponds to the expected spot price for time T given the information available at time t, which is summarised in the set Ω_t. This is true since entering a future contract (at least in theory) does not involve any initial payment, rather the contract is cash-settled at expiry in T. If market participants are indifferent to risk-taking, their willingness to pay today for an uncertain cash flow $S(T)$ in the future is exactly equal to the expected value of that cash flow. If that property does not hold, there would be possibilities for arbitrage. So the efficient market hypothesis may be seen as a generalisation of the law of one price (see Sect. 8.5.1) to trades at different moments in time. Obviously, if new information arises between t and T, the actual spot price at delivery may be different from the previously quoted future price. But any information already known at time t should be reflected in the then future price.[11]

Given that real-world agents are mostly risk-averse, the previous relationship will hold in reality only in the modified version

[10] The latter mostly foresee a physical settlement. But since there is a continuous secondary market for trading, the corresponding positions may be closed financially, and they may be used as financial asset.

[11] Note that different types of market efficiency may be distinguished following Fama (1970) according to the content of the information set Ω_t.

$$F(t,T) = \mathbb{E}[S(T)|\Omega_t] + \lambda(t,T) \tag{8.2}$$

Thereby λ is used to denote the risk premium paid for avoiding spot market risk. λ may be positive or negative, depending on whether the risk aversion of buyers or sellers prevails in the market. This risk premium is not directly observable, and different papers have come to different conclusions regarding the existence and height of that risk premium (cf. e.g. Bessembinder and Lemmon 2002; Benth et al. 2008).

If the changes in the information set are bounded in a certain probabilistic sense, we moreover have

$$\operatorname*{plim}_{t \to T}(F(t,T) - S(t)) = 0 \tag{8.3}$$

i.e. the future price converges to the spot price in probability as delivery approaches.

8.5.4 Implications of Storability

The previously established relationships between spot and future market prices are essential for understanding and analysing electricity market prices. Yet, it is also important to apprehend what they do **not** include: a link between the current spot price and current future price notations.[12] For storable commodities like crude oil or pure financial assets like stocks, such a relationship is established by the theory of storage. Although electricity is not storable, indirect storage possibilities like hydro reservoirs or battery storage may have similar effects.

If we think of a homogenous product with available storage capacities and storage costs $C^{\text{sto}}(T-t)$, then the so-called cash-and-carry arbitrage prevents the following situations:

$$F(t,T) - S(t) > C^{\text{sto}}(T-t) \tag{8.4}$$

$$S(t) - F(t,T) > -C^{\text{sto}}(T-t) \tag{8.5}$$

In the first case, buying at the current spot price and simultaneously selling at the future price would enable an arbitrage gain despite the physical storage costs $C^{\text{sto}}(T-t)$. Conversely, the second situation would allow selling physically now and replenishing later at costs given by the current future price. These considerations may be extended by considering the so-called **convenience yield**, i.e. the benefits of disposing of the commodity physically today. A positive convenience yield counterbalances storage costs and may lead to a negative effective cost term

[12] Or in the absence of future market quotes: a link between the current spot price and expected future spot prices adjusted for risk premia.

$C^{sto}(T-t)$ in the above inequality. This may then explain spot prices exceeding future prices ("**backwardation**").[13]

Together with Eq. (8.3), the previous inequalities imply—if storage costs tend to zero for small time intervals $T-t$,

$$\lim_{t \to T}(S(t) - S(T)) = 0 \tag{8.6}$$

Hence, storability leads to smooth price changes in the spot market.

8.5.5 Implications of Limited Storability

The previously established relationships do, in general, not hold for electricity prices. Notably, there is no reason why spot market prices for adjacent intervals in time should be close to each other. If demand (or inflexible supply, e.g. from renewables) changes from one interval to the next, the market prices may change abruptly. This is obvious in Fig. 8.5 for the case of the German power price. In Norway, by contrast, the available storage capacities enable arbitrage between subsequent hours, and the prices are much less volatile. A notable exception yet occurs during the first five days of the month – apparently, Norway imports the price volatility from the continent. This may be a consequence of higher demand in Norway, which is met by imported electricity or by peaking hydro units with high reservation prices (see Sect. 4.4.1.2).

So the key driver for short-term electricity price volatility is insufficient storage capacity. If available storage can enable a full arbitrage between hours of different demand, spot prices will be very stable, otherwise, they may fluctuate strongly. If residual demand is uncertain ex-ante, the relationship between current future prices and actual spot prices will also turn out to be less stable.

8.6 Future and Option Payoffs and Hedging of Physical Positions

As discussed in Sect. 8.5.2, futures and other derivatives have been primarily designed to enable owners of physical assets, such as power plants or retail customer contracts, to reduce their exposure to volatile spot prices – to "**hedge**" their price risk. To understand how this may be achieved, it is useful to consider first the payoffs at the expiry date that come along with futures and options. A bit of trader "slang" is useful for that purpose: a **long position** means that a trader has bought a contract. In the case of a future, this means that he is entitled to get the underlying commodity (or other tradeable assets) at the agreed expiry date T of the future at the price $F(t, T)$ agreed at trading time t. Or rather, since futures are usually settled financially, he will receive

[13] The opposite case with future prices exceeding spot prices is called "**contango**".

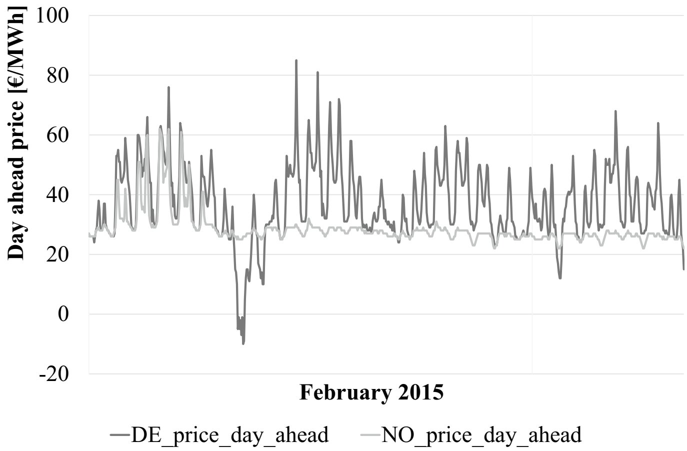


Fig. 8.5 Electricity prices in systems with little storage (Germany) and large storage (Norway). *Source* Own illustration based on data from www.epexspot.com and www.nordpool.com

financial compensation so that he may buy the underlying on the spot market and have total cost equal to $F(t, T)$. This is obtained through the **payoff function** $S(T) - F(t, T)$. Figure 8.6 illustrates this payoff function for the long position in a future as a solid line. For a **short position**, where the trader has sold the contract, the sign of the payoff scheme is reversed, graphically it is flipped horizontally (see dotted line in Fig. 8.6) – at least in an idealised world where transaction cost, bid-ask spreads and other market imperfections are disregarded.

Hence, selling a future contract is a convenient hedging strategy for a power producer with a fixed and predictable output. Without the hedge, revenues of the producer would be proportional to the spot price times the sold quantity: $R_0 = q \cdot S(T)$. Adding a short position in q futures leads to the total revenue term:

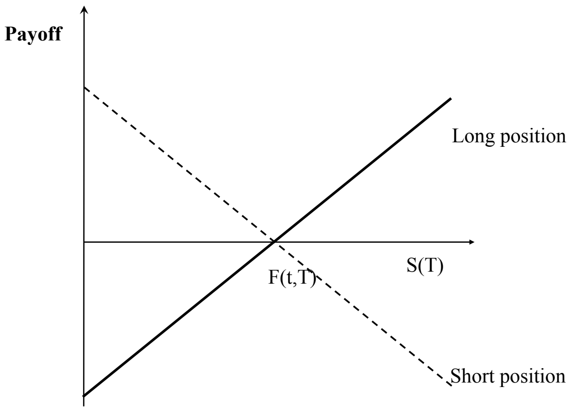


Fig. 8.6 Payoff functions for long and short positions of a future

$R_H = -q \cdot (S(T) - F(t,T)) + q \cdot S(T)$. The first term thereby represents the compensation obtained from the financial settlement of the future, and the second term is the spot procurement cost. After rearranging terms, we get $R_H = q \cdot F(t,T)$. Hence, the revenues no longer depend on the volatile spot prices.

The spot market price risk for a price-independent production quantity q is thus best hedged by entering a corresponding short position on the futures market. Or put in trader terms: an open long position resulting from the physical asset is closed by a corresponding short position on the future market. Obviously, this can also be done in the opposite direction: a retailer with physical delivery contracts for quantity q may hedge the price risk of this physical short position by entering a long position on quantity q on the futures market.

For many generation assets, the production comes at some variable cost c^{var} (see Sects. 4.3.2, 4.4 and 8.1). Consequently, the producer is better off if he does not sell at prices below c. The payoff obtained then on the spot market is described by the solid line in the shape of a hockey stick shown in Fig. 8.7. At prices below what is further on called the **strike price** X, the generation unit does not produce. This implies also that the payoff is zero. Beyond the strike price X, the operation margin (revenue minus variable cost) is $S(T) - X$ for each unit of production. This exactly corresponds to the payoff function of a call option with strike price X.

A **call option** provides the buyer (also called "holder") the right to purchase the underlying at a predefined strike price X before or at an expiry date T from the seller (also called "writer"). If the spot price $S(T)$ is below the strike price, the buyer will preferably not exercise the option but instead buy the underlying directly at the lower spot market price $S(T)$. With a financial settlement, the payoff function of the call option may thus be written $\max\{0, S(T) - X\}$, see Fig. 8.7. This financial derivative hence provides the buyer with protection against price increases in the underlying beyond the strike price X.

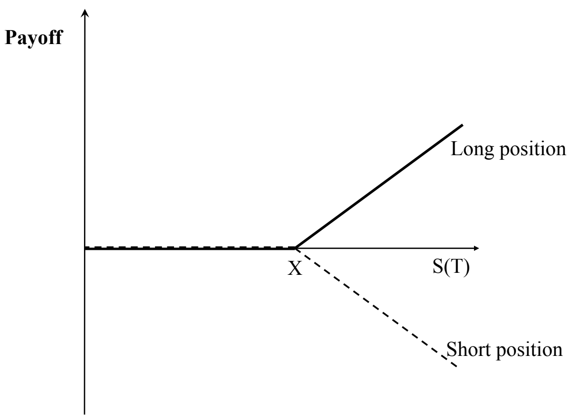


Fig. 8.7 Payoff functions for long and short positions of a call option

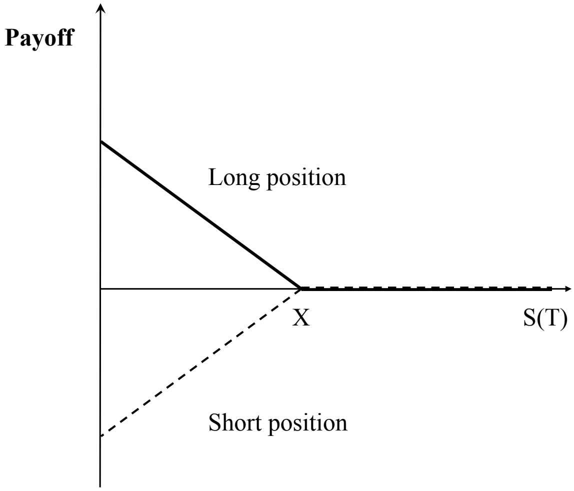


Fig. 8.8 Payoff functions for long and short positions of a put option

Instead of protection against upward price risks, protection against downward price risks may be required occasionally. This is provided by a **put option**. It provides the holder with the right to sell the underlying at a predefined strike price X before or at an expiry date T to the writer. The corresponding payoff functions are depicted in Fig. 8.8. For the holder of the option, it is given by the expression $\max\{0, X - S(T)\}$, whereas the writer of a put option has a payoff at the expiry of $-\max\{0, X - S(T)\}$.

With the payoff for the writer of a put option being always negative or zero (and similarly for a call option), this is obviously not a good deal for the option writer. This potential financial loss is compensated by an **option premium** paid by the buyer upfront at the signature of the option contract. There is an analogy to an insurance contract: the writer of a call option takes the risk of price increases from the option holder and therefore receives a premium. And conversely, the writer of a put option takes the risk of price drops and is similarly rewarded by a premium. This underlines that taking a short position in any option – i.e. being the seller or writer of the option – is usually not a risk-reducing but a risk-taking strategy. Moreover, it raises the question of how a fair option premium or option price may be computed on which seller and buyer could agree. Chapter 11 discusses this question in detail.

For a power plant operator, selling a call option may yet be a risk-diminishing strategy. The power plant itself is the physical equivalent of a long position in a call option with the strike price being equal to the variable costs of the plant – if all technical constraints like minimum operation times (see Sect. 4.4) are disregarded. If this long position is complemented by a short position on a financial call option with a similar strike price, the net risk may be reduced. Only if the strike price of the financial option is precisely equal to the variable costs of the power plant, the risk may be entirely eliminated. Given the diversity of real-world power plants, this implies that options with a broad range of strike prices should be traded. Yet this reduces the liquidity of trade on every single option. Together with other real-world complications, this prevents widespread option trading in the electricity markets so far. One of the other relevant issues is the different granularity of power plant operation (typically planned in time intervals of hours) and power derivatives (mostly traded at yearly or at least monthly granularity). The bridging of this gap will also be discussed in Chap. 11.

8.7 Further Reading

Hull, J. (2021). Options, Futures and Other Derivatives. 11th edition. Harlow et al.: Pearson.

The book Options, Futures and Other Derivatives give a detailed overview about different forms of derivatives and derivatives markets.

Stoft, S. (2002). Power System Economics – Designing Markets for Electricity. New York: Wiley.

The book Power System Economics provides a comprehensive introduction to the different aspects of the design of power markets.

Wilson, R. (2002). Architecture of Power Markets. Econometrica, 70, 1299–1340.

This paper discusses the design of spot and forward markets for electricity and different methods to mitigate market power.

8.8 Self-check of Knowledge and Exercises

Self-check of Knowledge

1. What are the main differences between the electricity market structure before and after the introduction of competition?
2. Which European energy exchanges do you know?
3. Plot the sequence of futures, spot and reserve markets and grid control operations.
4. Explain the differences between continuous trading and auction-based market clearing.
5. Where in the power markets do we typically find single-sided auctions and where two-sided auctions?
6. Compare the typical bidding behaviour in a clearing price auction and a pay-as-bid auction.

7. Why are balancing groups needed in the power sector?
8. Formulate the law of one price.
9. Formulate the efficient market hypothesis.
10. Explain what is meant by cash-and-carry arbitrage.

Exercise 8.1: Payoff Functions of Derivatives, Technologies and Portfolios
For a specific hour h, an energy company has bought an electricity future (long) with a price of 35 €/MWh and sold a call option (short) with a strike price of 35 €/MWh and an option premium of 1 €/MWh. Furthermore, the company owns a combined cycle gas turbine (CCGT) with techno-economic data according to Chap. 4. Draw the payoff functions of the future, the option, the CCGT (for 1 MWh each) and the entire portfolio (1 MWh of each component) for a spot price range between 30 and 150 €/MWh for hour h.

Exercise 8.2: Control Reserve and Imbalance Pricing
A TSO needs 1000 MW positive control reserve. The following 6 providers participated in the tendering scheme. Which providers will the TSO select in a multi-part auction if the selection of the winning bids is realised using the capacity rates (€/MW), whereas the activation is based on the energy rates (€/MWh)? Then, during the quarter 9.00–9.15, the TSO needs 100 MWh of positive reserve energy. Which suppliers will the TSO select to provide this positive reserve energy? Calculate the imbalance price (IP+) for these 15 min in a one-price system, assuming that average pricing is used. The control area of this TSO consists only of two balancing groups. Calculate the corresponding cash-flows for balancing energy if balancing group 1 had a shortfall of power of 100 MWh from 9.00 to 9.15, whereas balancing group 2 was well-balanced. How do the cash-flows for balancing energy look like if balancing group 1 had a shortfall of power of 150 MWh from 9.00 to 9.15, whereas balancing group 2 had a surplus of power of 50 MWh during this time?

Provider	Capacity (MW)	Capacity rate (€/MW)	Energy rate (€/MWh)
1	100	200	20
2	80	300	40
3	160	100	140
4	500	150	80
5	160	250	60
6	160	400	60

References

Benth, F. E., Cartea, A., & Kiesel, R. (2008). Pricing forward contracts in power markets by the certainty equivalence principle: Explaining the sign of the market risk premium. *Journal of Banking & Finance, 32*, 2006–2021.

Bessembinder, H., & Lemmon, M. L. (2002). Equilibrium pricing and optimal hedging in electricity forward markets. *The Journal of Finance, 57*, 1347–1382.

ECA. (2015). *European electricity forwards and hedging products—State of play and elements for monitoring*. Final Report to Acer. Economic Consulting Associates. Available at: https://extranet.acer.europa.eu/en/Electricity/Market%20monitoring/Documents_Public/ECA%20Report%20on%20European%20Electricity%20Forward%20Markets.pdf [Accessed May 28, 2022].

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*, 383–423.

Fama, E. F. (1991). Efficient capital markets II. *The Journal of Finance, 46*, 1575–1617.

Hirth, L., & Ziegenhagen, I. (2015). Balancing power and variable renewables: Three links. *Renewable and Sustainable Energy Reviews, 50*, 1035–1051. https://doi.org/10.1016/j.rser.2015.04.180

Hull, J. (2018). *Options, futures, and other derivatives* (10th ed.). Pearson.

Jevons, W. (1871). *The theory of political economy*. Macmillan.

Krishna. (2010). *Auction theory* (2nd ed.). Academic/Elsevier.

Malkiel, B. (1973). *A random walk down Wall Street*. Norton.

Ockenfels, A. (2018). *Marktdesign*. Available at: https://wirtschaftslexikon.gabler.de/definition/marktdesign-51491/version-274653 [Accessed December 21, 2018].

Roth, A. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica, 70*, 1341–1378.

Stoft, S. (2002). *Power system economics—Designing markets for electricity*. Wiley.

Vandezande, L. (2011). *Design and integration of balancing markets in Europe*. Dissertation: KU Leuven.

Wilson, R. (2002). Architecture of power markets. *Econometrica, 70*, 1299–1340.

9 Imperfect Competition and Market Power

In the market models developed in Chap. 7, producers and other agents (consumers, storage operators, etc.) are assumed to be price takers so that the assumptions of fully competitive markets are applicable (cf. Sect. 7.1.2). However, evidence in many European countries suggests that conventional generation is dominated by a very small number of prominent players, e.g. EDF in France. This raises concerns that large players may be willing and capable of strategically distorting market results away from competitive outcomes in favour of their profits. Such exercise of market power leads to imperfect competition and to both practically and scientifically challenging questions:

- How can (exercise of) market power be detected?
- How can it be incorporated into market models?
- What actions can and should be taken by the government and other agents?

These questions are not specific to electricity markets but arise in all kinds of markets, from traditional industries like steel and car making to the platforms of the Internet economy like Google and Facebook. Still, parts of the answers are dependent on the specifics of the markets considered. Therefore, we subsequently aim at combining general insights from economics with approaches specific to the power industry. We start by looking at possibilities to measure market power in Sect. 9.1 and then recall the standard game-theoretical models of imperfect competition in Sect. 9.2. Section 9.3 points at some applications of game-theoretical models to the wholesale electricity markets, while Sect. 9.4 extends the perspective beyond the area of wholesale markets to retail competition.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_9

Key Learning Objectives
After having gone through this chapter, you will be able to

- Describe the concepts used to identify, measure and mitigate market power.
- Apply key indicators for market power and simple models of imperfect competition to the wholesale electricity market.
- Discuss extensions of market models with strategic behaviour and their relevance in electricity markets.
- Analyse retail market strategies and equilibria under imperfect competition.

9.1 Indicators and Analyses of Market Power

In order to identify **market power**, different approaches and the use of different indicators are possible, cf. Twomey et al. (2005). Following a widely applied division in competition theory, indicators and analyses of market power may either assess the market structure, the market conduct (behaviour) or the market results (performance).

9.1.1 Indicators and Analyses of Market Structure

As market power increases with the firm size, a commonly used indicator is the market share m_i. Thereby not only the market share of the largest firm is considered but also the cumulative market shares, e.g. of the three or the five largest firms in the market. These are then labelled **concentration ratios** and denoted by CR_n, e.g. CR_3 for the cumulative share of the top three firms or CR_5 for the top five.

Another common measure of structural market power is the so-called **Hirshman–Herfindahl-index** (HHI). The market shares are hereby expressed in percentage points, and then, the sum of squared market shares is computed to obtain the HHI.

$$\mathrm{HHI} = \sum_i m_i^2 \tag{9.1}$$

By taking squares, the relevance of large firms is emphasised and the higher the HHI, the higher the (potential) market power. The maximum value for the HHI is 10,000 in the case of a single, monopolistic firm with 100% market share, and from values of 1800 onwards, a market is usually said to be highly concentrated.

The indices mentioned above are not specific to the electricity industry and are widely employed by competition authorities worldwide. However, given the non-storability of electricity, the use of such a static indicator may not be adequate to capture the role of market power in peak periods with little spare capacity. Specific indices have been developed to describe this aspect, notably the pivotal supplier index and the residual supplier index.

The **pivotal supplier index** indicates for each hour of the year, whether a specific supplier would be pivotal or not, i.e. if the hourly demand could be matched without that supplier. Therefore, it incorporates supply-side and demand-side characteristics, since in hours with low demand, no supplier usually is pivotal.

A disadvantage of the pivotal supplier index is the binary 0/1 assignment of the pivotal supplier status – being either true or not. Yet, when suppliers are close to being pivotal, they might already exert market power. Therefore, the **residual supplier index** is computed as the ratio of the market supply (sum of capacities K_j weighted with availabilities φ_{jt} except for the supplier i in question) and market demand D_t.

$$\mathrm{RSI_{it}} = \frac{\sum_{j;j \neq i} \varphi_{jt} K_j}{D_t} \tag{9.2}$$

Pushing even further, a complete residual demand curve is constructed in the so-called residual demand analysis. Yet this requires data that are not easily available to authorities in the European electricity markets.

9.1.2 Indicators and Analyses of Market Conduct

Typical behaviour when exerting market power includes physical or economic withholding of generation capacity. Physical withholding means that generation capacity is not bid at all into the market – pretexting, e.g. some necessity for maintenance works. Economic withholding, by contrast, corresponds to an output gap, i.e. at a given price, a unit produces less than would be profitable at that price. However, no single indicator is used to describe these phenomena and their detection typically requires manifold data and also model-based approaches.

9.1.3 Indicators and Analyses of Market Results

A frequently used indicator to evaluate market results is the so-called **Lerner index** LI, which sets the observed price markup over marginal costs C' in relation to the price p:

$$\mathrm{LI} = \frac{p - C'}{p} \tag{9.3}$$

Although approximate cost data are easily available, a precise estimate of the price markup, taking, e.g., start-up cost into account, requires a detailed competitive benchmark model, similar to the one required to identify economic withholding (cf. above). Sophisticated versions of the short-term market models discussed in Chap. 7 may thereby be used.

9.2 Simple Models of Imperfect Competition in Wholesale Markets

Whereas indicators of **market power** are primarily constructed to perform empirical analyses, a broad stream of literature has developed theoretical models to analyse market power in electricity markets and elsewhere over the last decades. These models are generally rooted in modern game theory, which has developed since the pioneering work of von Neumann and Morgenstern (1944) and Nash (1950) into a burgeoning field with multiple approaches of often considerable mathematical complexity. Yet, the literature on **imperfect competition** may even be traced back to nineteenth century mathematicians and economists Augustin Cournot and Joseph Bertrand. Their approaches are key elements in the classification of competition models shown in Fig. 9.1. Discussed initially in a framework with two players (duopoly), the Cournot and Bertrand models have been generalised to the case of multiple players (oligopoly) and rooted in the overarching concept of **Nash equilibrium** which forms the basis for most non-cooperative game-theoretical models. In the **Cournot model**, the strategic decision variable of the players is the sales quantity, whereas in the Bertrand model, the players decide on their sales prices. In the standard **Bertrand model**, the players tend to underbid each other and this process only stops when prices reach marginal costs, i.e. the same level as in **perfect competition**. This leads to the famous paraphrase of Bertrand competition: "Two is enough for competition".[1]

In the Cournot game, the players ration the quantities they bring to the market to increase their profit. Without any cartel formation or collusion, the profit-maximising strategies of the individual players lead to a non-cooperative equilibrium with higher prices and lower quantities than the competitive equilibrium.

A small example is considered to illustrate the key aspects of a **Cournot game**: this game includes N identical players with production cost $C(q_i)$ as a function of production q_i but without other constraints, such as capacity constraints. Demand is given by a linear inverse demand function, denoted as follows:

[1] Note that under certain circumstances, the Bertrand game becomes equivalent to the Cournot game. This is notably true, when the decision-making of the players is described as a two-stage problem: deciding on the capacities at the first stage and then on the sales price in the second stage, cf. Kreps and Scheinkman (1983).

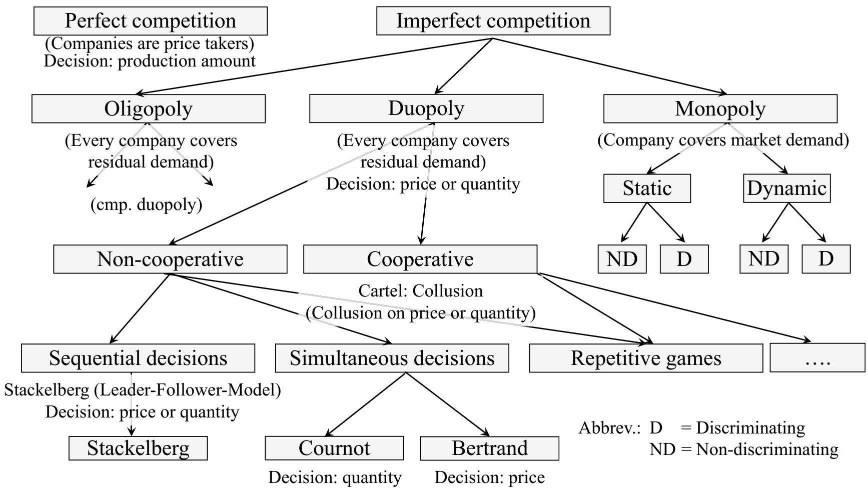


Fig. 9.1 Overview of simple models of competition

$$p = a - b \cdot D \tag{9.4}$$

All players i have the objective of maximising their profits π_i, given by the expression:

$$\pi_i = p\left(q_i | q_{j;j \neq i}\right) \cdot q_i - C(q_i) \tag{9.5}$$

They thereby optimise their profit given the decisions $q_{j;j \neq i}$ of the other players.

In equilibrium, demand equals the sum of all supplies and hence Eq. (9.4) may be rewritten as

$$p\left(q_i; q_{j;j \neq i}\right) = a - b \cdot \sum_j q_j \tag{9.6}$$

A general property of Nash equilibria is that all players achieve their respective optimum given the actions of the other players. Or to put it differently: all players select their reaction functions to respond to the action of the other players and from the set of all reaction functions, the equilibrium may be determined.

In the example, the reaction function may be obtained by inserting the price Eq. (9.6) into the profit function. This yields the following equation:

$$\pi_i = \left(a - b \cdot \sum_j q_j\right) \cdot q_i - C(q_i) \tag{9.7}$$

Then, first-order conditions for an individual profit maximum (and thus necessary conditions for an equilibrium) may be determined by taking the partial derivative:

$$\frac{\partial \pi_i}{\partial q_i} = \left(a - b \cdot \sum_j q_j\right) - b \cdot q_i - C'(q_i) \stackrel{!}{=} 0 \tag{9.8}$$

As the game is symmetric, a symmetric solution is to be expected, and this may be found by considering the total market quantity $Q = \sum_j q_j$ and the symmetric production quantities $q_i = Q/N$. One obtains the relationship:

$$a - b \cdot Q - b \cdot \frac{Q}{N} - C'\left(\frac{Q}{N}\right) \stackrel{!}{=} 0 \tag{9.9}$$

Under the additional assumption that the marginal cost C' is independent of the produced quantity, collecting and rearranging terms yield the solution:

$$Q = \frac{N}{N+1} \frac{a - C'}{b} \tag{9.10}$$

Inserting this result into (9.6) allows the price in equilibrium to be obtained:

$$p = \frac{1}{N+1} a + \frac{N}{N+1} C' \tag{9.11}$$

As expected, the oligopolists reduce the quantity compared to the competitive solution $Q = \frac{a-C'}{b}$. Correspondingly, the price in this equilibrium under imperfect competition is higher than in the competitive case, where p equals C'.[2]

It is worth noting that the solution of the Cournot game includes the perfect competition case as limiting case for $N \to \infty$. Also, the solution of the monopolistic case is obtained for $N = 1$.

The Cournot-Nash equilibrium is derived under the implicit assumption that all players take their decisions simultaneously. Alternatively, one may also consider that one player takes the lead (or several) and the remaining follow. This is the so-called **Stackelberg game**, after the German economist Heinrich von Stackelberg.

[2] In general, the inequality $a > C'$ holds. The economic interpretation is: a is the reservation price, i.e. the maximum willingness-to-pay of the consumers. If this reservation price does not exceed the marginal production cost, the quantity sold on the market will be zero. Then, the first-order-condition (9.13) will not hold and instead of an interior optimum the boundary solution $q_i = 0$ will be chosen.

9.3 Applications of Models of Imperfect Competition to Power Systems

Models of **imperfect competition** in electricity markets have mainly been developed to analyse strategic interaction in the wholesale market. Thereby six major types of approaches are subsequently briefly introduced (cf. also Weber 2005; Gabriel et al. 2013).

The first one makes use of the **Cournot-Nash framework** discussed in the previous section, cf. e.g. Andersson and Bergman (1995), Smeers (1997), Borenstein et al. (1999), and Helgesen and Tomasgard (2018). These models are used to analyse power market equilibria in the longer term. Thereby the assumption of electricity as a homogenous good is used, although possibly differentiated by location. Market equilibria are then expected to be determined through the capacity setting decisions of suppliers. Yet, the existence of a Cournot-Nash equilibrium with finite prices necessitates the own-price elasticity for electricity to be strictly negative. For prices that align with empirical observations, the price elasticity even has to be rather high. Most empirical studies, e.g. Labandeira et al. (2017) find significant negative own-price elasticities for electricity in the longer run, yet in the short run, the electricity price elasticity currently is very low or even non-existent – except for the limited number of consumers with specific price-based demand-side management measures (see Sect. 3.1.4) in place. Correspondingly, it is challenging to reconcile theoretical concepts and empirical observations in this framework.

A related approach for modelling strategic behaviour in electricity and other energy markets is the use of so-called **conjectured supply functions** (Day et al. 2002; Díaz et al. 2010). These allow the modelling of competition situations that are in-between pure Cournot-Nash equilibria and markets with full competition. Each strategic player's degree of oligopolistic behaviour is modelled by a scalar between 0 (full competition) and 1 (Cournot-Nash equilibrium). Usually, the model is calibrated by appropriate choices of these parameters to fit historical observations, cf. e.g. Spiecker (2013).

Another approach reflects more realistically the actual agent-behaviour in (auction-based) spot markets, namely the so-called **supply function equilibria**. This modelling approach for strategic interaction in the wholesale market is based on equilibria of firms bidding with supply and (possibly demand) curves into the wholesale market. Klemperer and Meyer (1989) developed this modelling framework, and Green and Newbery (1992) made an application to investigate the, then, freshly deregulated British electricity market. Bolle (2001), Hobbs (2001) and Green and Vasilakos (2010) are examples of further models based on that approach. It is thereby important to note that only in presence of demand uncertainty (or another source of uncertainty) supply function equilibria deviate from the traditional Cournot-Nash equilibrium (cf. Bolle 2001). In fact, the actual equilibrium given some prevailing uncertainties is undetermined, and it may range between the Cournot-Nash and the Bertrand equilibrium.

All problem classes considered so far in Chap. 7 and Sect. 9.2 except for the supply function equilibria may be viewed as special cases of a broad class of models called **mixed complementarity problems** (MCP), cf. Gabriel et al. (2013). MCPs are not formulated as optimisation problems but as a set of complementary constraints. In the case of optimisation models like those in Chap. 7, the complementary constraints correspond to the first-order conditions, the so-called **Karush–Kuhn–Tucker constraints**. For non-cooperative games, as those stated so far, the complementary constraints are obtained by taking the first-order conditions of the optimisation problems of each individual player plus market clearing constraints. Specialised solvers like the PATH solver (Ferris and Munson 2000) may be used to find solutions.

In more recent years, **bi-level models** have gained increased interest in operations research and electricity market modelling. They describe cases where the solution to an upper-level optimisation, e.g. by a strategic electricity market player, depends on the outcomes of a second, lower-level optimisation, which may represent another player's actions or the equilibrium in a market. Note that the above mentioned **Stackelberg game** (cf. Sect. 9.2) is a simple example of a bi-level programme. Applications of bi-level models to electricity markets include the participation of electricity retailers in a demand response market environment (Zugno et al. 2013), the optimal offering strategy of virtual power plants (Kardakos et al. 2016) or generation capacity expansion in competitive markets (Wogrin et al. 2011). Closely related to bi-level programmes are so-called **MPEC**s, standing for **mathematical programmes with equilibrium constraints**. In these models, the lower level consists of equilibrium constraints, which frequently correspond to the Karush–Kuhn–Tucker conditions of an optimisation problem. In that case, the MPEC is equivalent to a bi-level programme. A generalisation of MPECs is **equilibrium problems with equilibrium constraints (EPEC**) which include not only one but several optimisation problems at the upper level. An application to the electricity market field is (Hu and Ralph 2007). For EPECs, there may be no single equilibrium in pure strategies, and so these are among the games that are difficult to solve, although the Nash theorem asserts that a solution exists. However, this may be a solution in mixed strategies, i.e. the players choose with some probabilities between different strategies.

A rather different type of models used to investigate interactions between different players in the electricity market are so-called **agent-based simulation models**, cf. for reviews (Weidlich and Veit 2008; Ringler et al. 2016). In contrast to the game-theoretical approaches, these models are not designed to identify Nash-type equilibrium outcomes for electricity markets. Rather they typically use a combination of heuristic decision rules, simple market clearing algorithms and (albeit not always) reinforcement learning. Examples include Bunn and Oliveira (2001), Veit et al. (2009), Li and Shi (2012) or Deissenroth et al. (2017), with applications ranging from strategic behaviour of traditional generation companies to bidding behaviour of wind energy producers and integration of renewables into the electricity market. The strength of these approaches is certainly their flexibility, yet a weakness is the difficulty to provide rigorous foundations to the behavioural assumptions implemented.

9.4 Imperfect Competition in Retail Markets: Modelling Customer Switching Behaviour

Energy consumers are expected to play an essential role in the transition to a sustainable energy system. **Retail markets** are yet frequently neglected when analysing current electricity markets or future electricity systems. But with more decentralised electricity production and correspondingly more prosumers, the end-users of electricity and their choices deserve more attention.

Retail markets may be analysed from a variety of perspectives and using different conceptual approaches. From a marketing and sales perspective, the design of products, prices, placement and promotion may be put forward – these are the classical 4 Ps of marketing management (Kotler 2000, p. 64). In a more system-oriented approach, the dynamics in retail electricity markets could be analysed using diffusion models (cf. the seminal concepts by Rogers 1962; Bass 1969), notably when it comes to the adoption of innovations. Both perspectives build on the (implicit) assumption that electricity is more than a homogenous good traded at a uniform price,[3] as is usually assumed in simple models of wholesale electricity markets. Since the physical good is homogenous (in one place at one point in time), the heterogeneity of retail electricity products is rather a consequence of the commercial packaging: this may include the contract duration, the price structure (time-invariant vs time of use…), the billing frequency, the payment conditions, the combination with other services (e.g. electricity and gas bundle), the integration of self-produced electricity and so on.

When analysed with traditional models of **imperfect competition**, this emphasis on product heterogeneity also provides an interesting angle on retail markets. This third perspective, rooted in game theory, is the focus of the remainder of this section. It emphasises neither the perspective of a single company as in marketing nor the dynamics of the market as in diffusion and innovation models, but rather considers strategic interaction between various players and the resulting market equilibrium.

The starting point of the analysis is the observation that firms in retail markets usually set their sales prices and not the quantities sold, leading rather to a Bertrand-type of game (see Sect. 9.2). There are results stating the equivalence of price (Bertrand) and quantity (Cournot) competition under certain circumstances (cf. e.g. Kreps and Scheinkman 1983). Nevertheless, the analysis of the price competition on the retail market is worthwhile, if one takes as a starting point that not only prices influence the consumers' choices, but also further, less directly

[3] Uniform price at least at one point in time and in one given location (price zone), see Chap. 7.

observable characteristics of the offered products (e. g. service quality, billing frequencies, personal idiosyncrasies). This results in models of oligopolistic competition with product differentiation (e.g. Anderson et al. 1995). We subsequently show that this provides an approach that captures salient features of customer behaviour in competitive electricity retail markets.

9.4.1 Basic Model with One Retail Market Segment

To describe the competition with differentiated products, a non-cooperative game with N profit-maximising players i (suppliers) is considered (cf. Weber 2005, pp. 80–90). These provide a bundle of differentiated products, with one product y_i supplied per player. The total demand Y for the bundle of products is fixed, but demand will shift among products depending on the prices p_{i} offered by the different players.

The demand for the individual products y_i could be approximated by different types of functions. The use of a logistic function enables an at least partly analytical treatment along with lower and upper bounds of 0 and Y, respectively. This leads to the specification:

$$y_i = D_i(p_1, p_2, \ldots, p_N, Y) = \frac{e^{\alpha_i - \beta p_i}}{\sum_{j=1}^{N} e^{\alpha_j - \beta p_j}} Y \tag{9.12}$$

This functional form is also widely used as so-called multi-nomial logit model to describe discrete choices in other fields (cf. e.g. Train 2009). It can be rooted in a model of stochastic utility maximisation which is used both in psychology (e.g. Luce 1959) and microeconomics (cf. notably the work of Nobel Prize winner Daniel McFadden).[4]

For easier interpretation as well as notational convenience, the market share m_i relative to total demand $Y = \sum_{i=1}^{N} y_i$ is introduced.

$$m_i = \frac{y_i}{Y} = d_i\left(p_i; p_{j;j \neq i}\right) = \frac{e^{\alpha_i - \beta p_i}}{\sum_{j=1}^{N} e^{\alpha_j - \beta p_j}} \tag{9.13}$$

The price responsiveness of demand is thereby dependent on the parameter β. To obtain normal demand functions with negative own-price elasticity, β has to be positive. Larger values of β then correspond to a higher price responsiveness.[5]

[4] Seminal works are McFadden (1974, 1978). Early applications of the model include the choice of heating systems, notably Dubin and McFadden (1984).

[5] A straightforward generalisation of the model could be the introduction of individual price sensitivities β_i for the products offered by the different players.

The parameter α_i reflects the relative attractiveness of the product offered by player i given identical prices. The market share of product i at identical prices for all products increases with larger values of α_i. For $p_i = p$ for all i, one obtains as market shares:

$$m_i = \frac{e^{\alpha_i}}{\sum_{j=1}^{N} e^{\alpha_j}} \tag{9.14}$$

As in the Cournot game described in Sect. 9.2, all players i maximise their profit π_i.

$$\pi_i = p_i y_i(p_i) - C_i(y_i(p_i)) \tag{9.15}$$

Yet now, the decision variable is the sales price p_i, which is optimised accounting for the procurement or production cost $C_i(y_i)$, with corresponding marginal costs $C_i'(y_i)$ and taking into consideration the reaction in demand y_i in response to the price changes. The first-order condition then yields, after accounting for Eq. (9.12), the following equality:

$$\frac{\partial \pi_i}{\partial p_i} = y_i + \left(p_i - C_i'(y_i)\right)\frac{\partial y_i}{\partial p_i} = 0 \tag{9.16}$$

Using Eqs. (9.12) and (9.14), the following identity is obtained after some transformations:

$$\frac{\partial y_i}{\partial p_i} = -\beta_i y_i + \beta_i \frac{y_i^2}{Y} = -\beta_i y_i (1 - m_i) \tag{9.17}$$

Inserting this into the first-order condition leads to the equality:

$$y_i\left(1 - \beta_i\left(p_i - C_i'(y_i)\right)(1 - m_i)\right) = 0 \tag{9.18}$$

If the possibility $y_i = 0$ (market exit) is excluded, prices and market shares in the competitive equilibrium must fulfil the identity:

$$\beta\left(p_i - C_i'(y_i)\right)(1 - m_i) = 1 \tag{9.19}$$

For constant marginal costs $C_i'(y_i) \equiv c_i$, this is the equation of a hyperbole in the m_i–p_i-plane with asymptotes $p_i = c_i$ and $m_i = 1$ (see Fig. 9.2). This becomes more obvious when p_i is written as a function of m_i:

$$p_i = c_i + \frac{1}{\beta(1 - m_i)} \tag{9.20}$$

This is the inverse supply function for player i that describes the price setting $p_i = s_i^{-1}(m_i)$ of player i for a given market share m_i.

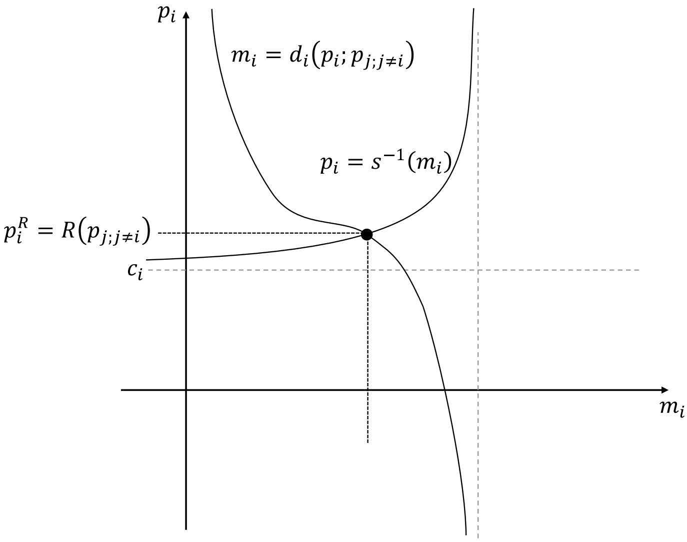


Fig. 9.2 Demand function $m_i = d_i(p_i; p_{j;j\neq i})$, inverse supply function $p_i = s_i^{-1}(m_i)$ and resulting price reaction $p_i^R = R(p_{j;j\neq i})$ in a non-cooperative game of retail competition. *Source* own illustration based on Weber (2005)

In this plane, the demand function (9.12) may also be depicted as a function of the own price p_i at given prices $p_{j,j\neq i}$ of the competitors. It corresponds then to a (rotated) sigmoid with asymptotes $m_i = 0$ and $m_i = 1$. The exact location of the demand function is thereby determined by the values of the parameters α_i, $\alpha_{j,j\neq i}$ and β depending on the parameter values and the prices of the other players. The demand increases with decreasing p_i, yet even at a price of 0, a player i will not reach a market share of 100% – this corresponds to an imperfect substitutability between the different products.

The hyperbolic inverse supply function $p_i = s_i^{-1}(m_i)$ and the (inverse) sigmoid demand function $p_i = d_{\boldsymbol{i}}^{-1}(m_i; p_{j;j\neq i})$ have a single intersection point which provides a unique price p_i as a function of the prices $p_{j;j\neq i}$ of the competitors. Computing this optimal price p_i for all possible combinations of $p_{j;j\neq i}$ then corresponds to determining the reaction function $p_i^R = R(p_{j;j\neq i})$ of player i to the prices $p_{j;j\neq i}$ set by its competitors.[6]

[6] It is not possible to give an explicit representation of this reaction function $p_i^R = R(p_{j;j\neq i})$, because this would necessitate to solve the following equation:

$$1 - \frac{1}{\beta(p_i - c_i(y_i))} = \frac{1}{1 + e^{-\alpha_i + \beta p_i}\left(\sum_{j;j\neq i} e^{\alpha_j - \beta p_j}\right)}$$

The left side thereby is obtained from solving Eq. (9.19) for m_i, and the right side corresponds to a rearrangement of the market share function (9.13), in which p_i appears only once. As p_i appears

The same reasoning may be applied for all the players i, which leads to a set of N equations with a unique solution.[7] If these first-order conditions are fulfilled, the second derivatives of the profit functions π_i with respect to the own prices p_i are negative. This indicates that profits are quasi-concave in own prices, and consequently, the unique solution corresponds to a profit maximum for each firm and hence a Nash equilibrium.

9.4.2 Extension to Several Retail Segments

The model presented in the previous section may be readily extended to the case of several customer groups or retail market segments s, e.g. households, commercial and industry (Weber 2005). Without providing the full formal details here, it may be just noted that we obtain as an equivalent of Eq. (9.19) the following generalised relationship:

$$p_{i,s} = C_i'\left(\sum_s y_{i,s}\right) + \frac{1}{\beta_s\left(1 - m_{i,s}\right)} \tag{9.21}$$

This result is remarkable in many respects:

- The price of supplier i in market segment s equals the marginal cost of that supplier (first term on the right hand side (RHS) of Eq. (9.21)) plus a markup (second term of the RHS).
- The marginal costs, as specified here, are dependent on the total quantity sold by supplier i across all market segments.
- By contrast, the markup does not depend on marginal costs but only on the market share $m_{i,s}$ and the parameter β_s. This parameter, as stated above, describes the price sensitivity of the customers in market segment s.

The last point has an interesting implication: the price charged to customers in equilibrium will depend on their price sensitivity. If customers easily switch suppliers – even for small financial gains – suppliers will offer them a better price. Customers who are not sensitive to prices will be charged higher prices.

This model result is in line with empirical observations. Operation margins for suppliers are tiny in segments with highly price-sensitive customers, e.g. large industrial customers. Higher margins are observed in other segments, e.g. for retail and small commercial clients, cf. BNetzA and BKartA (2018). But even within household customers, suppliers try to segment customer groups according to their price sensitivity. This leads to low prices offered on online shopping platforms

both inside an exponential function and outside, this is a so-called transcendental equation, for which explicit solutions are only available in special cases.

[7] For the symmetric case with constant marginal costs, the existence and uniqueness of the market equilibrium are proven by Anderson and de Palma (1992).

compared to the prices charged to customers via other retail channels. Even separate brands may be created to specifically target different customer groups (e.g. Yello as a second brand of EnBW). Similar marketing and sales strategies can also be found in other sectors, e.g. in the field of mobile phone services. However, it is important to be aware of such aspects when prices and offers on the end-user markets are interpreted.

Also in terms of identifying and measuring market power, this result has an interesting implication. The optimal pricing Eq. (9.21) includes a markup over marginal cost, and when constructing a **Lerner index** $\mathrm{LI}_{i,s}$ along the lines of Eq. (9.3), we get the result:

$$\mathrm{LI}_{i,s} = \frac{p_{i,s} - C_i'}{p_{i,s}} = \frac{1}{\beta_s\left(1 - m_{i,s}\right)} \cdot \frac{1}{p_{i,s}} \tag{9.22}$$

At first sight not surprisingly, the oligopolistic game leads to a price markup. Yet a closer look reveals that even for the limiting case $m_{i,s} \to 0$ a strictly positive markup inversely proportional to β_s persists. So in this case even the limiting case of perfect competition will not lead to a complete absence of price markups. Rather only markups beyond the competitive outcome of $\frac{1}{\beta_s}$ should be considered as a sign of imperfect competition.

9.5 Workable Competition and Market Monitoring

In view of the various indicators and models established to describe imperfect competition, the question of the policy implications arises. Thereby it has to be clear that the concept of **perfect competition** can generally not be the guideline and objective of competition policy and competition authorities in practice, given the underlying strong and idealistic assumptions. Rather, practitioners usually refer to the concept of **workable** or **effective competition** (e.g. Mecke 2018; Bender et al. 2011). This concept starts by distinguishing the three layers of market structure, market conduct (also called market behaviour) and market results (cf. Sect. 9.1). Workable competition is then a combination of these three layers that leads to a dynamic competitive process, where first-moving innovators may gain a competitive advantage. Yet, this advantage is eroded over time through competitive pressure emanating from competitors' (imitating or alternative) behaviour.

With this complex definition referring to the dynamics of competition, it is obvious that none of the indicators discussed before will on its own provide perfect evidence on the existence or absence of abuse of market power. Instead, an investigative process is needed to obtain evidence regarding the different indicators, to analyse the evidence found and eventually to draw conclusions and take appropriate action to improve the state of competition.

A key issue at the beginning of such an investigative procedure is the delimitation of the relevant market. This is also true for electricity markets: Is the relevant market for electricity the national one, or does a pan-European market exist? Depending on the delimitation, the indicators on market structure will provide quite different results. E.g. the generation market share of EDF as the largest generator is around 80% in France, but only around 20% if the entire European Union is considered.

For investigative processes into the state of competition to be handled appropriately, an institution in charge of these procedures is furthermore needed. These are the **competition authorities**, which in many leading economies worldwide are established as state agencies with a strong legislative mandate, but independent of the government. Examples are the Federal Trade Commission (FTC) in the US and the Bundeskartellamt in Germany. A notable exception to the rule of independence from the executive branch is the Directorate General and the Commissioner for Competition in the European Union.[8]

An appropriately designed and competent competition oversight for the electricity sector also has to consider the specifics of the sector. The question then is whether this oversight should be transferred to the general competition authority or to some unique entity for the electricity sector. In most jurisdictions with deregulated electricity markets, the task is eventually split. The general competition authority handles general aspects of competition, such as the control of mergers and acquisitions in the electricity sector. In contrast, more operational types of anti-competitive behaviour, such as price markups or capacity retention, may be handled by electricity market specific entities. Yet sensible differences arise between the centrally organised markets relying on an ISO as in the US (cf. Sect. 10.8) and the decentralised markets as established in Europe. The ISO as a market operator has access through its daily operations to most market-relevant data of the firms such as bids, availabilities and start-up costs. Typically, it then has its own market monitoring unit that regularly investigates firms' anti-competitive behaviour, or it delegates this task to an external consultant. There is no such central market operator in the European markets, and the power exchanges and other trading platforms are themselves privately-owned companies. They are mandated to check and report fraudulent trading behaviour, yet do not have the resources and authority to check whether the bids include price markups. Anti-competitive operational behaviour is therefore primarily only investigated when corresponding allegations are put forward. There have been extensive so-called sectoral inquiries, e.g. at the European level (EC 2007) in the first decade of the twenty-first century, which were carried out in close cooperation between the general competition authority and the energy market regulators. Moreover, monitoring reports on major market developments, including competition aspects, are presented annually by the regulators (e.g. BNetzA & BKartA 2018).

[8] This may be explained by the historical evolution of the European institutions, which have only gradually evolved from intergovernmental agencies into an executive branch of the European Union. And still today, the European Union has only a limited mandate provided by its member states and not full executive and legislative competences—although these competences are large in the fields of the internal (European) market and competition issues therein.

Besides establishing an independent competition authority, the facilitation of **market entry** is a second general recipe that should be applied in view of workable competition – also in the electricity industry. Not only the effective market entry of new competitors improves the market outcomes, but also the pure threat of a market entry by some foreign or domestic competitor exerts pressure on the incumbents to operate their business efficiently and to offer competitive prices to their customers, cf. e.g. Baumol et al. (1982).

Looking at the mathematical and numerical models used to describe market outcomes in imperfectly competitive electricity markets (cf. previous sections), the two general recipes derived for competition policy have a clear implication: if the role of competition authorities and market entry are not considered in the models, the model outcomes rather indicate the direction of possible detrimental effects of market power and provide an upper bound to their strength than a realistic estimate. This has to be kept in mind when interpreting the results of such models.

9.6 Further Reading

The basic concepts of oligopolistic competition are explained in most economic textbooks.

Varian, H. (2019). Intermediate Microeconomics – A Modern Approach. 9th edition. New York: WW Norton.

This is a seminal text by Hal Varian that provides an overview on multiple topics at an intermediate mathematical level.

Gabriel, S. A., Conejo, A. J., Fuller, J. D., Hobbs, B. F., & Carlos, R. (2013). Complementarity Modelling in Energy Markets. New York: Springer.

This book introduces the rather general concept of complementarity problems along with applications to energy systems. Also further related problem classes such as mathematical programmes with equilibrium constraints (MPECs) are discussed.

Green, R., & Newbery, D. (1992). Competition in the British Electricity Spot Market. Journal of Political Economy, 100, 929–953.

This is the original paper by Richard Green and David Newbery applying the concept of supply-function equilibria to then newly liberalised British electricity markets. Various sequels have extended the original setting to cover, e.g. intermittent renewable infeed, e.g.:

Green, R., & Vasilakos, N. (2010). Market behaviour with large amounts of intermittent generation. Energy Policy, 38, 3211–3220.

9.7 Self-check of Knowledge and Exercises

Self-check Questions

1. Which three-part sequence is typically used in competition theory for analysing market power? What are indicators used to quantify market power at the three levels?
2. Name the two basic game-theoretic models used to model imperfect competition and discuss their advantages and shortcomings, notably in applications in the power sector.
3. Why is competition in retail markets frequently imperfect and how may it be modelled?
4. Describe institutional aspects and guiding principles of market monitoring in practice.

Exercise 9.1: Indicators of market structure
Figure 9.3 gives market shares for electricity generation in Great Britain in 2019. Use these data to compute the concentration ratios CR_1, CR_3 and CR_5 and the Hirshman–Herfindahl index (HHI). For the computation of the HHI assume that the shares of others are subdivided equally among 25 small players.

Is this market highly concentrated according to general standards? May international competition foster competition?

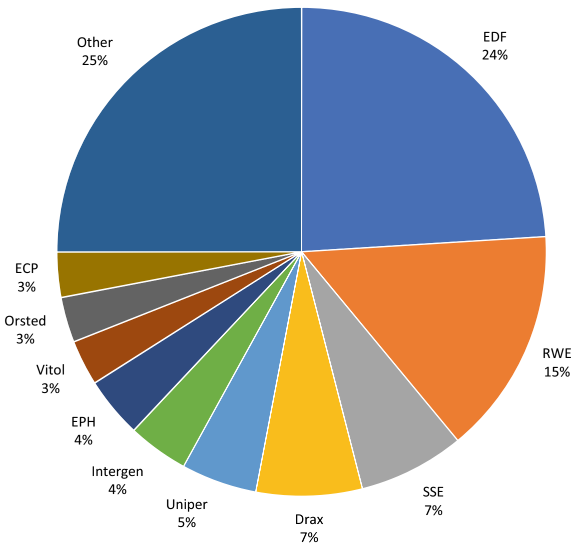


Fig. 9.3 Wholesale electricity generation market shares by companies in GB 2019. *Source* Ofgem (2021)

Exercise 9.2: Cournot-Nash equilibrium

Imperfect competition is frequently assumed for the fossil energy carriers oil and gas. Therefore, we apply the Cournot model of Sect. 9.2 to a simplified model of gas supply to EU27.

Considered producers and corresponding production and transport cost:

- Russia: 1.3 ct/kWh
- Norway: 1.2 ct/kWh
- Northern Africa: 1.5 ct/kWh.

Characteristics of demand:

Annual imports D_0: 4200 TWh.

Average import price p_0: 2 ct/kWh.

Own-price elasticity η: 0.4.

To determine the parameters used in the inverse demand function (9.4), you may use the following calibration formulas:

$$b = -\frac{D_0}{p_0 \cdot \eta} a = p_0 \cdot \left(1 - \frac{1}{\eta}\right) \tag{9.23}$$

Note: These formulas may be derived by combining Eq. (9.4) and the standard definition of the elasticity $\varepsilon = \partial q / \partial p \cdot p/q$ for the calibration point (D_0, p_0).

(1) Develop an Excel model for the described gas market. Implement the formulas given in Sect. 9.2 and above and insert the appropriate parameter values. Note that the Eq. (9.9) and the following ones have to be modified somewhat to accomodate for different marginal costs for the different producers. Hint: you may add up the individual first-order conditions from Eq. (9.8) over all market participants to obtain one equation for the overall quantity Q.

(2) Determine the Cournot quantity and prices for this case.

(3) Compute additionally the produced quantities for each country. What do you notice? Also determine the contribution margins [ct/kWh] and the profit in billion € for each producer.

(4) What happens to quantities and prices if an additional importer (e.g. LNG from the US with production costs of 2.6 ct/kWh) enters the market?

(5) Restart from the "base case" (without LNG from the US). Analyse what happens if you change the elasticity parameter to 1, 0.1 and 0.01. What are the impacts on prices and quantities?

Exercise 9.3: Retail competition

The UK regulator OFGEM closely monitors the competition in the British energy markets. Key results for the British retail electricity market in 2020 are given in Table 9.1.

Table 9.1 Average tariff prices and electricity supply market shares by suppliers in GB in 2020

Supplier	Supplier's cheapest tariff (customer bill per year for medium consumption in GBP)	Market share (%)
British gas	968	18.2
E.ON	859	12.0
EDF	878	10.7
Scottish power	918	9.1
Npower	878	6.5
Shell energy	857	2.7
OVO energy	880	15.3
Utility warehouse	906	1.9
Green network energy	856	1.4
Bulb	915	5.7
Avro energy	817	1.6
Octopus energy	895	5.1

Source Ofgem (2021)

Use these data to estimate the parameters for the imperfect competition model discussed in Sect. 9.4.1. Assume thereby that the data reflect a market equilibrium situation and that all suppliers can procure electricity at similar, constant cost c_0 from the wholesale market.

(1) Show that the market equilibrium condition (9.19) may be reformulated under the assumptions above as

$$p_i = c_0 + \frac{1}{\beta} \cdot \frac{1}{1 - m_i} + \varepsilon_i \tag{9.24}$$

(2) Run a linear regression using the above data for the model:

$$p_i = a + b \cdot \frac{1}{1 - m_i} + \varepsilon_i \tag{9.25}$$

(3) Use the results of the regression to determine the value of the parameters c_0 and β in Eq. (9.24).
(4) What would be in that market equilibrium the expected price charged by a small player with market share (close to) zero?
(5) Give reasons why the results of this small econometric exercise may not provide an accurate indication regarding the propensity of GB customers to switch suppliers.

References

Anderson, S. P., & de Palma, A. (1992). The logit as a model of product differentiation. *Oxford Economic Papers, 44*, 51–67.

Anderson, S. P., de Palma, A., & Nesterov, Y. (1995). Oligopolistic competition and the optimal provision of products. *Econometrica, 63*, 1281–1301.

Andersson, B., & Bergman, L. (1995). Market structure and the price of electricity: An ex ante analysis of the deregulated Swedish Electricity Market. *The Energy Journal, 16*, 97–110.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science, 15*, 215–322.

Baumol, W. J., Panzar, J., & Willig, R. D. (1982). *Contestable markets and the theory of industry structure*. Harcourt.

Bender, C. M., et al. (2011). "Effective competition" in telecommunications, rail and energy markets. *Intereconomics, 46*, 4–35.

BNetzA & BKartA. (2018). *Monitoringbericht*. Available at: https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Energie-Monitoring-2018.pdf [Accessed May 31, 2019].

Bolle, F. (2001). Competition with supply and demand functions. *Energy Economics, 23*, 253–277.

Borenstein, S., Bushnell, J., & Knittel, C. R. (1999). Market power in electricity markets: Beyond concentration measures. *The Energy Journal, 20*, 65–88.

Bunn, D. W., & Oliveira, F. S. (2001). Agent-based simulation—An application to the new electricity trading arrangements of England and Wales. *IEEE Transactions on Evolutionary Computation, 5*, 493–503.

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

Deissenroth, M., Klein, M., Nienhaus, K., & Reeg, M. (2017). Assessing the plurality of actors and policy interactions: Agent-based modelling of renewable energy market integration. *Complexity*, art. 7494313.

Díaz, C. A., Villar, J., Campos, F. A., & Reneses, J. (2010). Electricity market equilibrium based on conjectural variations. *Electrical Power System Research, 80*, 1572–1579.

Dubin, J. A., & McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica, 52*, 345–362.

EC. (2007). *DG competition report on energy sector inquiry (SEC(2006)1724*, January 10, 2007. Available at: https://ec.europa.eu/competition/sectors/energy/2005_inquiry/full_report_part1.pdf [Accessed June 11, 2022].

Ferris, M. C., & Munson, T. S. (2000). Complementarity problems in GAMS and the PATH solver. *Journal of Economic Dynamics and Control, 24*, 165–188.

Gabriel, S. A., et al. (2013). *Complementarity modelling in energy*. Springer.

Green, R. J., & Newbery, D. M. (1992). Competition in the British electricity spot market. *Journal of Political Economy, 100*, 929–953.

Green, R., & Vasilakos, N. (2010). Market behaviour with large amounts of intermittent generation. *Energy Policy, 38*, 3211–3220.

Helgesen, I., & Tomasgard, A. (2018). An equilibrium market power model for power markets and tradable green certificates, including Kirchhoff's Laws and Nash-Cournot competition. *Energy Economics, 70*, 270–288.

Hobbs, B. F. (2001). Linear complementarity models of Nash-Cournot competition in bilateral and POOLCO power markets. *IEEE Transactions on Power Systems, 16*, 194–202.

Hu, X., & Ralph, D. (2007). Using EPECs to model bilevel games in restructured electricity markets with locational prices. *Operations Research, 55*, 809–827.

Kardakos, E. G., Simoglou, C. K., & Bakirtzis, A. G. (2016). Optimal offering strategy of a virtual power plant: A stochastic bi-level approach. *IEEE Transactions on Smart Grid, 7*, 794–806.

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

Kotler, P., Keller, K. L., & Chernev, A. (2022). *Marketing management* (16th ed., Global ed.). Pearson.

Kreps, D. M., & Scheinkman, J. A. (1983). Quantity precommitment and Bertrand competition yield Cournot outcomes. *The Bell Journal of Economics, 14*, 326–337.

Labandeira, X., Labeaga, J. M., & López-Otero, X. (2017). A meta-analysis on the price elasticity of energy demand. *Energy Policy, 102*, 549–568.

Li, G., & Shi, J. (2012). Agent-based modeling for trading wind power with uncertainty in the day-ahead wholesale electricity markets of single-sided auctions. *Applied Energy, 99*, 13–22.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

McFadden, D. L. (1978). Modelling the choice of residential location. In A. Karlquist (Ed.), *Spatial interaction theory and residential location* (pp. 75–96). North Holland.

Mecke, I. (2018). *Wirksamer Wettbewerb*. Available at: https://wirtschaftslexikon.gabler.de/definition/wirksamer-wettbewerb-51067/version-274271 [Accessed November 2, 2019].

Nash, J. (1950). *Non-cooperative games* (Unpublished PhD Thesis). Princeton University.

Ofgem. (2021). *Retail market indicators*. Available at: https://www.ofgem.gov.uk/data-portal/retail-market-indicators [Accessed January 9, 2021].

Ringler, P., Keles, D., & Fichtner, W. (2016). Agent-based modelling and simulation of smart electricity grids and markets – A literature review. *Renewable and Sustainable Energy Reviews, 57*, 205–215.

Rogers, E. M. (1962). *Diffusion of innovations*. Free Press of Glencoe.

Smeers, Y. (1997). Computable equilibrium models and the restructuring of the European electricity and gas markets. *The Energy Journal, 18*, 1–31.

Spiecker, S. (2013). Modeling market power by natural gas producers and its impact on the power System. *IEEE Transactions on Power Systems, 28*, 3737–3746.

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge University Press.

Twomey, P., Green, R., Neuhoff, K., & Newbery, D. (2005). *A review of the monitoring of market power* (Cambridge Working Papers in Economics 0504). University of Cambridge.

Veit, D. J., Weidlich, A., & Krafft, J. A. (2009). An agent-based analysis of the German electricity market with transmission capacity constraints. *Energy Policy, 37*, 4132–4144.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour*. Princeton University Press.

Weber, C. (2005). *Uncertainty in the electric power industry – methods and models for decision support*. Springer.

Weidlich, A., & Veit, D. (2008). A critical survey of agent-based wholesale electricity market models. *Energy Economics, 30*, 1728–1759.

Wogrin, S., Centeno, E., & Barquin, J. (2011). Generation capacity expansion in liberalized electricity markets: A stochastic MPEC approach. *IEEE Transactions on Power Systems, 26*, 2526–2532.

Zugno, M., Morales, J. M., Pinson, P., & Madsen, H. (2013). A bilevel model for electricity retailers' participation in a demand response market environment. *Energy Economics, 36*, 182–197.

Electricity Markets in Europe 10

After the theoretical explanation of markets and their functioning in the preceding chapters, this chapter introduces the different European electricity markets and explains the basic principles of these markets and how they are interlinked with each other. Since liberalisation, rules of the markets have been adjusted several times and adapted to new fundamental situations. A good example is the introduction of intraday markets, which got relevant with higher shares of renewable energies and the need to balance their uncertain day-ahead forecast in a further market before real-time balancing.

This chapter starts with an overview of spot markets, including day-ahead and intraday markets and cross-border trading. Spot markets generally act as a reference for the other markets. But other markets are also crucial for a proper operation of the electricity system. The role of the different markets and mechanisms – derivative markets, control reserve markets, provision of system services, capacity mechanism and congestion management – are explained in the adjoining subchapters. Furthermore, retail markets and their functioning are described. Besides sections on retail contract types, competition in retail markets and energy poverty, key ratios (self-supply, autonomy) for the characterisation of decentralised energy sources are also introduced. As electricity markets in Europe are differently designed than in North America, this chapter ends with a comparison of the design of European and North American markets.

Key Learning Objectives

After having gone through this chapter, you will be able to

- Describe the basic principles of European electricity markets and how they are interlinked with each other.
- Describe the difference between continuous and spot market trading making use of the example of intraday and day-ahead markets.
- Understand how cross-border trading is organised in Europe.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_10

- Explain what power reserves are distinguished and how these are organised.
- Describe how congestion management is organised and which methods are used for capacity allocation and alleviation.
- Explain the basic functioning of retail markets and corresponding contract types.
- Understand competition on retail markets and how retail prices are formed.
- Explain the indicators self-supply, grid parity and level of autonomy.
- Explain the difference between European and US power markets.

10.1 Spot Markets

When electricity markets in Europe were deregulated, spot market trading usually occurred the day before delivery. Correspondingly, the term spot market is frequently used synonymously to **day-ahead markets** (cf. Sect. 10.1.1). But more recently, intraday trading has become increasingly important (Sect. 10.1.2) and also cross-border trade has been developed much further (Sect. 10.1.3).

10.1.1 Day-Ahead Markets

The market price in the day-ahead market is determined by matching offers from the supply and demand side. Supply is primarily provided by generators and demand is stemming mainly from energy utilities and large retailers who serve end consumers (see, e.g., in Sect. 7.1.4). In an **auction-based market** (see Sect. 8.3), a supply and demand equilibrium and the corresponding market clearing price is determined, usually on an hourly – partly also on a quarter-hourly – time interval for the 24 h of the following day. As prices are calculated for every single hour of a day, in general, 24 prices for electricity on the next day are determined.

Trade on the day-ahead market is generally organised with a fixed closing time, e.g. 12 o'clock on EPEX SPOT, where all collected bids are matched and a unique market price for every hour of the following day and physical delivery at a given location is derived. It is thereby not only possible to trade single hours of the following day but also to submit so-called **block bids**.[1] Multiple hours of a day can be combined to block bids and various standardised products combining several hours (e.g. base- or peak-load) are already defined on the energy exchanges. These combined orders refer to different hour contracts for the same day and delivery

[1] Block bids and multi-part bids are summarized under the term complex bids. For multi-part bids see Sect. 10.8.

location and have to be either buy or ask bids. All orders of block bids are either executed entirely or not executed at all (so-called Fill-or-Kill criterion). This type of bidding is especially relevant for less flexible technologies and smaller portfolios, where the operator has to decide to turn on a unit for several consecutive hours to fulfil minimum operation hour requirements (see Sect. 4.4.1.3). Even if the market price for a single hour is below the accepted price of a block bid, this single hour may be accepted as part of a block bid. However, this will only happen if the average price over the combined hours is higher than the bid for the hour block. Else the bid will not be executed at all. This allows, for example, to commit the whole capacity of a plant for eight consecutive hours instead of risking a shut-down of the plant in hours with low electricity prices (in the case of bidding every single hour separately). This is an advantage for the market participants yet comes at the expense of higher computational complexity for the energy exchange. Notably, handling such bids requires the introduction of binary variables, leading to a mixed-integer model for the market clearing (see also Sect. 4.4.1.3).[2]

Several **day-ahead markets** for electricity exist in Europe: electricity, e.g., from Austria, France and Germany is traded on the EPEX SPOT in Paris, electricity from Scandinavian countries on the Nord Pool. Further energy exchanges are Omie for the Iberian Peninsula, PolPX for Poland, PXE for Czech Republic, Hungary, Romania and Slovakia. Also, Great Britain and Ireland initially traded their electricity purely nationally.

Over the years, electricity markets have undergone several significant developments. Besides changes in the underlying generation mix (such as the increased penetration of renewables and the nuclear phase-out in countries like Germany), significant changes have been made in regulation and market design. The introduction of the European emission trading system in 2005 (see Sect. 6.2.4.1) has created additional interdependencies and affected the pricing in electricity markets. Also, the introduction of negative prices and the establishment of competitive procurement procedures for reserves (see Sect. 10.3) have affected price formation and market outcomes. Furthermore, provisions have been made for the rare cases when the day-ahead auction leads to tight market conditions (very low or very high prices). Then a second auction is held, which allows participants to adjust their bids to facilitate market clearing. However, the most significant changes over the last two decades have been the introduction of intraday markets and an increasingly international market coupling, as discussed in the following sections.

10.1.2 Intraday Markets

Day-ahead markets can be considered the reference markets in most power systems and result in an allocation close to that at delivery or real time. However, after closing the auction, new information can emerge within the time frame between the

[2] This may even lead to paradoxically rejected block bids (for details see, e.g., Madani and Van Vyve 2014).

closure of the day-ahead market and physical delivery. Forecast updates for weather-dependent renewables like wind and solar are one source of new information. With higher penetration of these renewables, significant deviations may arise between the forecasted and the final feed-in. Yet, power plant outages or unexpected changes in demand also induce information updates, which cannot be handled in day-ahead markets. So-called **intraday markets** therefore provide the possibility to react to changes in forecasts over time by allowing market participants to trade electricity close to physical delivery.

Whereas in the Scandinavian countries, some intraday trade has been practised for more than twenty years in the ELBAS market and on the regulatory power market platform run by the Scandinavian TSOs, mediated trading in continental Europe only started around 2010. It has been expanding considerably, at least in Central West Europe (Germany, Benelux and France). Contrarily to the day-ahead market, it is organised as continuous trading, which enables an immediate reaction to updated information, be it a power plant outage or an updated wind forecast.

Nowadays, intraday trading is possible in many European countries until 60 min or even less before delivery. Day-ahead and intraday markets for electricity are organised as double-sided markets, allowing buyers and sellers to submit offers. In the day-ahead market for electricity, buyers and sellers submit their offers to a power exchange, which determines the price periodically, in general once per day for the 24 h of the following day. The power exchange uses the submitted sell orders to construct a supply function and the submitted buy orders to find a demand function for electricity. The intercept of both functions results in the price. In contrast to that, continuous trading delegates the clearing process to the market participants. The power exchange only provides a platform that gives information to the participants and enables them to conclude transactions with each other. This platform is usually the "**open order book**" with limit orders for purchases and sales (see Sect. 8.3). Through market orders, i.e. orders with unlimited prices, traders may then directly and on a continuous time scale execute trades – or alternatively include their own limit orders to increase the available volume in the order book. In general, European intraday markets are organised as **continuous trading**, sometimes there is also an **opening auction**, e.g. in Germany.

10.1.3 Cross-Border Trading

As with other goods, trade in electricity between two or more countries will increase the welfare of all participating countries as market participants have access to a larger market for sales or purchases (cf. also Sect. 7.2).[3] As electricity trade is only possible via cross-border power lines, two questions are crucial when

[3] Each country engaging in trade will increase its overall economic surplus. In the simple models of Sects. 7.2 and 7.3, this corresponds to a decrease in cost – when costs are adjusted for the value of imports and exports. Yet not necessarily all market participants within the countries will benefit from cross-border trading. Typically, producers in high-price markets lose profits whereas consumers benefit from lower prices. The opposite is true in low price areas.

organising electricity trade between two countries: How should the trade be organised? And how much transport capacity between countries can be considered for trading?

Allocation methods like first-come-first-served and/or pro-rata, which were applied before deregulation and in the early years thereafter fail the criteria of being non-discriminatory and market-based. Today, the applied methods may be distinguished by several aspects (see also Sect. 10.6). Two main mechanisms to allocate capacity between two neighbouring markets can be distinguished: explicit versus implicit auctions. While some years ago, explicit allocation schemes were state-of-the-art in Europe, nowadays implicit auctions are dominating. **Explicit auctioning** means that the right to cause a power flow over interconnections between countries is auctioned to the market separately and independently from the marketplaces for electricity trading. Hence, traders have to buy a transmission right to implement a trade between two countries and they have to enter two separate trades on the markets of the two countries. The transmission operators generally care about the contracts and allows for exchange as long as physical interconnection capacities are not exceeded. Capacity may be auctioned at different time scales, e.g. in annual, monthly and daily auctions. Such trading requires a low degree of integration and coordination between the involved grid and market operators. Furthermore, there is no necessity for a common trading platform nor simultaneous clearing of day-ahead markets between countries under such a market mechanism. However, as prices for both markets and the price for the transmission right are not known ex-ante, this mechanism may lead to an inefficient result. Frequently, interconnector capacity is not fully used and cross-border traders have failed to correctly anticipate the price spread between countries, sometimes resulting in acquiring the transmission capacity in the "wrong" direction. Despite this being recognised for day-ahead markets, long-term trading of cross-border capacities is still based on explicit auctioning in Europe. Also here, a coordinated capacity calculation was introduced in Europe. The Network Code on Forward Capacity Allocation establishes rules for long-term cross-border capacity assignments. These are intended to enable market participants to secure capacity on cross-border lines up to several years in advance. These capacities often correspond to physical transmission rights, as opposed to financial transmission rights used in nodal markets (see Sect. 10.6.1). By allowing such deals a long time in advance hedging against congestions is enabled. A so-called **joint allocation office (JAO)** has been established, which provides a single auction platform and point of contact to facilitate the purchasing and selling of transmission capacity.

In contrast to these explicit auctions, **implicit auctioning** means that cross-border capacities are included in a centralised clearing of local power exchanges. Market clearing for transmission rights and electricity occurs simultaneously and the resulting prices per area reflect both the cost of energy and congestion. If no congestion occurs, prices in the two countries are equal (cf. Sect. 7.2). In consequence, market participants only submit bids to the marketplaces in their respective country. As part of the auction process, cross-border bids and, as a result, deliveries are automatically generated by the system, which aligns the resulting prices of the two countries as far as available capacities allow. If, for example, separate auctions in Germany and France have as a result that the daily price in Germany is lower than in France, the affected power exchanges automatically generate a delivery from Germany to France which either completely aligns the price or, if this is not possible, at least utilises all short-term cross-border capacities from Germany to France. In consequence, the best possible alignment of short-term prices is achieved (see Sect. 7.2). Further aspects of organising trade between countries, such as market coupling versus market splitting, volume coupling versus price coupling, available transfer capacity versus flow-based methodologies and zonal versus nodal prices are discussed in the section on congestion management in electricity markets (see Sect. 10.6).

10.2 Derivative Markets

Besides trading on the spot markets, the European markets also offer numerous possibilities to trade **derivatives**. Table 10.1 provides an overview of the exchanges offering derivative trading for electricity market products in European countries and the corresponding trading volumes.

Several points are worth noting:

- Derivative trading is possible in most countries. The only exceptions are Malta and Cyprus and the South-East European countries Bulgaria, Croatia and Slovenia. Derivative trading is usually organised as continuous trading with opening and closing auctions, as is typical for financial markets.
- In some countries, multiple exchanges propose trades in power derivatives, even if there is only one spot market operator. There has been substantial consolidation across Europe and beyond in the active exchanges. Besides the Germany-based European Energy Exchange (EEX), the US-based Intercontinental Exchange (ICE) and NASDAQ are the most important players. NASDAQ notably has overtaken the power derivative trading in the Nordic and Baltic markets.
- The trading volume on OTC markets exceeds in many countries the volume traded on power exchanges. The opposite is yet true in the Nordic and Baltic markets.
- Options are much less traded than forwards and futures – in some countries, even no option products are offered. And when they are offered, trading is lower by a factor of 20 or so than the trades in forwards and futures. Nevertheless, the concept of options and the techniques for their valuation are useful for dealing with flexibilities (cf. Chaps. 8 and 11).
- The churn rate is a useful measure to compare trading volumes among countries of different sizes. It corresponds to the quotient of the trading volume and the annual consumption. These churn rates strongly vary among countries. In general, trading volumes and churn rates are the highest for large countries.

Table 10.1 Derivative trading for electricity in the EU, Norway Switzerland

Country	Active exchanges[a]	Trading volume [GWh]			Churn rate[d] (all products together) (%)
		Exchange traded futures	OTC traded forwards	Options	
Austria	**EEX**, ICE, NASDAQ	150	246	4	635
Belgium	**EEX**, ICE	13	6	–	23
Bulgaria	–	–	–	–	0
Croatia	–	–	–	–	0
Cyprus	–	–	–	–	0
Czech Republic	PXE	18	123	–	249
Denmark	NASDAQ	74	54	?[b]	408
Estonia	NASDAQ	223	16	?[b]	564
Finland	NASDAQ	110	80	?[b]	237
France	EEX	83	786		198
Germany	**EEX**, ICE, NASDAQ	1233	2026	30	635
Greece	EEX	0.3			0
Hungary	PXE, HUPX	7	144		434
Ireland	SEM	?[c]	?[c]		0
Italy	EEX, **GME**, ICE, IDEX	161	205		127
Latvia	NASDAQ	16	12	?[b]	415
Lithuania	NASDAQ	20	15	?[b]	393
Luxembourg	**EEX**, ICE, NASDAQ	15	24		635
Malta	–	–	–	–	0
Netherlands	EEX, ICE, **NASDAQ**	118	205		303
Norway	NASDAQ	315	229	?[b]	497
Poland	PXE, **POLPX**	163	78		194
Portugal	OMIP	6	–	–	13
Romania	EEX	0.1	13		33
Slovakia	PXE	1	0		4
Slovenia	–	–	–	–	0
Spain	**OMIP**, EEX	31	224	0.3	105
Sweden	NASDAQ	310	225	?[b]	428
Switzerland	EEX	1	155		265
United Kingdom	ICE, NASDAQ	5	301		97

[a] The incumbent market operator is marked in **bold**. In most cases, it has by far the largest market share
[b] No separate indication is given in the source
[c] According to the source, anecdotal evidence points at low volumes
[d] In this context, the churn rate is a measure of market liquidity. A churn rate of 200% means that the traded volume of power is equal to two times the electricity consumption on the observed markets
Source ECA (2015, pp. 107–112) and own adaptations

But derivative trading is also vital in the Nord Pool market (Nordic and Baltic countries).

- A churn rate of 1000% is sometimes postulated as a minimum requirement for liquid trading – which makes it attractive for pure financial players. No European power derivative market reaches this churn rate and according to the numbers provided, even a churn rate of 500% is only reached in the (then existing) common market area Germany–Austria–Luxemburg and in Estonia. In recent years, power derivative trading has even seen further slight decreases. This is partly attributable to additional regulatory requirements related to European directives put into place in the aftermath of the global financial crisis of 2008.[4] Here the balance between the necessary preservation of financial stability and overregulation of relatively small players (in the context of financial organisations) has to be found.

10.3 Management of Reserves

As discussed in Sect. 5.1.4.2, various so-called **ancillary services** or **system services** are required to operate the electricity network. Among those, reserves used for **frequency control** become more and more procured on a market basis. In recent years, the European regulation has not only issued harmonised definitions of different reserve power categories, but also the European Network Codes[5] streamline the corresponding operational procedures in view of cross-border competitive procurement. These define three processes along with the corresponding types of **reserves**:

FCP: Frequency Containment Process. As indicated by its name, this process aims at maintaining the grid frequency within an acceptable range around its set-point of 50 Hz. This is done by automatically activating the so-called **FCR**: **Frequency Containment reserves**.[6]

FRP: Frequency Restoration Process. Whereas the FCP aims at limiting frequency deviations, the **FRP** has the objective to re-establish the frequency at 50 Hz while at the same time also restoring the inter-area power flows to their scheduled values. Thereby two types of reserves are used:

[4] EMIR: European Market Infrastructure Regulation. MiFID II: Markets in Financial Instruments Directive. REMIT: Regulation on Wholesale Energy Market Integrity and Transparency. While the first two are applicable to a broad range of financial derivatives, the last one specifically applies to the energy sector and imposes increased reporting requirements on energy traders.

[5] Cf. Guideline on electricity transmission system operation CR 2017/1485 (EC 2017a) and Guideline on electricity balancing CR 2017/2195 (EC 2017b).

[6] Cf. Sect. 5.1.4.2 for this and the other reserve categories.

- **aFRR**: **Frequency Restoration Reserves** with **automatic activation**
- **mFRR**: **Frequency Restoration Reserves** with **manual activation**.

RRP: Reserve Replacement Process. As the third step in reserve management, the RRP may re-establish the previously activated reserves. This is done using so-called **RR: Replacement Reserves**.

Note that the *RRP*, in contrast to the first two processes, is not mandatory across the EU. For example, since the deregulation of the electricity market in late 1990, Germany has not had a process for reserve restoration. Furthermore, these processes are supplemented in the European Network Codes by processes for exchanging reserves between TSO areas. The TSOs handle these processes, yet the assets used to provide the reserves are due to **unbundling** requirements (see Sect. 6.1) usually not under the direct control of the TSOs. Hence, the question arises of how these resources may be procured. Over the years, more and more market-based procurement mechanisms have been established and they have increasingly become international, too.

For these markets, the following three general challenges arise:

- the close **coordination** needed between short-term grid operation and generation (or more generally flexibility) operation
- the **technical restrictions** relating to the provision of **reserves** and the energy provision in (conventional and other) power plants
- the avoidance of excessive **market entry barriers** and the related danger of abuse of **market power**.

At the same time, several important design choices for reserve markets have to be made:

1. **Product design**
2. **Procurement periods**
3. **Prequalification requirements**
4. **Auction design**
5. **Bid remuneration approach**
6. **Auction timing**.

1. Product design: Since imbalances between feed-in and off-take from the grid may occur in both directions, one has to decide whether there should be one symmetric product including both reserves for upward and downward regulations, or instead separate products for upward regulation (positive reserve) and downward regulation (negative reserve). In continental Europe, FCR is procured as a single symmetric product since rapid activation in both directions is required. By contrast, all other reserves are procured using separate products to enable bids from different types of flexibilities, both from the demand and supply sides. So this contributes to lowering the market entry barriers.

2. Procurement periods: From a TSO perspective, a long procurement period is advantageous since reserve capacities are secured long in advance. However, in terms of the economic efficiency of the reserve markets, short procurement periods with as short as possible lead-times are advantageous. With short lead-times, suppliers better know both their units' availability and the opportunity costs associated with reserve provision. Similarly, short procurement periods also enable more targeted bids and will lead to a higher efficiency of the reserve markets (cf. Just 2011). Over time, we have seen a gradual reduction of the procurement periods – for aFRR in Germany from half-yearly bids down to four-hour block bids. This has fostered market entry and increased competition.

3. Prequalification requirements: Participation in the reserve markets usually necessitates the corresponding units undergo a prior prequalification. Thereby the TSOs notably check that the technical equipment of the units enables them to follow the activation signals received from the TSO. Very restrictive prequalification rules may contribute to a higher technical reliability of the power system. Yet, they also form barriers to market entry, especially for smaller and unconventional reserve providers such as storages and demand response. Therefore, tight prequalification requirements may contribute to a higher reliability but may reinforce the position of the incumbents and may enable them to exercise market power.

4. Auction design: The reserve markets operate as **single-sided multi-unit auctions**, i.e. with multiple sellers and the TSO(s) as a single buyer. This is different from the day-ahead electricity market with its double-sided auction (cf. Sect. 10.1). The central question is whether the providers should be paid for the capacity, energy or both. A capacity payment is analogous to an option premium, paid on financial markets, as it is received independently of the actual use of the reserve. The energy payment then corresponds to the strike price that is received when the flexibility is used.

Empirical evidence on the paid compensation is mixed. In Germany, FCR is only remunerated on a capacity basis, whereas aFRR and mFRR are paid both for the capacity and the energy they provide. In the Nordic countries, where predominantly mFRR was used in the past, it has been procured on the so-called regulating power market and remunerated purely on an energy basis.

These at first sight inconsistent findings may be explained by the differences in (opportunity) costs faced by the relevant providers. In the case of FCR, the symmetric product design and the over time rather balanced activation of positive and negative reserves imply that there are little energy costs. Capacity costs arise since the capacity may not be marketed on the power market. Since FCR has to be provided from spinning units, capacity costs may also arise because the units have to be kept in operation during periods with prices lower than variable costs – these are the opportunity costs associated with the must-run condition. Those also may arise when thermal power plants provide positive FRR. At the same time, the split products imply that positive energy costs will arise when positive reserves are activated. Conversely, the activation of negative reserves implies savings in fuel consumption for conventional power plants (since they are producing less) or additional electricity offered for consumption in case of demand response. Hence, a negative energy price for these reserves should be expected, i.e. paid by the reserve provider to the TSO. The Nordic system is hydro-dominated, and reserves are also mainly provided by hydropower plants (see also Sect. 8.5.5). At the same time, generation capacity is not scarce in these systems, rather the energy stored in the reservoirs. Therefore, it seems natural that capacity is not priced there but energy.

5. Bid remuneration approach: An aspect related to auction design is the payment principle. In multi-unit auctions, either each unit may receive the same (marginal) price – as in the day-ahead electricity market. Or each unit is paid its own bidding price – this is then called "**pay-as-bid**" instead of "**pay-as-cleared**". In the past, reserve procurement auctions have been frequently held using the pay-as-bid approach. The current European regulation advocates pay-as-cleared. At first sight, this may be considered inefficient since some providers are paid more than they are asking for. The advantage of pay-as-cleared is that suppliers have fewer incentives to submit non-cost-based bids. In fact, in a pay-as-bid market, suppliers base their prices on their best estimate of the marginal bid, so these markets may be named "guess the clearing price" (cf. Cramton and Stoft 2006; Swider and Weber 2007). This will foster collusion, i.e. anti-competitive behaviour, and raise entry barriers for newcomers and small firms in general since those have typically fewer competencies in forecasting the price.

6. Auction timing: Another key aspect for efficient market operation is the timing of reserve auctions relative to the day-ahead electricity market. A co-optimisation of energy and reserve provision will lead to the best results from an overall system perspective. And this is the approach implemented in competitive US markets (cf. also Sect. 10.8). In Europe, the markets are cleared separately, not least since the power exchanges and OTC trading platforms are institutionally separated from the grid operators. In contrast, the American Independent System Operators (ISOs) have joint responsibility for market and system operation. In continental Europe, the procurement auctions for reserves have traditionally been held before the day-ahead power market. This ensures sufficient liquidity on the reserve power markets[7] since unit commitment decisions have not yet been taken and units selected on the reserve power markets may sell their resulting must-run generation on the day-ahead market. At the same time, the unit commitments obtained as an outcome of the reserve market may be inefficient given the results of the day-ahead market. If rescheduling is not possible through portfolio-internal swaps or a secondary market, this will result in inefficiencies. Therefore, a later selection of power plants for reserve provision is, in principle, advantageous. It is typically implemented in markets where unit commitment decisions are less constraining and also further limitations to reserve provision like ramping constraints do not play an as important role – i.e. the Nordic countries. In other countries, there is increasingly the possibility for units not retained in the procurement auction for reserve capacities to submit a so-called free energy bid, i.e. to propose un-scheduled capacities after the day-ahead electricity market for activation as reserves.

[7] If overall supply adequacy is satisfied; see Sect. 10.5.

10.4 Provision of Other System Services

Besides load–frequency control, grid operators must perform other regular system operation tasks and be prepared to cope with fault and emergency situations. The most important tasks in normal operation are (see also Sect. 5.1.4):

- **voltage control** and
- **congestion management**.

In fault and emergency states, the following tasks have to be performed:

- **short-circuit management** and
- **restoration of supply**.

In future power systems with high shares of power electronics-based generation technologies such as solar and wind power, also the following tasks are expected to gain importance:

- **very short-term frequency stabilisation** and
- **grid (or frequency) forming**.

To perform these tasks, grid operators make use of the ancillary services discussed in Sect. 5.1.4.2. Subsequently, the possibilities for market-based procurement of these services are discussed.

Voltage control ensures that voltage at all grid nodes remains within predefined bounds. Voltage control is mainly performed by adjusting the reactive power infeed locally. Like congestion management, it is hence a task to be performed locally in the grid. Thus, the use of standard short-term markets without local discrimination is not adequate. At the same time, conventional generation units can provide reactive power within a broad range at relatively low costs. Therefore, the provision of reactive power is either mandatory for large-scale power generation units in European electricity systems or its provision and remuneration is dealt with through bilateral contracts. In addition, devices installed and operated directly by grid operators may be used for voltage control, e.g. tap-changing transformers or FACTS, cf. Sect. 5.1.4.2. In this case, market-based solutions to voltage control become even more questionable.[8]

Congestion management is dealt with in more detail in Sect. 10.6 and therefore not considered further here.

Short circuit management, as described above, relies on the overcurrent induced by faults and provided by the conventional large-scale generators. This mechanism is more or less a by-product of the technical characteristics of

[8] A profound overview of market mechanisms and remuneration concepts for voltage control is given in Hinz (2017, Chap. 3).

conventional generators and their operation mode. Correspondingly, no market mechanism is currently in place to handle this ancillary service, instead it is specified in the grid connection codes.

Restoration of supply capabilities are required to cope with the (hopefully unlikely) event of large-scale disruption in electricity supply. In such a case, units are needed that are able to black start. Grid operators (usually TSOs) have to ensure the availability of sufficient generation capacities with black start capabilities in their control area, and they will make emergency plans on how to rebuild an operating grid after such a large-scale failure gradually. As this ancillary service is rarely needed and requires appropriate location of resources and close coordination with the TSO, it hardly may be procured through short-term markets. It is rather more appropriately handled through negotiated bilateral contracts where competitive bids from distributed generation may play an increasing role in the future.

As discussed above, conventional generation units currently provide vital services to cope with the tasks mentioned earlier. With the transformation of the electricity system towards a system based predominantly on renewable energies, new concepts and solutions have to be developed to secure the stable operation of the grid (cf. Sect. 12.3). Moreover, some additional issues arise where separate ancillary services have not been defined so far.

Very short-term frequency stabilisation is currently provided by the **instantaneous reserve** that results from the **inertia** in the system (cf. Sect. 5.1.4.2). In conventional systems, the rotating masses of generators and turbines in large-scale units provide sufficient inertia to dampen frequency drops in the very short run and achieve an instantaneous reaction to imbalances. However, as wind turbines, solar panels and batteries are based on electronic DC-AC converters, they do not provide such inertia to the system. Consequently, it may be required to treat instantaneous reserve as an additional reserve category with a corresponding procurement market. This reserve could then, e.g., be procured from very short-term storage – yet the actual delivery of this service also requires methods of measurement that detect very rapidly upcoming imbalances (cf. e.g. MIGRATE 2018).

Grid (or frequency) forming is an additional requirement that synchronous generators currently deal with in large-scale conventional power plants. By rotating at the pre-specified synchronous frequency (50 Hz in Europe), the synchronous generators provide a sinusoidal voltage signal of precisely this frequency. The converters in power electronics-based systems are also capable of following an externally defined regular frequency signal. Without additional control concepts, they will not be able to generate or stabilise such a frequency signal. Hence, forming the grid frequency is an additional task to be performed by some grid elements in future grids. However, the regulatory and market framework for the provision of these ancillary services still has to be investigated.

10.5 Capacity Mechanisms

There are serious concerns about whether an **energy-only market** (EOM), where only the produced electricity is remunerated (€/MWh), can provide sufficient investment incentives to ensure (long-term) supply adequacy. **Supply adequacy** is thereby understood as the ability of the system to meet the energy requirements of all consumers, so-called long-term supply security. In contrast, supply security is the ability of the system to withstand disturbances, so-called short-term supply security (cf. e.g. Oren 2003).

In contrast to regulated electricity markets, no specific player is responsible for the provision of an adequate level of supply adequacy[9] in a liberalised energy-only market. This responsibility is handed over to market signals. A company will invest as soon as the expected rate of return due to market prices satisfies the individual investment profitability requirements.

In an EOM, times of scarcity are compulsory. In such scarcity hours, market prices are higher than the short-run marginal costs of the price-setting power plant. In reality, the market prices will be rather a function of the level of scarcity: the scarcer the market is, the higher the prices are. In real scarcity, the equilibrium price will be set by the demand side (see Fig. 7.1, outmost right demand curve). In such a situation, the market price will correspond to the **willingness-to-pay** (WTP) of the last served customer. This allows owners of power plants to recover their fixed operational and capital costs (see Sect. 7.4). However, due to imperfections of electricity markets, especially the lack of demand elasticity and the limited possibilities to control the real-time electricity flows (cf. e.g. Stoft 2002, pp. 14–16 and Joskow and Tirole 2007), the functioning of energy-only markets in reality is often seriously questioned. Furthermore, scarcity prices might be rather seldom. For example, they might only appear during a few hours of a freezing winter due to a significantly increased heat demand. Therefore, the corresponding revenue streams are highly uncertain. This holds especially true for electricity markets characterised by overcapacities during most hours of the year, e.g. due to an increasing share of fluctuating renewable electricity production, which is incentivised by additional support schemes (see Sect. 6.2.4.2), and market coupling activities. Therefore, an investment in new generation units will result in a high risk for the investor, and such volatile revenues may discourage investments.[10] Furthermore, very high price spikes – even if they only arise for a few hours of the year – are sometimes seen to be politically unacceptable, resulting in governmental intervention by introducing price caps (upper limits) for electricity prices, which might lead to the so-called missing-money problem and according to a discouragement of potential investors (see, e.g., Cramton and Ockenfels 2012; Hogan 2005 and Sect. 7.4.1).

[9] In this context, first the question has to be answered, how an adequate level of supply adequacy is to be defined (quantitatively).

[10] This might especially be the case for traditional energy companies that are often characterized to be extremely risk-averse. On the role of risk aversion for generation investment see e.g. Neuhoff and De Vries (2004).

In general, a possibility to manage the volatility of wholesale electricity prices and the corresponding risks would be an increased forward contracting between generators and retail companies or load-serving entities (LSE), which would buy these contracts for their customers (see de Vries and Hakvoort 2004, pp. 7–9). With the help of such long-term contracts, electricity customers could be protected against very high and power generators against very low electricity prices. Yet such contracts might increase the costs and therewith also the prices for electricity sold by companies participating in these long-term markets. This might reduce their sales volumes as long as customers are free to choose their retailer. Furthermore, as power flows to specific customers can currently hardly be limited (without limiting all power flows in the corresponding district), even customers bound by contract to another retailer (and not having bought these long-term contracts) would possibly be supplied in the case of scarcity.

One possibility to avoid the flaw caused by the limited possibility to control power flows is to allow competition only on the generation and not on the retail side (see Newbery 2002, pp. 30–32). Then the obligation to order an adequate level of supply adequacy would be given to load-serving entities. These entities would have to be the supplier for all customers in a specific region and would purchase supply adequacy for them. The obvious disadvantage of such a solution is that customers may not freely choose their suppliers.

Another solution could be to let customers become an active part of the system, which has only partly been realised in electricity markets. This could be achieved with the help of so-called capacity subscriptions (see, e.g., Grande et al. 2001), which would allow customers to choose the desired level of reliability. Again the lack of real-time control of power flows has so far hindered the use of such contracts, with the help of which consumers could choose their reliability level according to their needs. In the future, installing smart meters with load-limiting devices could remedy this flaw, transforming the good **reliability** into a **private good** (see Table 6.4). On the other hand, such a solution means that customers would have to estimate their capacity demand, which would require that consumers have to look into the details of their power consumption in a much more intensive way (which is not very likely for most (small) customers).

Another opportunity to provide less risky revenue streams for power plant operators is to establish remuneration mechanisms to influence the installation of electricity generating capacity (Joskow 2008), so-called **capacity mechanisms**. By establishing an additional revenue stream (€/kW) for power plant operators, the need for high wholesale electricity prices to provide sufficient investment incentives will be reduced or even totally eliminated. These remuneration mechanisms can be differentiated according to various aspects; most important seem to be at least the following two:

- targeted versus market-wide remuneration mechanisms; whereas only selected technologies would be addressed with a targeted mechanism, all units (e.g. generation, storage, demand side) could participate in a market-wide mechanism.

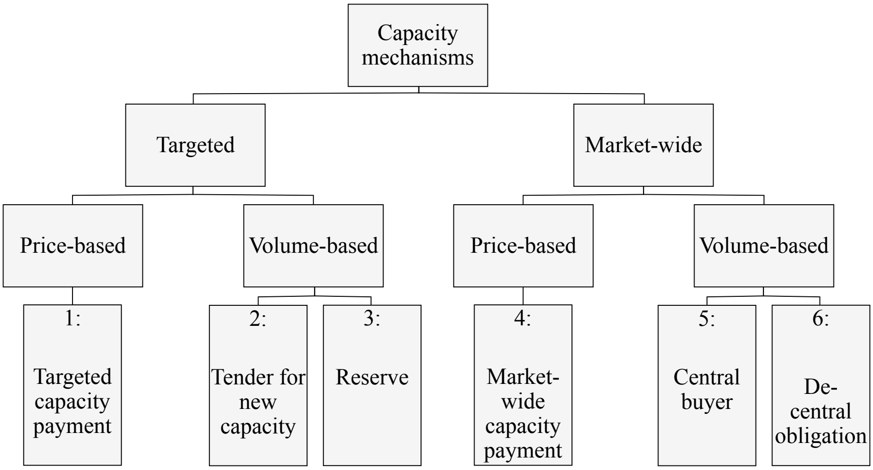


Fig. 10.1 Different forms of capacity mechanisms. *Source* Own illustration based on EC (2016, p. 50)

- volume-based versus price-based mechanisms; in a volume-based mechanism, the regulator sets the required capacity and the price is fixed through the market. Conversely, the remuneration price is exogenously set in a price-based mechanism.[11]

Figure 10.1 shows the classification of capacity mechanisms used by the European Commission (cf. EC 2016, p. 50). An overview of capacity remuneration mechanisms in place in Europe and the USA is given in Bublitz et al. (2019).

A straightforward and easy way to establish a new revenue stream for the provision of electricity generating capacity is that an authority pays a fixed price per megawatt of installed (or available) capacity (see price-based mechanisms 1 and 4 in Fig. 10.1). The challenging questions here are which units are eligible for this payment (all or only those fulfilling specific criteria) and how the price per megawatt is determined.

Within the category of volume-based mechanisms, many different concepts exist (for more information see, e.g., ACER 2013; Bublitz et al. 2016; Höschle 2018; de Vries 2004):

- **Tender for new capacity** (2 in Fig. 10.1): by establishing a tender, the construction of new units, e.g. new power plants, is supported in order to establish the additionally needed capacity. Once the installations have been connected to the grid, they can either be integrated into the energy markets or be further supported through a power purchase agreement.

[11] Depending on the slope of the supply and demand functions for capacity, the effects of an error in setting the price or the quantity lead to a bigger deviation from the equilibrium (see e.g. Oren 2000).

- **Reserve** (3 in Fig. 10.1): in a so-called strategic reserve, the required capacity is contracted by the (transmission) system operator, e.g. using a competitive tendering process. The strategic reserve is held back from spot and control reserve markets and is only used in emergency cases. This means the reserve will be activated through instructions of the system operator when there is a shortage of generation capacity in the market or when a given threshold concerning wholesale electricity prices is exceeded. Typically, this will result in rather old power plants to be transferred into a strategic reserve. This capacity mechanism can lead to the problem that more and more capacities are needed in the strategic reserve, once power plants leave the regular markets to become part of this reserve (a so-called slippery slope). Another kind of capacity reserve is the so-called operating reserve. Here the system operator contracts the required capacity by using frequent, e.g. daily, auctions.
- **Central buyer** (5 in Fig. 10.1): in such a centralised capacity market, a central buyer is responsible for calculating and procuring the required capacity. The dispatch of the corresponding units is not within his field of responsibility. To open the concept for new units, the bidding process for the capacities should be realised a considerable time before the delivery period. In so-called forward capacity markets, the capacity required is tendered some years in advance, giving multi-year contracts to new units and yearly contracts to existing units (see, e.g., Cramton and Stoft 2006; Bhagwat et al. 2017). In general, the challenge of the central buyer concept is to calculate the capacity demand, which should be sufficient to secure supply adequacy. To realise this, the central buyer might develop a so-called downward sloping capacity demand curve, which starts with a price cap that is often derived from the costs for new peak power plants (Cost of New Entry, **CONE**) and ends with a price of zero for the maximal demand level (see Fig. 10.2 and, e.g., Höschle 2018, pp. 46–48). The intersection of this demand curve and the supply curve, determined, e.g., with the help of a (Dutch) auction, leads to the capacity market clearing price and the contracted capacity.

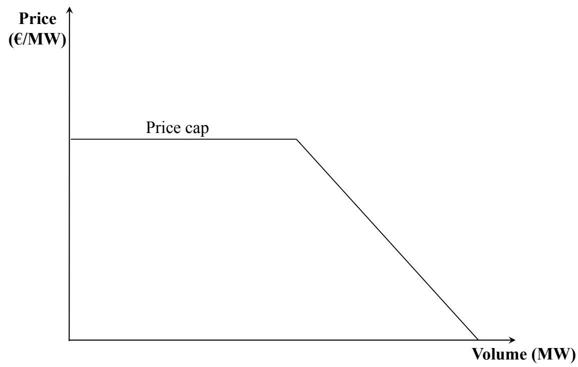


Fig. 10.2 Simplified typical capacity demand curve. *Source* Own illustration based on Bhagwat et al. (2017)

If generators have sold their capacity but are not producing during scarcity situations, they have to pay a penalty. Sometimes this incentive to produce during scarcity situations is seen to be insufficient, which has led to an important refinement of centralised capacity markets: the so-called **reliability options** (see, e.g., Pérez-Arriaga 1999). Under such a concept, the capacity owners sell call options to the central buyer, giving him the right to be compensated by the difference between the electricity wholesale price and the option's strike price. Therefore, the central buyer has not only to determine the demand of capacity needed, but also the strike price, which should be higher than the highest marginal cost of the capacity units available. Reliability options lead to an incentive to make capacities, whose electricity production has been sold via a call option, available during scarcity situations.

- **Decentral obligation** (6 in Fig. 10.1): an obligation is placed on all load-serving entities to secure the total capacity they need to meet their consumers' demand. In contrast to the central buyer model, a central planning authority is not required, and a central bidding process is not needed. Instead, individual contracts between load-serving entities and capacity providers might be negotiated.

The fact that the costs of excess capacity are typically much lower than the costs of undersupply supports the introduction of a capacity mechanism, yet the problems of designing these mechanisms are manifold. Obviously, there is a risk of regulatory failure. The ongoing redesign of capacity mechanisms worldwide illustrates the challenge to define adequate levels for the different parameters (e.g. the price in price-based and the demand in volume-based instruments). Furthermore, the cross-border effects of capacity mechanisms, e.g. due to lower wholesale prices in countries with capacity mechanisms, have to be considered.

10.6 Congestion Management in Electricity Markets

Congestion in the electricity system means that the existing power lines cannot realise all electricity flows requested by market participants. Accommodating these flows could violate existing physical (thermal) limits of the lines and transformers of TSOs or DSOs. But even if a congestion is identified, this will not necessarily mean that thermal limits are violated because the system operators detect a congestion with the help of load flow calculations considering the N-1 criterion[12] (so-called contingency analysis; see Sect. 5.1.4). There are manifold possibilities to classify mechanisms for **congestion management**, e.g. one might differentiate:

[12] This means that even in the case of a failure of one system element, a stable system operation is feasible.

- according to the technical **reason responsible for the bottleneck** into active power induced and reactive power induced congestion management,
- according to the **location** of the congestion into inter-zonal and intra-zonal congestion management,
- according to the **voltage level**, in which the congestion happened, into DSO or TSO congestion management,
- according to the **consideration of physical electricity flows** into flow-based and not flow-based congestion management, or
- according to the **"lead time"** into capacity allocation and congestion alleviation methods (cf. Androcec and Wangensteen 2006).

Subsequently, we structure the discussion along the distinction between capacity allocation and congestion alleviation methods.

10.6.1 Capacity Allocation Methods

Capacity allocation methods (also called long-term[13] or ex-ante[14] congestion management methods) aim at allocating the existing transmission (or distribution) capacity in an optimal way a certain period before the capacity is physically used. To allocate existing capacities, first, the maximum available capacity needs to be estimated. Therefore, limits for physical electricity transfer have to be calculated. Depending on the level of incorporation of physical power flows, the calculation procedure might be very demanding. Rather simplified calculations consider only characteristic load flow situations at typical days between two regions, assuming a direct pathway. Here the cross-border exchange is gradually augmented until the given restrictions are reached. To avoid overestimating the existing capacities while using such simplified methods, a reliability margin is subtracted from the calculated total transfer capacity (TTC), leading to the so-called net transfer capacities (NTC).

Flow-based methods try to consider the existing physical restrictions of the grid in a more elaborated way. An AC load flow or a linearised DC load flow model can be used to calculate the input data for flow-based market mechanisms. With the help of such a model, nodal Power Transfer Distribution Factors (PTDFs) are estimated (see Sects. 5.1.2.3 and 7.3). Nodal PTDFs show the influence of a change of the power infeed at a specific node on the power flows at critical branches in the grid. Depending on the power flow model used, AC PTDFs or DC PTDFs are computed (see Sect. 5.1.2.3). Furthermore, so-called Generation Shift Keys (GSKs) are estimated, which are used to predict how the production of a generation unit is affected by changes in the balance of the zone in which this unit is located. These GSKs are then used to transform the nodal PTDFs into zonal PTDFs. In addition to zonal PTDFs, an estimation of the still available transmission capacity for each

[13] A (very) long-term measure not considered in this chapter would be to build new transmission capacities.
[14] Ex-ante in this context means before clearing of the (day-ahead) market.

critical branch – the so-called Remaining Available Margin (RAM) – is needed. For calculating these parameters, information about the power grid is again necessary, underlining the need to involve the TSOs (and possibly also DSOs) into the process. The use of PTDFs makes it possible to represent power flow restrictions in a more realistic way than by using NTC values. Nevertheless PTDFs are still an approximation and only valid under certain assumptions, e.g. static framework conditions (e.g. no topology changes (Duthaler et al. 2008)).

In general, the available transmission capacity can then be allocated via non-market-based mechanisms (e.g. priority rules like "first-come, first-serve(d)") or market-based ones (e.g. auctioning of transmission capacities). In an explicit auctioning mechanism, separate markets for electricity and transmission capacity are put into place (cf. Sect. 10.1.3). This means that market participants have to submit bids for electricity and for physical transmission rights (PTR). As there are two separate auctions, bidders will lack information about the other commodity prices, which may lead to inefficient utilisation of transmission capacities. In particular, it can happen that market players do not use the PTR they purchased by an auction; e.g. they only purchased a PTR to prevent the use of it by their competitors, so that netting of opposite electricity flows[15] cannot be realised. This can be avoided by introducing compensation payments if the rights are not used or the obligation to return the PTR in the case of not using it ("use it or lose it"). The auction process will lead to revenues obtained by the TSO that have to be taken for building new capacities or reducing use-of-system charges.

So-called implicit auctions might be used to avoid the information problem of explicit auctions. Under such a scheme, electricity and transmission capacities are traded together. Market participants in different regions do not have to trade transmission capacity but just make bids for buying or selling electricity at their exchange. The different markets are coupled (so-called **market coupling**), so that the orders from different markets can be exchanged. If sufficient transmission capacity is available, the wholesale electricity price will be the same in the different markets.[16] Otherwise, the electricity price difference shows the cost of congestion (cf. Sect. 7.2). To realise this, the exchange operators in the different regions must have information about the restrictions of the grid to consider the available transmission capacity in the market clearing process.

Market coupling leads to the same results as a **market splitting** mechanism, another form of implicit auctioning. The difference to market coupling is that there is only one exchange operator responsible for the different markets under a market splitting regime. As soon as congestions appear, the system splits into different markets. These concepts are also called **zonal pricing**, where different prices arise between zones (as soon as transmission capacities are scarce), but a uniform electricity price is maintained within the zone. Zonal borders should correspond to the bottlenecks of the specific transmission situations, meaning that zonal cuts

[15] Netting means that electricity flows over the same line in opposite directions offset each another.
[16] The interested reader is referred to a more detailed description of the market coupling optimisation problem, which can, e.g., be found in Ringler (2016, p. 109).

depend on the transmission situation and should therefore be set dynamically. The current zonal pricing approach in Europe does not consider these dynamics but consists of (static) zones normally corresponding to national borders.

Considering all existing congestions – not only those between different zones as in zonal pricing – leads to the concept of **nodal pricing** (also called **locational marginal pricing** (**LMP**)). The principle of nodal electricity pricing may notably be traced back to Schweppe et al. (1988). These prices not only include generation costs but also the costs of transmission losses and congestions. A node in this context might be every location where electricity is fed into or withdrawn from the grid. Nodal prices represent the locational value of electricity (cf. Sect. 7.3), setting, on the one hand, the right incentives for investment decisions and guarantying, on the other hand, the optimal dispatch. As there is a need for a central system operator in charge of clearing the market considering network constraints, this kind of market design requires a central dispatch market (cf. Sect. 10.8). An obstacle for implementing nodal prices might be the corresponding distributional effects related to the fact that the prices might differ substantially at two different nodes (potentially even located next to each other).[17] So-called Financial Transmission Rights (FTRs) can be introduced to hedge against such price differences. FTRs are typically allocated with the help of auctions[18] and give their owners the right to receive payments according to the congestion rent, if a congestion and for this reason different electricity prices occur (see, e.g., Kunz et al. 2016).

10.6.2 Congestion Alleviation and Redispatch

Congestion alleviation methods (which might also be called short-term or ex-post[19] congestion management methods) aim to manage expected congestions on a shorter time frame (see, e.g., Kunz 2013), typically after the clearing of the day-ahead market. On a short-term basis, e.g. based on results of their grid operation planning, system operators can partly alleviate congestions by grid-specific measures, like **topology changes** (**switching operations**), to directly influence the load flow. By switching transmission lines on and off or by using, e.g., **flexible AC transmission systems** (**FACTS**; see Sect. 5.1.3.2), the power flow can be actively channelled through the existing network. In addition to these rather technical, non-costly measures, market-related measures like redispatch and countertrading can be used.

[17] In some markets, nodal prices are only used for the generation side, whereas on the consumption side the nodal prices are aggregated e.g. to zonal prices. Aggregating nodal prices, e.g. across a region, is often used to limit consumer price risk exposure (cf. Neuhoff and Boyd 2011, pp. 7–8).

[18] In these auctions, it has to be guaranteed that only feasible FTRs are issued by running the so-called Simultaneous Feasibility Test (SFT) (for more information see e.g. Hedman et al. (2011)).

[19] Ex-post in this context means after clearing of the (day-ahead) market. Sometimes this form of congestion management is also called curative congestion management, which might lead to confusion as curative actions can also be seen as post-fault actions (see, e.g., Hoffrichter et al. 2019).

In the case of **redispatch**, the system operator relieves transmission system overloads by giving instructions to installations located in front of and behind the congestion to adjust their production or demand; e.g. generators located in front of the congestion (in the so-called surplus region) have to reduce their output (negative redispatch), generators behind the congestion (in the so-called deficit region) have to increase their output (positive redispatch). It should be mentioned that by redispatching generation units, the (transmission) system operator directly intervenes in power plant decision-making, which seems to be more or less the opposite of what unbundling aims at (see Sect. 6.1). To illustrate the related costs of such a redispatch, a market comprising the region A and the region B is assumed in the following (cf. Nüßler 2012, pp. 12–18). In a one-price market, the grid is seen as a copper-plate (in other words, congestions are not considered when clearing the market), and therefore, market prices always have to be the same in both regions. Due to generation units with lower marginal costs, electricity will be exported from region A to region B (see diagram on the left in Fig. 10.3). If the resulting load flow exceeds the existing transmission capacity, the transmission system operator must adjust electricity generation in both regions (see diagram on the right in Fig. 10.3). Compared to the situation without congestion, generators located in region A (in the surplus region) have to decrease their output, generators in region B (in the deficit region) have to increase their production (in Fig. 10.3 by an amount equal to the distance between E and F). Savings partly compensate costs for increasing the output in the deficit region due to the reduction of production in the surplus region. Yet, the overall result is additional costs compared to a situation without congestions (see "Additional costs" in Fig. 10.3).

Whereas in this form of redispatch (based on costs) only the directly connected costs would be reimbursed, a so-called market-based redispatch would compensate the redispatch by prices determined on a competitive basis. But, as the contribution of a unit to relieve a congestion strongly depends on the location of this unit, such a market has to take place on a local level, opening up possibilities to act strategically

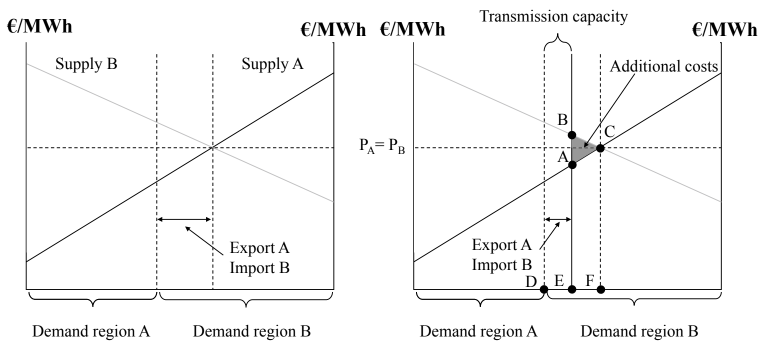


Fig. 10.3 Market prices without (left) and with congestions (right). *Source* Own illustration based on Nüßler (2012, p. 15)

(cf. Hirth et al. 2019). Nevertheless, a market-based redispatch may provide some incentives to build power plants at locations where they are needed from a grid perspective. However, these incentives are neither as strong nor as consistent as under a nodal market design. To have sufficient installations that can be used for redispatch in the deficit region, the system operator has to have enough capacity for upward regulation. This could lead to the introduction of an additional "redispatch reserve" (see, e.g., the so-called Grid Reserve in Germany), which could even been seen as a kind of a capacity mechanism (see Sect. 10.5).

Another possibility to relieve the congestion is **countertrading**, where the (transmission) system operator counter-trades against the direction of the congested flow (e.g. between two bidding zones) to reduce the flow over the line.

If the described grid-specific and market-related measures are not sufficient to guarantee the stability of the grid, the (transmission) system operator can adjust the feed-in and outtake as a final measure. Within this **feed-in management** scheme, the reduction of renewable generation is called "**curtailment**". Since installations based on renewable energy are mostly connected to the distribution grid, distribution grid operators (DSO) are affected mainly by renewable curtailment. As in many electricity systems, renewable energies enjoy priority access (see Sect. 6.2.4.2) to the grid, curtailment of renewables is seen as a last resort to relieve congestions. As electricity from renewable energies hardly causes any emissions and variable costs, the regulation often demands that system operators (using curtailment to relieve congestion) minimise the amount of curtailed renewable energy (cf. e.g. Schermeyer et al. 2018). However, it has to be emphasised here that curtailment may still be the least-cost option compared to alternatives such as grid expansion. Many studies seem to agree that some level of curtailment is economically advantageous (cf. e.g. Moser 2015; Schreiber et al. 2021, pp. 191–193), and at least in Germany, this is also foreseen according to the current regulation (cf. Bundestag 2022, §11).

Most Western European countries are still one-price zone, and price differences between zones (countries) reflect cross-border congestion. As congestions within countries (internal congestion) have increased during the last years, e.g. due to more electricity transport, (transmission) system operators had to intensify the practice of congestion management, mainly using redispatch and feed-in management to relieve the congestion. Using flexibilities provided, e.g. by demand-side applications (see Sect. 3.1.5), like cooling installations and cross-sector applications, like power-to-heat units, might help enhance congestion management in the future (cf. Chap. 12).

10.7 Retail Markets

The sales of electricity from electricity retailers to final customers are organised in retail markets. The range of end-use consumers and their electricity consumption is very heterogeneous, starting from small households via small businesses up to energy intensive industries, so a unique market with one price for all customers does not exist. In contrast to most spot markets, retail markets are based on bilateral contracts between electricity retailers and end-use consumers. The typical elements of retail contracts are discussed in Sect. 10.7.1. Section 10.7.2 is devoted to competition and prices in the retail market. The following two sections touch upon two topics that are attracting increasing attention in the context of the transition towards sustainable electricity systems: Sect. 10.7.3 addresses the issue of energy poverty, whereas Sect. 10.7.4 focuses on self-supply of customers based on rooftop PV or other distributed technologies.

10.7.1 Retail Contract Types

In general, retail contracts may be negotiated bilaterally (for large customers) or may be based on standardised offers by the suppliers (in the case of smaller clients). Yet even if negotiated bilaterally, contracts are primarily based on three kinds of price components as discussed earlier in the context of electricity tariffs (cf. Sect. 3.1.6):

- Prices for the connection per month or year, this is often called the service price or the base price for retail customers,
- Prices for the power or capacity measured in kilowatts (kW) or megawatts (MW),
- Prices for the electric work or electrical energy measured in kilowatt-hours (kWh) or megawatt-hours (MWh).

In most European countries, retail markets for households, small commercial customers as well as larger industries can be distinguished reflecting their different needs for power and energy:

- Larger and especially energy-intensive industries partly purchase their electricity directly on spot markets, ergo not on specific retail markets. However, this necessitates an exact forecast of hourly day-ahead electricity demand profiles. As a consequence, the hourly power demand of the following day has to be estimated by the customer (cf. Sect. 3.1.5) and deviations of this forecast are billed with the (possibly high) costs of balancing energy (see Sects. 8.2 and 8.4). An energy management system is necessary to handle this process and can only be operated economically if large amounts of (electrical) energy are purchased.
- In contrast to large energy consumers, households may consume electricity in relation to their needs without being priced for the capacity used. Of course, the power off-take is limited by the technical limits of the building connection, yet there is no direct tariffing of power for households. However, the service fee (or base fee of the contract) can be seen as a power price for being connected to the

given maximum power of the technical system.[20] In general, household consumers are charged for consumed electrical energy (in kWh) during a year (or a month) and the already mentioned additional monthly service fee (or base fee). Since European energy markets have been liberalised, household customers can select their energy contract from several energy retail companies. In general, web-based market platforms give an overview about available tariffs for different consumption levels (comparable to the communication market) and household customers select their energy retail companies according to various criteria (e.g. price per kWh, service charge, origin of power (e.g. green electricity), etc.).

- Contracts for larger commercial customers, including shops, services, etc., comprise prices for power and energy. These contract types are usually applicable above a certain consumption threshold, e.g. more than 100,000 kWh in Germany. Hence, companies have an incentive to reduce their peak power to avoid paying huge sums on the capacity price, and they also have an incentive to reduce energy consumption. Prices for capacity and energy depend for commercial customers on the utilisation rate of the power connection. Among two companies with the same yearly amount of consumed electricity, the company with higher power peaks (and thus higher capacity charges) and correspondingly a lower utilisation rate will generally be charged with higher costs. This kind of pricing is plausible from a technical and economic point of view: the higher peak power may necessitate a higher technical power input resulting in higher costs, which the customer must cover. From an economic point of view, the higher peak power offers more flexibility to the customer, which also justifies these higher prices.[21] Due to the various components of electricity prices, tariffs for larger customers may also depend on the grid level they are connected to (see Sect. 6.1.4 for principles of network pricing).

10.7.2 Competition on Retail Markets and Retail Prices

Competition in retails market is different from the wholesale competition as the primary action variable for retailers are the sales prices (cf. also Sect. 9.4). Retailers usually set their various above-mentioned price components, notably their base or service price and energy price (kWh) (and the capacity price for larger customers). Additionally, they may offer premiums for switching and provide specific products like green or local electricity. Thus, it is not a homogenous market with a single price as on wholesale markets. Another difference between retail and wholesale markets is the time granularity. Typical retail contracts are set up for a delivery

[20] In some European countries, contracts are offered for different power levels. In consequence, the base rate is higher for higher power peaks. In Germany, DIN 18015-1 “Planning of electrical systems in residential buildings” (DIN 2020) regulates the specifications of electrical house connections. The standard assumes a power requirement of 14.5 kW for a (standard) residential unit.

[21] Yet in principle coincident-peak and non-coincident-peak charges should be distinguished to reflect different cost drivers (cf. Sects. 3.1.6 and 6.1.4.4).

period of at least one year. In view of spot market procurement, the annual quantities have to be transformed into hourly quantities using load profiles. For pricing purposes, these load profiles are then combined with expected hourly price profiles, called **hourly price forward curves** (short: **HPFC**). These are discussed in detail in Sect. 11.2. Several authors have shown that the competitiveness of retail markets strongly depends on the switching rate of customers. This is also discussed along with further aspects of competition in retail markets in Sect. 9.4.

Exemplary per-unit costs of electricity are depicted in Fig. 10.4 for different types of customers in Germany. These significantly vary due to unequal price components. Large energy consumers are generally connected at high-voltage levels resulting in lower grid fees (as they only have to bear costs of the extra-high-voltage and high-voltage grid). For example in Germany, they are also often exempted from surcharges financing renewable energies (EEG-levy). In contrast, household customers are connected to low-voltage levels and thus have to carry the costs of all grid levels[22] and the full levy for renewable support. As depicted in Fig. 10.4, household prices are in the order of 30 €ct/kWh, while prices for energy-intensive industry (with privileged treatment[23]) are – with less than 6 €ct/kWh – by five times lower in Germany.

Comparing **electricity prices** in Europe, huge differences can be observed, with the highest **household prices** in Denmark and Germany reaching 30 €ct/kWh and the lowest in Lithuania and Bulgaria attaining approximately 12 and 10 €ct/kWh, respectively (cf. Fig. 10.4). The share of taxes and levies in household prices also varies enormously between the different European countries, being highest in Denmark with approximately 67% and lowest in Malta with only 5%. The differences in taxes and levies result from different value-added taxes, electricity taxes and levies for renewable energies. Besides absolute electricity prices, the relative share of electricity costs compared to the average (net) income is an indicator of how much of the income has to be spent on electricity. This share varies from 0.7% in Luxemburg and 1% in the Netherlands up to 2.5% in Latvia and 3% in Bulgaria. Prices for **industrial customers** are in all countries (except Malta) lower than household prices and range in a magnitude between highest 14 €ct/kWh in Malta and lowest 5.9 €ct/kWh in Finland for medium-sized customers with a consumption between 500 and 2,000 MWh.

[22] Of course, this statement depends on the regulation of power grids. In most European countries, costs are passed-on from higher voltage levels to lower voltage grids resulting in higher costs for consumers at low-voltage levels (cf. Sect. 6.1.4). This can be justified by the fact that the customers at the lower levels additionally make use of the (electricity transport) services at the higher levels.

[23] Depending on their electricity consumption, energy-intensive industries can apply for exemptions from non-energy related cost components of the electricity price such as the renewable levy.

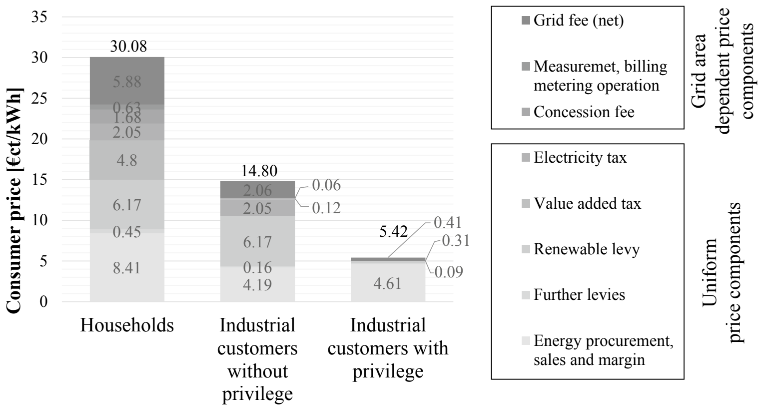


Fig. 10.4 Average composition of retail prices for electricity in Germany in 2015. *Source* Own illustration based on data from Bundesnetzagentur (2015)

10.7.3 Energy Poverty

An issue related to retail markets and retail electricity prices is **energy poverty** which describes the lack of access to modern energy services. In a global perspective, energy poverty raises serious and growing public health concerns related to indoor air pollution due to the use of polluting and less energy-dense fuels. Also, physical injury during fuelwood collection, and lack of refrigeration and medical care in areas that lack electricity are major issues (cf. Sovacool 2012). Energy poverty is primarily a severe problem in developing countries.

However, energy poverty can also be defined more broadly than the lack of access to modern energy: Bozarovski and Petrova (2015) formulate the following condition for energy poverty: "the inability to attain a socially and materially necessitated level of domestic energy services". With it, energy poverty not only refers to the situation of large numbers of people in developing countries but is also a major concern across the EU, where about 50 million or approximately 10% of the population say they are struggling to pay their energy bills.[24] High electricity or energy prices in relation to net income of households may result in energy poverty. In the last years, awareness of energy poverty has been rising in Europe and several EU institutions have identified it as a policy priority. For example, one third of the Bulgarian population is in arrears with their utility bills, according to the EU Energy Poverty Observatory, launched at the beginning of the year 2018. Energy poverty is considered an increasing social problem in the European Union, especially since the economic crisis in 2009 tended to worsen energy poverty in Europe (cf. Oliveras et al. 2021), continued by the corona crisis and the rise in energy prices in the wake of the Russia-Ukraine war.

[24] Cf. EU Energy Poverty Observatory: https://www.energypoverty.eu/.

In countries like Germany and France, the public debate touches related issues often under the somewhat broader term of "**affordability**" of energy and energy services. In the context of the transformation towards a sustainable energy system, this implies that both cost efficiency and distributional aspects of decarbonisation policies have to be taken into account (cf. Sect. 6.2.3).

10.7.4 Self-supply, Grid Parity and Level of Autonomy

With increasing household electricity prices and decreasing prices for PV systems, distributed electricity production from photovoltaics is cheaper than procuring electricity from the grid. Self-supply by small electricity generators at customers' locations (e.g. photovoltaics) substitutes electricity purchase in small and decentralised systems. In this context, the term **grid parity** describes the fact that an alternative energy source (e.g. photovoltaics) can generate power at levelised cost (LCOE) of electricity lower than the price of electricity purchased from a (grid-based) supplier. In this case, the LCOE is compared to the retail price of grid-delivered power, which includes not only generation costs but also further upstream cost components like grid fees, renewable levies, taxes, etc. (see Sect. 10.7.2). Retail prices are (much) higher than wholesale electricity prices and it is unclear which price shall be used as a benchmark for grid parity. As a consequence, different kinds of grid parity can be distinguished, depending on what is taken as a benchmark for retail prices:

- A **first phase** of grid parity is achieved when an alternative energy source can generate power at lower LCOE than the price of purchasing power, including taxes and levies on electricity prices. This grid parity was reached for utility-scale solar in 2011 and in 2012 for rooftop solar PV in Germany. In 2014, grid parity for solar PV systems was already reached in most European countries due to further decreasing LCOE of PV.
- A **second phase** of grid parity is attained when an alternative energy source can compete with the purchase price for electricity without taxes and levies. Hence, the costs of small and decentralised electricity production from an alternative source have to be lower than costs for production, transmission and distribution. Alternatively, some definitions of this second phase of grid parity take prices of industrial or commercial sectors as a benchmark. As shown in Fig. 10.5, electricity prices for medium-sized industries are roughly in the same order of magnitude as prices for households without taxes and levies in many countries. Due to rapid price decreases for PV modules, solar power generation has already reached this second phase of grid parity in a wide variety of locations or will reach it in the next few years.
- For the **third phase** of grid parity, different definitions exist. Japan's New Energy and Industrial Technology Development Organization (NEDO) defines the third phase of grid parity when an alternative energy source can compete with the cost of conventional power generation. This allows competition

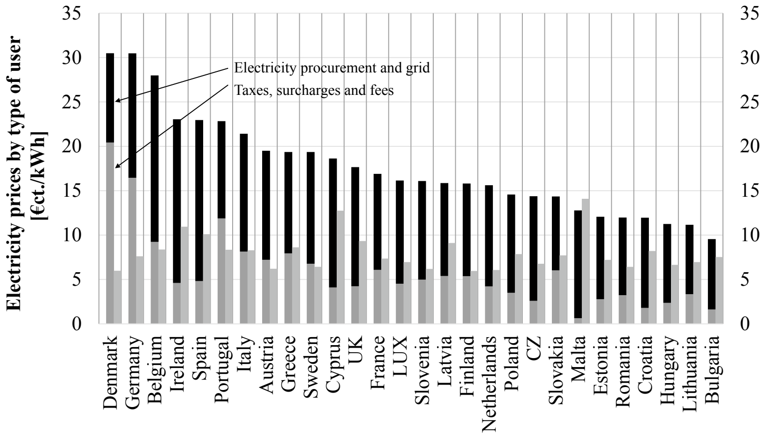


Fig. 10.5 Electricity prices in EU28 in 2017 for medium-sized households (left bars) and medium-sized industry (right bars) (Households with annual consumption between 2500 and 5000 kWh and industry with annual consumption between 500 and 2000 MWh). *Source* Own illustration based on data from Eurostat

between conventional and also large-scale alternative resources based on LCOE, but then the availability of the plants and the weather-dependability of the alternative resource is not considered. An alternative definition of this third phase of grid parity reflects more literally the term "grid parity": this definition refers to a competition between the purchase of electricity from the grid and a demand-driven, self-sufficient provision of electricity from the decentralised alternative system. In the above definitions of (first and second phases of) grid parity, it is always neglected that the generation from the alternative source depends on an external and not influenceable factor, the availability of the renewable source, which is in general dependent on the weather (e.g. solar radiation). A backup system is necessary to obtain electricity from an alternative source that matches the demand, which could be a battery system. Hence, the third phase of grid parity is sometimes also defined so that the decentralised alternative source, including the balancing system (e.g. PV, including sufficient battery capacity), competes with the purchase of electricity from the grid (without taxes and levies). According to this definition, the third phase of grid parity has not yet been reached. Competitiveness of decentralised systems strongly depends on the development of storage cost and future CO_2 prices.

The three phases of grid parity serve more as a rough guide when, at what level, and in which country alternative sources for electricity are becoming competitive. As the first phase of grid parity is reached in most (or even all) countries, the share of self-produced electricity is getting more and more a variable being optimised by the customers. This can be achieved by optimising the size of the PV field and the storage capacity at the point of consumption (with regard to the consumption level).

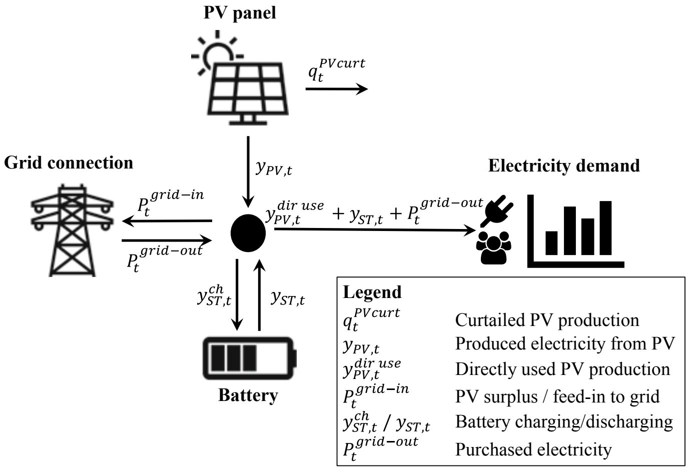


Fig. 10.6 Schematic illustration of PV electricity production and its usage in a household

To characterise this kind of decentralised electricity production, there are several indicators: The **level of power autonomy** describes how much of the electricity consumed locally is taken directly or physically from the installed photovoltaic system. This indicator is sometimes also referred to as the **rate of self-sufficiency** (cf. e.g. Dietrich and Weber 2018). The level of power autonomy considers the coincidence of production and demand and describes the share of PV electricity that is directly or indirectly consumed within the household: a PV electricity production $y_{\mathrm{PV},t}$ may be used directly in the household $y_{\mathrm{PV},t}^{\mathrm{dir\,use}}$ up to the current household load D_t.[25] The surplus is yet fed into the grid $P_t^{\mathrm{grid-in}}$ or used to charge a battery storage $y_{\mathrm{ST},t}^{\mathrm{ch}}$, if available (see also Fig. 10.6):

$$y_{\mathrm{PV},t} = y_{\mathrm{PV},t}^{\mathrm{dir\,use}} + y_{\mathrm{ST},t}^{\mathrm{ch}} + P_t^{\mathrm{grid-in}} \tag{10.1}$$

Conversely, the instantaneous household demand D_t may be met by the directly consumed PV production $y_{\mathrm{PV},t}^{\mathrm{dir\,use}}$, the discharging of stored electricity $y_{\mathrm{ST},t}$ and the outtake from the grid $P_t^{\mathrm{grid-out}}$:

$$D_t = y_{\mathrm{PV},t}^{\mathrm{dir\,use}} + y_{\mathrm{ST},t} + P_t^{\mathrm{grid-out}} \tag{10.2}$$

Accordingly, the **level of power autonomy** *LPA* is calculated by:

[25] Note that the notation used here is aligned as far as possible to the notation used in Sect. 4.4 and Chap. 7.

$$\mathrm{LPA} = \frac{\sum_{t=0}^{T}\left(y_{\mathrm{PV},t}^{\text{dir use}} + y_{\mathrm{ST},t}\right)\cdot\Delta t}{\sum_{t=0}^{T} D_t\cdot\Delta t} \tag{10.3}$$

A 25% level of power autonomy means that one-fourth of the electricity consumed comes from the photovoltaic system. Sometimes, the **level of energy autonomy** *LEA* is used as a further, more virtual indicator describing how much electricity consumed is on average provided by the photovoltaic system on a yearly basis. This indicator is based on annual values and neglects that power from the PV system may be available when there is no demand and that the PV surplus is fed into the power grid. The level of energy autonomy is per definition higher than the level of power autonomy and is calculated by:

$$\mathrm{LEA} = \frac{\sum_{t=0}^{T} y_{\mathrm{PV},t}\cdot\Delta t}{\sum_{t=0}^{T} D_t\cdot\Delta t} \tag{10.4}$$

According to the schematic illustration in Fig. 10.6 and as described in Eq. (10.1), the total PV production $y_{\mathrm{PV},t}$ used to calculate energy autonomy neglects battery losses and considers the surplus feed into the grid within the balance.

In contrast to the levels of power and energy autonomy, the **rate of self-consumption** *RSC* describes the share of self-consumed solar electricity in the total solar electricity produced. Thereby also the curtailed production $q_{\mathrm{PV},t}^{\mathrm{curt}}$ is considered in the denominator. The rate of self-consumption is calculated by:

$$\mathrm{RSC} = \frac{\sum_{t=0}^{T}\left(y_{\mathrm{PV},t}^{\text{dir use}} + y_{\mathrm{ST},t}\right)\cdot\Delta t}{\sum_{t=0}^{T}\left(y_{\mathrm{PV},t} + q_{\mathrm{PV},t}^{\mathrm{curt}}\right)\cdot\Delta t}. \tag{10.5}$$

A self-consumption rate of 50% means that half of the self-produced solar electricity is consumed directly on site.

Other things being equal, an increased size of the PV panel decreases the rate of self-consumption, as there are limits to the direct use of the produced electricity (see Eq. 10.2). The level of power autonomy yet increases for larger PV installations, given that the denominator in Eq. (10.3) remains constant. Through intelligent planning concerning the demand profile, the size and orientation of the PV field, and the size of the battery system, the self-consumption rate of and consequently the profitability of a photovoltaic system can be optimised.

A 100% self-consumption rate is technically achievable through sufficient storage capacity as long as the solar production remains substantially below the annual demand. Yet a 100% level of power autonomy is challenging to reach at reasonable cost given today's prices. A 100% level of self-consumption requires storing the surplus of PV production and consuming this surplus in times of low or no PV production. For a 100% level of power autonomy, shifting production from summer (when PV production is high) to winter (when energy demand is highest) is necessary. This would require a large decentral storage for shifting energy from one season to another which is, due to the size and the low full-load hours of storage utilisation, far beyond being economically attractive. Moreover, this could also be provided by sector coupling and power-to-heat with a seasonal thermal storage system. In contrast to a 100% level of power autonomy, a 100% level of energy autonomy could be easily achieved by dimensioning the solar field in a magnitude that the energy production is larger than the yearly energy consumption, neglecting time-dependencies of production and consumption.

Experiences from typical households with PV systems in central Europe show that self-consumption is in the order of magnitude of 30% without battery systems. However, self-consumption can strongly vary as it depends on the size, orientation and location of the plant and the size of the household. Thereby the level of self-consumption depends on the simultaneity of PV production and energy consumption. This simultaneity can be relaxed by installing a decentralised storage capacity. The economic attractiveness of such a storage system depends on the electricity costs compared to the PV LCOE and the storage costs. Hence, solar field size and battery size can be optimised concerning the own consumption profile. At today's battery prices and retail electricity prices of about 30 €ct/kWh in Germany as well as PV LCOE in the range of less than 10 €ct/kWh, the self-consumption rate in Germany can be increased up to 70% while still being close to a profitable investment in single-family houses (cf. Dietrich and Weber 2018).

The self-consumption of locally produced electricity may not be restricted to single-family houses but could also be extended to multi-dwelling buildings and larger areas or districts of homes ("Energy communities" in the wording of the EU legislation). Tenant electricity (also called district electricity) is locally produced electricity offered to private or commercial tenants. The model is equally suitable for condominium communities. The tenant electricity is produced in "immediate proximity" to the rental property and does not need to be routed through public networks. So grid charges and concession levies could be avoided making the model economically attractive for consumers. In Germany, security of supply is not affected by the purchase of tenant electricity. If additional power is needed, it can be obtained via the public grid. Conversely, surplus electricity from the tenant electricity system can be fed into the public grid under the terms of the subsidy scheme (renewable energy act). However, the success of such tenant electricity models strongly depends on the regulatory framework conditions in the different countries, especially whether additional taxes, levies and grid fees can be avoided and whether costs for the electricity grid have to be paid per kWh and not per connection point. Also, transaction costs are likely to be considerably higher for multi-stakeholder arrangements than for PV installations owned and operated by one household. Discussion is still ongoing, whether and how such models should be promoted, especially as the state loses revenues from taxes on the other side. And also, levies have to be borne by other customers. Furthermore, there is an ongoing discussion about completely off-grid energy communities, which have backup capacities instead of the public grid to ensure security of supply.

10.8 Markets in Europe Versus North America

Whereas electricity market design also comprises elements like the design of incentive mechanisms for greenhouse gas emission reduction and renewable energies (see Sects. 6.2.3 and 6.2.4), the subsequent considerations focus on the design of the core electricity market segments: short-term electricity markets, long-term electricity markets, markets for reserve energy (as the most important ancillary service) and congestion management. Table 10.2 summarises key design elements of these submarkets of the electricity market that are typical for most European and some North American electricity markets (for more details see, e.g., Baldick 2017; Chaves-Avila 2014; Ehrenmann 2018; Roques 2018, 2019; Ockenfels et al. 2008; Grimm et al. 2008); thereby it is worth mentioning that in both regions also electricity markets with quite different designs exist.

The European electricity markets are typically characterised by a high level of decentralisation with different exchanges, whereas competitive markets in the USA rely on a central and powerful institution – the so-called **Independent System Operator** (**ISO**). In European markets, players are free to trade the electricity bilaterally or send their bids to one of the power exchanges (like EPEX SPOT), which execute auctions and facilitate continuous trading. As these auctions represent multi-unit auctions (cf. Sect. 8.3) with uncertain demand, theoretical results obtained for single-unit auctions, like the revenue equivalence theorem (Vickrey 1961) cannot be transferred (cf. Stoft 2002, pp. 100–101). This market design is also called MinISO or exchange model or bilateral market, or auction market model. The bids of the market players are based on their strategies (notably obtained through **self-scheduling**, cf. Sect. 4.4, i.e. self-unit commitment and self-dispatch) and the market clearing process of the day-ahead market results in one (zonal) market price, e.g. for every hour of the following day for the respective market area. Besides bids for one specific hour of the following day (simple bid), day-ahead markets usually offer the opportunity to put more complex (non-convex) bids on the market. A prominent example of a complex bid is a so-called block bid (cf. Sect. 10.1.1). Such a block bid has to be entirely accepted or fully rejected (Fill-or-Kill criterion). It is not possible to accept the bid only for some of the hours.

The day-ahead markets are complemented by the possibility of trading electricity continuously in intraday markets to consider new information, which was not available day-ahead. Qualified market players can also decide to make bids to provide reserve capacity and reserve energy via different reserve markets (taking into account their opportunity costs), where the TSOs are single buyers (see Sect. 10.3). Clearing these energy and reserve markets typically results in zonal prices for the respective market area, mostly a country. To integrate European electricity markets, the markets for different countries are coupled (**Multi-Regional Coupling**) by exchanging information between European exchanges and using the available transmission capacities for electricity transport between countries for exports and imports to equalise differences in electricity market prices (see Sect. 7.2). If the available transfer capacity is sufficient, this will lead to precisely

Table 10.2 Elements of electricity market design and their specifications in typical European and North American electricity markets[a]

Elements	Subelements	Europe	North American ISO markets
Long-term markets	Electricity trading	Forward and future markets	Forward and future markets
	Capacity trading	Partly capacity mechanisms	Mostly capacity mechanisms
Short-term markets	Day-ahead market	– Auction for hours of the following day – Bids of market players, in case of generation and storage portfolios based on self-scheduling – Paradoxically rejected bids possible – In principle physical fulfilment	– Central security-constrained scheduling for hours of the following day by ISO – Side payments – Financial settlement
	Intraday/real-time market	– Continuous trading – Bids of market players, e.g. based on self-dispatch considering own portfolio	Central security-constrained dispatch for the next 5 min (base point for generators)
	Forms of bids	– Physical bids – Block bids as complex bids	– Complex multi-part bids (techno-economic parameters of generators) – Physical and virtual bids
	Spatial dimension	Zonal (often countries)	Nodal (Locational marginal prices)
Balancing aspects	Responsibility	Balancing responsible parties	Independent system operator
Reserve markets	Integration of energy and reserve markets	– Separate markets – Opportunity cost bidding – TSOs are single buyers	Co-optimisation of energy and reserve markets (day-ahead and real time)
Network aspects	Congestion management	– Interzonal constraints considered in the market clearing process – Intrazonal congestions mostly to be solved by TSO (e.g. by redispatch)	Integration of network constraints into short-term market clearing (day-ahead and real time)

[a] Despite the general classification, there are some exemptions, e.g. Poland has a central dispatch with a nodal clearing. While capacity trading is common in USA, Texas has an EOM with an operating reserve demand curve in place.
Sources Baldick (2017), Chaves-Avila (2014), Ehrenmann (2018), Roques (2018), Roques (2019), Ockenfels et al. (2008) and Grimm et al. (2008)

the same market price in the two countries. Grid congestions between countries are considered in the market process with the help of flow-based market coupling, whereas internal congestions can hardly be handled by the market coupling. After the market clearing process, the TSOs responsible for network operation must make load flow calculations and apply different measures for congestion management (e.g. **redispatch**) in the case of congestions (see Sect. 10.6).

In contrast to the European electricity market design, offers[26] to buy or sell electricity and capacity in North American markets like PJM (Pennsylvania, New Jersey, Maryland Interconnection) have to be given to a **central pool**. The responsible ISO then clears the whole market centrally. This system design is also called MaxISO or the pool model. As the ISO is also in charge of operating the transmission grid, although not necessarily being the owner of the grid, the ISO as a central institution has all information needed to calculate the optimal solutions for the energy and reserve markets simultaneously, considering electricity grid constraints (so-called co-optimisation; see, e.g., Papavasiliou 2016). Therefore, ISOs can execute security-constrained scheduling calculations day-ahead and security-constrained dispatch calculations nearly in real-time (e.g. for every five minutes) to determine locational marginal prices (LMP) for every node in their system. The results of the security-constrained dispatch calculations show the level of operation for all units in the system at five-minute intervals, which means that reserve capacity is only needed for intra-five-minute variability (Baldick 2017). Financial transmission rights (FTR) may be used under this market design to hedge against different marginal prices in different locations (see Sect. 10.6). To conduct security-constrained unit commitment and dispatch calculations, ISOs need a huge amount of (truthful) information, e.g. about the costs and production constraints of the production, storage and consumption units (multi-part bids) and of the transmission assets. The optimal results of the central calculation can lead to situations, where, e.g., a generation unit is “in the market”, although the nodal prices at the location of this installation are not sufficient to cover the production costs, leading to the necessity of compensation payments, so-called side payments.

The two market designs emphasise different aspects of competitive markets. The European framework has been fostering on the freedom of market players to enter into contracts and to dispose individually of their assets, limiting the role of the monopolistic grid operators. In the USA, the emphasis has rather been on getting the incentives for the market players right, especially since nodal pricing internalises grid constraints and thus does not incentivise power producers (and consumers) to externalise costs on the grid (see also Wilson 2002).

In view of the transition towards a sustainable electricity system, both aspects are highly important. Yet for Europe, the major issue is whether the intertwined European and national legislators are willing to engage into the major institutional overhaul necessary to establish locational marginal pricing. According to the European treaties, the European Commission has an important role to play to enable trade and competitive markets across Europe. Yet its mandate is limited when it comes to imposing new international institutions.

[26] In some day-ahead markets also virtual bids are allowed. This means that the virtual day-ahead bid to sell or buy electricity has to be offset by a corresponding bid to buy or sell the electricity back on the real-time market.

10.9 Further Reading

Meeus, L. (2020). The Evolution of Electricity Markets in Europe. Cheltenham: Edward Elgar.

This book describes how Europe has experienced the evolution of modern electricity markets since liberalisation in the mid-90s. The author explains the sequence of electricity markets in Europe from wholesale to balancing markets. He also discusses forward transmission markets, capacity mechanisms, redispatching and flexibility markets.

EC. European electricity market reports. https://data.europa.eu/euodp/en/data/dataset/european-electricity-market-reports.

The European Commission publishes quarterly reports on European gas and electricity markets and energy prices in Europe. These quarterly reports give a good overview of recent developments of the energy markets and analyse main developments on the markets and interactions between countries.

10.10 Self-check of Knowledge and Exercises

Self-check of Knowledge

1. Explain how a unique market price within one hour is determined.
2. Name at least three day-ahead markets for electricity (electricity exchanges) in Europe.
3. What are block bids?
4. Explain the main difference between day-ahead and intraday markets, primarily focusing on the aspect of continuous trading.
5. How is cross-border trading organised? Explain the difference between explicit and implicit trading of cross-border capacities.
6. Name and explain at least two system services.
7. Name and explain at least three different types of capacity mechanisms. What is the purpose of capacity mechanisms?
8. What is redispatch?
9. Explain the difference between ex-ante and ex-post congestion management.
10. Explain how market prices in two regions are determined with the help of the standard supply and demand model in two cases: with and without congestion.
11. What are retail electricity markets and which general price components can be distinguished on these markets?
12. Explain the average composition of retail prices at the example of Germany for different customers.
13. What is the difference between energy and power autonomy? What autonomy is easier to be achieved?

Exercise 10.1: Power Markets

1. In the context of electricity markets, characterise futures markets according to the purpose the markets serve and classify them according to the time between the commercial transaction and the physical delivery of the service.
2. Name elements of electricity market design and their specifications in typical European and North American electricity markets.
3. Can generators cover their fixed costs if the market price always equals the (physical) marginal costs of the price-setting technology? Distinguish in your answer base, medium and peak-load technologies.
4. Discuss the role of capacity markets in this context. How do capacity markets impact the market price?

Exercise 10.2: Market Clearing and Redispatch

In a European country (one market area), the following generation capacities are available:

- 3 CCGT with a capacity of 300 MW_{el} each, located in the northern part of the country,
- 3 OCGT with a capacity of 100 MW_{el} each, located in the southern part of the country,
- 2 coal power plants with a capacity of 800 MW_{el} each, located in the northern part of the country,
- 1 coal power plant with a capacity of 800 MW_{el}, located in the southern part of the country and
- wind power plants with a capacity of 400 MW_{el}, located in the northern part of the country.

For the techno-economic characterisation of the technologies, please refer to Chap. 4. The fuel costs for coal are assumed to be 12 €/MWh_{th}, the fuel costs for natural gas 45 €/MWh_{th} and the CO_2 price is 30 €/t.

(a) Calculate the clearing price of the day-ahead market assuming perfect competition for an hour with wind feed-in of 200 MW. The demand in this hour is estimated to be 2900 MW, equally distributed between the different parts of the country.
(b) Considering the results of (a), the TSO identifies a congestion between the northern and the southern part of the country in his day-ahead congestion forecast. So, the TSO decides to demand a redispatch in the amount of 300 MW_{el} to avoid this congestion. Which additional costs will arise?

Exercise 10.3: Self-supply and Level of Energy Autonomy of a PV System
A PV system with a peak capacity of P_p = 4 kW will be installed on a building with an annual electricity demand of 4000 kWh. The PV modules can be oriented to the south or split to an east and west orientation.

The following data are given:

Assumptions	2010	2017
Capacity [kW]	4	4
Full-load hours South [h/a]	976.25	976.25
Full-load hours East–West [h/a]	828.75	828.75
Grid feed-in[a] South [kWh/a]	*	2600
Grid feed-in East–West [kWh/a]	*	1500
Specific investment PV [€/kW]	3500	1500
Interest rate [%]	3	0.1
Subsidy tariff [€/kWh]	0.41126	0.12
Electricity price [€/kWh]	0.23	0.3
Load [kWh]	4000	4000
Lifetime [a]	20	20

[a] Grid feed-in = Produced electricity from PV − self-consumption

(a) Determine the share of self-consumption and the level of energy autonomy for the systems in 2017. What is the influence of the orientation on the level of self-consumption?
(b) Calculate the electricity price at which an east–west orientation is preferable compared to a south exposure in 2017?
(c) The south-oriented PV system shall be expanded to include a battery storage system with a storage capacity of 4 kWh. The electricity consumption from the grid decreases to 1.361.78 kWh/a. Calculate the share of self-consumption and the level of energy autonomy.

References

ACER. (2013). *Capacity remuneration mechanisms and the internal market for electricity*. European Union Agency for the Cooperation of Energy Regulators (ACER).

Androcec, I., & Wangensteen, I. (2006). Different methods for congestion management and risk management. In *9th International Conference on Probabilistic Methods Applied to Power Systems*. Available at: https://www.researchgate.net/publication/224704006_Different_Methods_for_Congestion_Management_and_Risk_Management [Accessed August 19, 2022].

Baldick, R. (2017). *Seams, ancillary services and congestion management: US versus EU electricity markets*. Available at: https://medium.com/florence-school-of-regulation/seams-ancillary-services-and-congestion-management-us-versus-eu-electricity-markets-ab9ad53a16df [Accessed December 21, 2018].

Bhagwat, P., Marcheselli, A., Richstein, J., Chappin, E., & De Vries, L. (2017). An analysis of a forward capacity market with long-term contracts. *Energy Policy, 111*, 255–267.

Bozarovski, S., & Petrova, S. (2015). A global perspective on domestic energy deprivation: Overcoming the energy poverty–fuel poverty binary. *Energy Research & Social Science, 10*, 31–40.

Bublitz, A., Fraunholz, C., Zimmermann, F., & Keles, D. (2016). *Overview on market design options for European electricity markets*. Reflex report. Available at: http://reflex-project.eu/wp-content/uploads/2017/11/D5.1_Overview_on_market_design_options_update.pdf [Accessed May 26, 2022].

Bublitz, A., et al., (2019). A survey on electricity market design: Insights from theory and real-world implementations of capacity remuneration mechanisms. *Energy Economics, 80*, 1059–1107.

Bundesnetzagentur. (2015). *Monitoringbericht 2015*. Report, Bonn. Available at: http://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Bundesnetzagentur/Publikationen/Berichte/2015/Monitoringbericht_2015_BA.pdf [Accessed May 26, 2022].

Bundestag. (2022). *Gesetz über die Elektrizitäts- und Gasversorgung (Energiewirtschaftsgesetz vom 7. Juli 2005) zuletzt geändert durch Art. 2 Gesetz v. 23.5.2022*. Deutsche Bundesgesetze und -verordnungen.

Burke, B., Merrill, H., Schweppe, F., Lovell, B., McCoy M., & Monohon, S. (1988). Trade off methods in system planning. *IEEE Transactions on Power Systems, 3*(3), 1284–1290. https://doi.org/10.1109/59.14593

Chaves-Avila, J. P. (2014). *European short-term electricity market designs under high penetration of wind power*. Dissertation: TU Delft. Available at: https://doi.org/10.4233/uuid:cbc2c7bd-d65e-4f4c-9f88-2ab55c11def8 [Accessed May 26, 2022].

Cramton, P., & Ockenfels, A. (2012). Economics and design of capacity markets for the power sector. *Zeitschrift für Energiewirtschaft, 36*, 113–134.

Cramton, P. C., & Stoft, S. (2006). *The convergence of market designs for adequate generating capacity with special attention to the CAISO's resource adequacy problem*. A White Paper for the Electricity Oversight Board. Available at: https://hepg.hks.harvard.edu/files/hepg/files/cramton_stoft_0406.pdf?m=1523368470 [Accessed May 26, 2022].

de Vries, L. (2004). Policy framework for the stabilization of investment in generating capacity. In *Proceedings of the 19th World Energy Congress*, London, UK.

de Vries, L. J., & Hakvoort, R. A. (2004). *The question of generation adequacy in liberalised electricity markets*. Available at SSRN: https://ssrn.com/abstract=600503 [Accessed May 26, 2022].

Dietrich, A., & Weber, C. (2018). What drives profitability of grid-connected residential PV storage systems? A closer look with focus on Germany. *Energy Economics, 74*, 399–416.

DIN. (2020). *DIN 18015-1: 2020-05. Electrical installations in residential buildings—Part 1: Planning principles Beuth*.

Duthaler, C., Emery, M., Andersson, G., & Kurzidem, M. (2008). Analysis of the use of PTDF in the UCTE transmission grid. In *Power System Computation Conference*, Glasgow. Available at: https://core.ac.uk/download/pdf/147964029.pdf [Accessed August 19, 2022].

EC. (n.d.). *European electricity market reports*. Available at https://data.europa.eu/euodp/en/data/dataset/european-electricity-market-reports [Accessed: June 13, 2022].

EC. (2016). *Commission staff working document accompanying the Report from the Commission: Final Report of the Sector Inquiry on Capacity Mechanisms SWD/2016/0385 final*. Available at https://eur-lex.europa.eu/legal-content/DA/TXT/?uri=CELEX:52016SC0385 [Accessed August 19, 2022].

EC. (2017a). *Commission Regulation (EU) 2017/1485 establishing a guideline on electricity transmission system operation*. European Commission. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R1485 [Accessed May 26, 2022].

EC. (2017b). *Commission Regulation (EU) 2017/2195 establishing a guideline on electricity balancing*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2017.312.01.0006.01.ENG&toc=OJ:L:2017:312:TOC [Accessed January 11, 2021].

ECA. (2015). *European electricity forwards and hedging products—State of play and elements for monitoring*. Final Report to Acer. Economic Consulting Associates Limited. Available at:

https://www.acer.europa.eu/en/Electricity/Market%20monitoring/Documents_Public/ECA%20Report%20on%20European%20Electricity%20Forward%20Markets.pdf [Accessed May 26, 2022].

Ehrenmann, A. (2018). Response Electricity market design and the green agenda. In *Transforming Energy Markets 41st IAEE International Conference, Groningen*.

Grande, O., Doormann, G., & Wangensteen, I. (2001). Peaking capacity in restructured power systems: experience and solutions. In *2001 IEEE Porto Power Tech Proceedings* (Cat. No. 01EX502).

Grimm, V., Ockenfels, A., & Zoettl, G. (2008). Strommarktdesign: Zur Ausgestaltung der Auktionsregeln an der EEX. *Zeitschrift für Energiewirtschaft, 32*, 147–161. https://doi.org/10.1007/s12398-008-0020-7

Hedman, K. W., Oren, S. S., & O'Neill, R. P. (2011). Optimal transmission switching: Economic efficiency and market implications. *Journal of Regulatory Economics, 40*, 111–140. https://doi.org/10.1007/s11149-011-9158-z

Hinz, F. (2017). *Voltage stability and reactive power provision in a decentralizing energy system* (Vol. 12). Series of the Chair of Energy Economics and Dissertation: Technische Universität Dresden. Available at: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-229585 [Accessed May 26, 2022].

Hirth, L., Schlecht, I., Maurer, C., & Tersteegen, B. (2019). *Kosten- oder Marktbasiert? Zukünftige Redispatch-Beschaffung in Deutschland*. Neon & Consentec (im Auftrag des BMWi). Available at: https://www.bmwk.de/Redaktion/DE/Publikationen/Studien/zukuenftige-redispatch-beschaffung-in-deutschland.html [Accessed May 26, 2022].

Hoffrichter, A., Kollenda, K., Schneider, M., & Puffer, R. (2019). Simulation of curative congestion management in large-scale transmission grids. In *Bucharest: 2019 54th International Universities Power Engineering Conference (UPEC)*. Available at: https://doi.org/10.1109/UPEC.2019.8893627 [Accessed May 26, 2022].

Hogan, W. (2005). *On an 'Energy only' electricity market design for resource adequacy*. Center for Business and Government, John F. Kennedy School of Government, Harvard University, Cambridge (MA). Available at: https://scholar.harvard.edu/whogan/files/hogan_energy_only_092305.pdf [Accessed May 26, 2022].

Höschle, H. (2018). *Capacity mechanisms in future electricity markets*. Dissertation: KU Leuven. Available at: https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1731489&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&fromSitemap=1 [Accessed August 18, 2022].

Joskow, P., & Tirole, J. (2007). Reliability and competitive electricity markets. *The RAND Journal of Economics, 38*, 60–84.

Joskow, P. L. (2008). Lessons learned from the electricity market liberalization. *The Energy Journal, 29*, 9–42.

Just, S. (2011). Appropriate contract durations in the German markets for on-line reserve capacity. *Journal of Regulatory Economics, 39*, 194–220.

Kunz, F. (2013). Improving congestion management: How to facilitate the integration of renewable generation in Germany. *The Energy Journal, 34*, 55–78.

Kunz, F., Neuhoff, K., & Rosellón, J. (2016). FTR allocations to ease transition to nodal pricing: An application to the German power system. *Energy Economics, 60*, 176–185.

Madani, & Van Vyve. (2014). Minimizing opportunity costs of paradoxically rejected block orders in European day-ahead electricity markets. In *11th International Conference on the European Energy Market (EEM14)*. Available at: https://doi.org/10.1109/EEM.2014.6861237 [Accessed May 26, 2022].

Meeus, L. (2020). *The evolution of electricity markets in Europe*. Edward Elgar.

MIGRATE. (2018). *Massive integration of power electronic devices*. Project funded under the EU Horizon 2020 scheme. Available at: https://www.h2020-migrate.eu/_Resources/Persistent/b955edde3162c8c5bf6696a9a936ad06e3b485db/19109_MIGRATE-Broschuere_DIN-A4_Doppelseiten_V8_online.pdf [Accessed May 26, 2022].

Moser, A. (2015). *Systemstudie zum Einspeisemanagement erneuerbarer Energien*. Wissenschaftliche Studie im Auftrag der EWE Aktiengesellschaft, RWTH Aachen.

Neuhoff, K., & Boyd, R. (2011). *International experiences of nodal pricing implementation*. Climate Policy Initiative. Available at: https://www.climatepolicyinitiative.org/wp-content/uploads/2011/12/Nodal-Pricing-Implementation-QA-Paper.pdf [Accessed May 26, 2022].

Neuhoff, K., & de Vries, L. (2004). *Insufficient incentives for investment in electricity generation*. Cambridge Working Papers in Economics CWPE 0428. Available at: https://core.ac.uk/download/pdf/142112.pdf [Accessed May 26, 2022].

Newbery, D. (2002). Regulatory challenges to European electricity liberalisation. *Swedish Economic Policy Review, 9*, 9–43.

Nüßler, A. (2012). *Congestion and redispatch in Germany. A model-based analysis of the development of redispatch*. Dissertation: University of Cologne.

Oliveras, L., Peralta, A., Palència, L., Gotsens, M., López, M., Artazcoz, L., Borrell, C., & Marí-Dell'Olmo, M. (2021). Energy poverty and health: Trends in the European Union before and during the economic crisis, 2007–2016. *Health & Place, 67*, 102294.

Oren, S. (2000). Capacity payments and supply adequacy in competitive electricity markets. In *VII Symposium of Specialists in Electric Operational and Expansion Planning*, Curitiba. Available at: https://oren.ieor.berkeley.edu/workingp/sepope.pdf [Accessed May 27, 2022].

Oren, S. (2003). *Ensuring generation adequacy in competitive electricity markets*. Energy Policy and Economics, University of California Energy Institute, UC Berkeley. Available at: https://escholarship.org/uc/item/8tq6z6t0 [Accessed June 18, 2022].

Papavasiliou, A. (2016). Co-optimization of energy and reserves. In FPM Meeting, Paris. Available at: https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.545499.de/fpm-paris-2016-10-04-papavasiliou.pdf [Accessed May 27, 2022].

Pérez-Arriaga, I. J. (1999). Reliability and generation adequacy. *IEEE Power Engineering Review, 19*, 4–5.

Ringler, P. (2016). *Erzeugungssicherheit und Wohlfahrt in gekoppelten Elektrizitätsmärkten*. Dissertation: Karlsruhe Institute of Technology (KIT). Available at: https://publikationen.bibliothek.kit.edu/1000064573/4015819 [Accessed May 26, 2022].

Roques, F. (2018). *Market design for electricity—Comparing the US and EU*. Available at: https://www.youtube.com/watch?v=uD8j1Mmc3Ws [Accessed September 16, 2021].

Roques, F. (2019). *Wholesale power market design—Key issues and principles for an efficient decarbonisation*. In IX Electricity Security Advisory Panel Workshop, Paris. Available at: https://iea.blob.core.windows.net/assets/imports/events/240/S3_4_Roques.pdf [Accessed May 27, 2022].

Schermeyer, H., Vergara, C., & Fichtner, W. (2018). Renewable energy curtailment: A case study on today's and tomorrow's congestion management. *Energy Policy, 112*, 427–436.

Schreiber, S., Zöphel, C., & Möst, D. (2021). Optimal energy portfolios in the electricity sector: Trade-offs and interplay between different flexibility options. In D. Möst, et al. (Eds.), *The future European energy system—Renewable energy, flexibility options and technological progress*, pp. 177–198. Springer.

Sovacool, B. (2012). The political economy of energy poverty: A review of key challenges. *Energy for Sustainable Development, 16*, 272–282.

Stoft, S. (2002). *Power system economics—Desigining markets for electricity*. Wiley.

Swider, D. J., & Weber, C. (2007). Bidding under price uncertainty in multi-unit pay-as-bid procurement auctions for power systems reserve. *European Journal of Operational Research, 181*, 1297–1308.

Vickrey, W. (1961). Counterspeculation, auctions, and ompetitive sealed tenders. *The Journal of Finance, 16*, 8–37.

Wilson, R. (2002). Architecture of power markets. *Econometrica, 70*, 1299–1340.

Valuing Flexibilities in Power Systems as Optionalities

11

The concepts of flexibilities and optionalities in electricity systems have become increasingly popular over the last two decades. There are two major but distinct drivers for this development: the first one is related to the financial trading of electricity products on future and other derivative markets. In that context, it has apparent merits to consider flexibilities in physical assets, like power plants analogously to financial contracts with embedded flexibilities. The latter are named options, and hence, it has become popular to consider power plants, storages and other assets as "real options".

The other driver is increasing shares of fluctuating renewables that are expected to dominate in the future sustainable energy systems. Here, a lack of flexibilities is perceived as a potential challenge: increasing shares of renewables imply, other things being equal, higher uncertainties due to growing forecast errors. And at the same time, they go along with decreasing shares of controllable conventional power plants.

The two perspectives on flexibilities have somewhat different starting points, and dealing with them in a common framework is not an easy exercise. The most striking difference is that the real options perspective takes prices as exogenous to the decision-maker. In contrast, the second perspective takes a system view, where prices are necessarily a result of interactions between system elements – as in the fundamental equilibrium models of Sect. 7.1. A complete synthesis of these two perspectives is beyond the scope of this textbook. Yet, some elements are put forward after a concise introduction to the financial perspective on flexibilities, i.e. real options. We start thereby by modelling prices as stochastic processes (cf. Sect. 11.1). Then, we introduce the concept of the hourly price forward curve to link future and spot prices in electricity markets in Sect. 11.2. Section 11.3 uses these concepts to value simple options, followed by a digression to financial options and the seminal Black–Scholes model in Sect. 11.4. Section 11.5 discusses the

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-030-97770-2_11.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_11

merits and limits of the Black–Scholes model for electricity market modelling, whereas Sect. 11.6 describes an approach to model thermal and hydropower plants as options in view of valuation. Section 11.7 then applies this approach, and Sect. 11.8 finally comes back to how to bridge the gap between the asset valuation and the system perspective.

Key Learning Objectives
After having gone through this chapter, you will be able to

- Describe and apply key stochastic processes that are used to model price changes in energy and other markets.
- Explain the concept of the hourly price forward curve and how it is used to price electricity supply contracts.
- Discuss key concepts underlying the valuation of options using methods from mathematical finance.
- Discuss the concept of real option and apply a simple valuation model for a thermal power plant.

11.1 Prices as Stochastic Processes

For financial assets like stocks, the price reflects the value attributed to that asset in the market. This price may change over time. E.g. if a company announces unexpected losses, the price of its shares on the stock exchange will go down. Mathematically, the price of an asset may then be described as a stochastic process, i.e. a sequence of realisations of a stochastic variable. One may wonder: why is the price considered a **stochastic process**? This is closely related to the efficient market hypothesis (see Sect. 7.2.5). If a market is efficient, it uses all available information at time t (the information set Ω_t) to determine the asset price. Any new information arriving after time t may change the price. But it would not be **new** information if it did not come as a surprise, i.e. randomly, from the perspective of time t. Put differently: with hindsight (ex-post), we may pretend that we knew before, but ex-ante, we as rational decision-makers will include all available information (even vague expectations, etc.) in our decisions and valuations.

To describe stochastic processes in general, it is helpful to start with a straightforward process that may serve as the basis for multiple generalisations, namely the **Wiener process**. The Wiener process may be best understood as a kind of a **random walk** in continuous time. A random walk consists of a sequence of steps Δz_k that are taken during subsequent time intervals k of length Δt and are randomly and independently chosen. Additionally, we impose for the mathematical description that the steps correspond to stochastic variables ε that are normally distributed with zero mean and standard deviation proportional to $\sqrt{\Delta t}$. This leads to the following mathematical description:

$$\Delta z_k = \varepsilon \quad \varepsilon \sim N(0, \sqrt{\Delta t}). \tag{11.1}$$

Applying standard rules of calculus for normally distributed random variables, it can be shown that for any time interval $T = K \cdot \Delta t$ (i.e. composed of K time steps Δt), the following relationship holds

$$z_{t+T} - z_t = \sum_{k=1}^{K} \Delta z_k \sim N(0, \sqrt{K \cdot \Delta t}). \tag{11.2}$$

That means that for a time interval of arbitrary length T, the change in the **stochastic process** variable z_k is still normally distributed with mean zero and standard deviation $\sqrt{T}$. The process is hence "self-similar", independently of the time granularity considered.

This property may then be generalised to infinitesimal time steps dz, leading to the formulation:

$$\mathrm{d}z = \lim_{\Delta t \to 0} \Delta z_k. \tag{11.3}$$

The so defined dz is then the (infinitesimally small) increment of a Wiener process $z(t)$ and using a somewhat loose mathematical notation, we may write $\mathrm{d}z \sim N\left(0, \sqrt{\mathrm{d}t}\right)$. Besides being normally distributed, the increments dz are independent of each other, again irrespective of the time scale considered. The stochastic process variable $z(t)$ itself is then described as a stochastic integral of the increments

$$z(t) - z(0) = \int_0^t \mathrm{d}z. \tag{11.4}$$

One application of this Wiener process in physics is the description of the random movement of particles in a (non-flowing) gas or liquid. This movement was first observed by Scottish nineteenth century scientist Robert Brown and is also known as Brownian motion.

The self-similarity of the Brownian motion becomes apparent in Fig. 11.1, where one single realisation of the Brownian motion is depicted at different discretisation levels. The highest discretisation in the top panel includes 2000 time steps of equal length, whereas the middle panel highlights 100 discrete steps. And the bottom panel is further zoomed out with just 5 discrete steps over the same overall time period. Yet, at each discretisation level, the process includes random steps upwards and downwards of different size. Another property that is also visible is the absence of any mean-reverting effect. The observed realisation of the random process moves

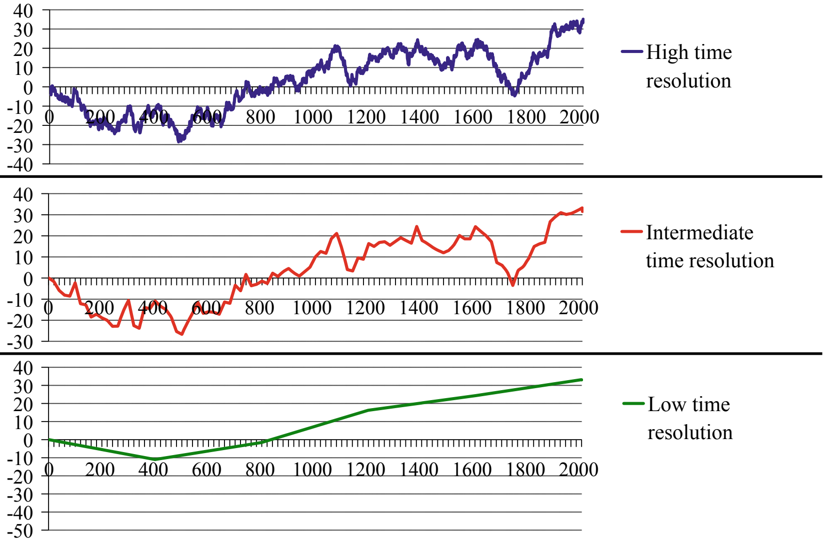


Fig. 11.1 One realisation of a Wiener process observed at different scales of time discretisation

away from the starting value of 0. Independently of the level attained, the probability of going up or down the next step remains unchanged. This property is related closely to the fact that the resulting time series is "non-stationary". We will come back to this point after introducing some generalisations.

A straightforward generalisation of the Wiener process is to introduce a drift – in physics, the equivalent would be an (average) flow direction – and a scaling of the stochastic component so that it may be of arbitrary variance. This leads to the following definition of a **generalised Wiener process** dx:

$$\mathrm{d}x = a \cdot \mathrm{d}t + b \cdot \mathrm{d}z. \tag{11.5}$$

The (positive or negative) parameter a is called the drift rate – it is an average rate of change of x over time. The positive parameter b is named variance rate – although it rather scales the standard deviation (i.e. the square root of the variance) of the stochastic process x.

The impact of the drift rate becomes obvious in Fig. 11.2. With a positive drift rate, the stochastic process moves on average upwards – although this does not preclude that certain increments are negative. As indicated by Eq. (11.5), the overall change is the sum of the deterministic drift part (first term) and the stochastic process part (second term) and the sign depends on the sign and magnitude of the stochastic realisation.

Suppose a price process is expected to oscillate around some average value. In that case, an alternative specification is required for a stochastic process since neither the Wiener process nor a fortiori its generalisation tend to return to some

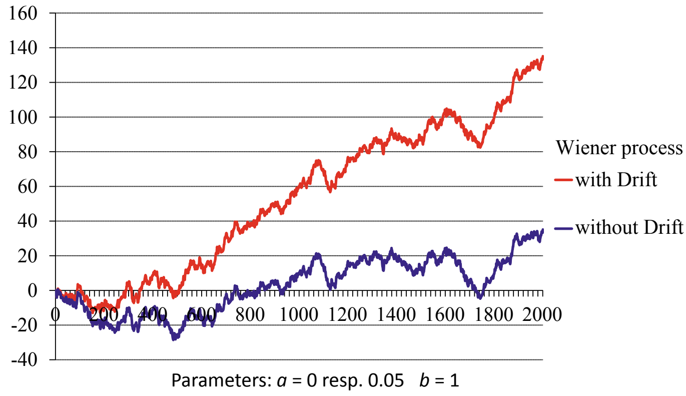


Fig. 11.2 Realisations of a Wiener process with and without drift with identical stochastic components

prespecified mean value. Equation (11.6) specifies a so-called mean-reversion process, also called **Ornstein–Uhlenbeck process**:

$$\mathrm{d}x = \kappa \cdot (\mu - x) \cdot \mathrm{d}t + \sigma \cdot \mathrm{d}z. \tag{11.6}$$

The stochastic second term consists again of a Wiener process multiplied by a standard deviation parameter σ. So the difference lies in the deterministic first term, which includes the factor $(\mu - x)$, which is positive when x is smaller than μ and negative in the opposite case. With a positive factor κ (called mean-reversion rate), this induces a tendency for x to return to the mean value μ. The higher the mean-reversion rate κ, the faster the return to the equilibrium value μ – similar to the pull-back force of a mechanical spring. Yet again, we have a stochastic component superposed on this mean-reversion component, and thus, the resulting incremental changes may go in both directions as illustrated in Fig. 11.3.

As a last relatively simple stochastic process, we introduce the so-called **geometric Brownian motion** (or GBM for short). It is notably used in standard finance models to describe the movement of stock prices $S(t)$. The increments dS of this stochastic process are described by the following stochastic differential equation:

$$\mathrm{d}S = \mu \cdot S \cdot \mathrm{d}t + \sigma \cdot S \cdot \mathrm{d}z. \tag{11.7}$$

Besides the use of different symbols both for the stochastic process variable and the parameters, there are two salient differences of this equation compared to the one describing the generalised Wiener process (Eq. 11.5), namely the multipliers "$\cdot S$" in both the deterministic first term and the stochastic second term. Rewriting the previous equation slightly, we get the following formulation:

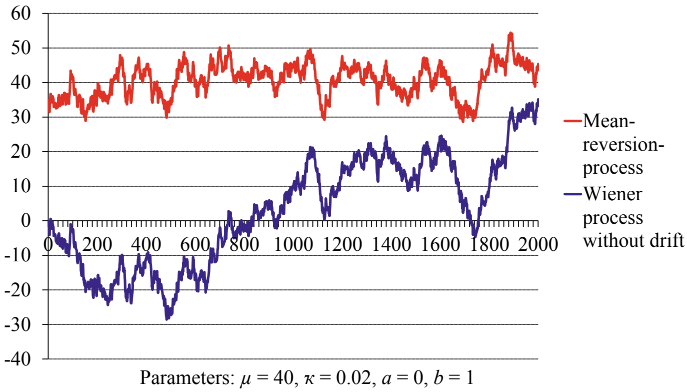


Fig. 11.3 Realisations of a mean-reversion process and a Wiener process with identical stochastic components

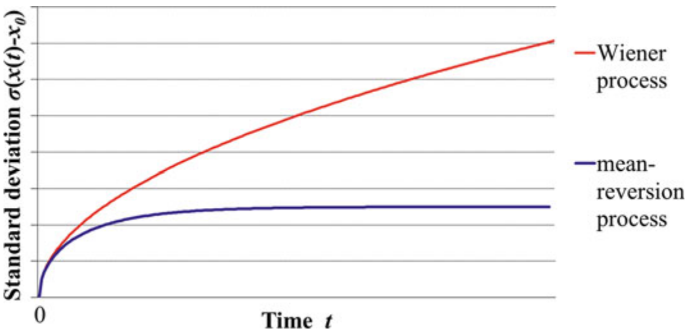


Fig. 11.4 Standard deviation of stochastic process increments as a function of the time step length for Wiener and mean-reversion processes

$$\frac{\mathrm{d}S}{S} = \mu \cdot \mathrm{d}t + \sigma \cdot \mathrm{d}z. \tag{11.8}$$

This highlights that the relative changes in prices $S(t)$ are composed of a mean rate of change μ and stochastic deviations around that mean with standard deviation σ. This is considered appropriate for stock prices because it implies that the expected return on the currently invested capital is independent of the current stock price. E.g. with t measured in years and $\mu = 0.07$, the expected annual return will be somewhat above 7% (due to compound interest effects), independently of whether the current share price is 50 or 500 €. So this may be easily connected to standard asset pricing models like the seminal **capital asset pricing model (CAPM)**. Closer mathematical scrutiny also reveals that prices under the geometric

Brownian motion (GBM) model will always remain positive if the starting price is positive. This property seems very obvious for stock prices. Yet, it is less for electricity prices, where technical constraints and market regulations have induced repeatedly negative prices, particularly in market areas with high proportions of renewables (cf. Sect. 10.1).

This indicates that a pure transposition of approaches developed in the mathematical finance literature to electricity markets may not be adequate. On the other hand, one must acknowledge that electricity is the clear exception to the rule – negative prices are almost unthinkable for storable commodities like oil and gas.[1] And even the GBM model may be considered a reasoned choice for these commodities, as notably the **Hotelling model** of price formation for exhaustible resources suggests a constant return on assets (cf. Sect. 2.3), at least at constant interest rates.

Before proceeding further, four observations are essential:

1. Mathematical finance mainly defines stochastic price processes in continuous time, as sketched above. This enables an elegant analytical treatment using stochastic calculus. Alternatively, stochastic processes may be defined in discrete time, as is current practice in econometrics. The mathematical treatment, especially for valuation purposes, is then in general less elegant. Yet when it comes to numerical estimation and simulation procedures, a discretisation of continuous time is required, and computational techniques for discrete problems have rapidly evolved over the last few decades. Hence, both approaches have their merits, and it is worth considering in applications which approach is more convenient.
2. There exist many more general and more complicated stochastic process specifications than those discussed above. Directions that have been explored by research notably include:

   - Time-varying mean: Especially, when it comes to modelling electricity spot prices, the time-varying scarcity of electricity should be captured by time-varying parameters, e.g. a time-varying mean in a mean-reversion process.
   - Time-changing volatility: In discrete time, so-called **GARCH processes** (cf. Bollerslev 1990) have become very popular to describe the volatility-clustering observable in stock and other asset prices. Several specifications like the **Heston model** (Heston 1993) exist in continuous time, which capture shifts between periods with weak and strong price changes.
   - Increments that are not normally distributed, e.g. jumps. An interesting, general model class in that field are so-called Lévy processes (Bertoin 1996), which build on independent and identical increments yet drop the normality

[1] There was an exemption during the beginning of the Corona crisis in April 2020, as oil demand suddenly sharply decreased resulting in negative prices for the US standard oil variety WTI (West Texas Intermediate). In fact, the strong demand shock coincided with a lack of spare physical storage at the delivery point – and this combination drove prices below zero given that WTI futures are settled physically, contrarily to the common practice mentioned in Sect. 8.6.

assumptions. Any **Lévy process** may be decomposed in a Brownian motion, a drift term and a pure jump process.

- **Multi-factor processes**: Prices may be driven by more than one influencing factor, e.g. electricity prices by fuel prices and scarcity of generation capacities. Correspondingly, different stochastic processes may also be needed to describe actual price characteristics, e.g. different time constants for mean reversion or a combination of mean-reverting and non-stationary components (cf. below). If some of these price components are not directly observable, we are in the presence of so-called "latent variables" which pose additional challenges in identification and estimation.

3. A fundamental property of stochastic processes is stationarity respectively its absence. This is closely linked to the **stationarity** of time series in econometrics. A stochastic process $x(t)$ is (strictly) stationary, when the distribution of $[x(t_1+\tau), x(t_2+\tau)\ldots x(t_k+\tau)]$ is independent of τ, i.e. notably, the mean and the variance of $x(t)$ are independent of t. This is the case for the mean-reverting process described above but neither for the (generalised) Wiener process nor the geometric Brownian motion. An important implication of stationarity is that the price uncertainty remains bounded when the time step length is extended (cf. Fig. 11.4). That means that even for several years ahead, prices under a mean-reverting process only have a limited range of expected values. Whether this is an appropriate property has to be checked in each application.
4. There are multiple links between stochastic price models and neighbouring disciplines like econometrics and control theory worth exploring in more advanced modelling. As with finance models, one should be thoughtful and precise when adapting approaches, e.g. from control theory to pricing issues. Societal and economic systems are made up by persons who make purposeful, individual decisions. And these may hence be described by relationships similar to those governing technical systems only under specific assumptions.

11.2 Hourly Price Forward Curves to Link Future and Spot Prices

As discussed in Sect. 8.5, future contracts are usually written at time t for delivery at time T. Yet for electricity futures, delivery is generally not specified for one single point in time but rather over a time interval $\tilde{T} = [T_1 T_2]$, e.g. a month or a year. The question then arises how the price $\tilde{F}(t, \tilde{T})$ for the future contract over the interval $\tilde{T}$ links to the future prices $F(t, T)$ at different points in time T with $T \in \tilde{T}$. Theoretically, we may argue that such a future market is not complete, meaning that not every **idiosyncratic risk** in each hour of the delivery period may be insured (or hedged) through a specific trading product. Practically, this goes along with the fact that there is not one unique rule to derive the single hour prices $F(t, T)$ from the observed prices $\tilde{F}(t, \tilde{T})$. Practitioners have, therefore, designed various approaches to overcome the gap and to construct what is known as the **hourly price forward curve** (**HPFC**). The two most important methods are those based on

(a) econometric procedures or
(b) the **typical day** approach.

Both methods take observed past spot prices as the basis for constructing a time profile of electricity prices. This profile is then adjusted to the current level of the future prices. In such a way, **arbitrage-free** hourly expected prices are obtained which may then be used to value both delivery contracts to final customers and generation profiles. Note that the obtained prices are future prices for short (hourly) periods and need to be adjusted by the adequate risk premium to obtain expected spot prices (cf. Sect. 8.5.4).

We subsequently focus on the typical day method, which may be summarised in the following five steps:

1. **Define the typical time segments $s \in S$ to be used for the analysis.**
   Example: each hour of the day, differentiated by day of the week, constitutes a separate time segment. Hence, there are a total of 168 (24 × 7) different time segments.
2. **Select the historical observation period $\tilde{T}_H$ to be used for the establishment of** the HPFC.
   Example: the three preceding calendar years.
3. **Define the mapping function $m(t)$ linking historical observations $\tilde{T}_H$ and future time steps $\tilde{T}$ to the typical time segments**.

$$m : \begin{array}{ccc} \tilde{T}_H \cup \tilde{T} & \mapsto & S \\ \tau & \rightarrow & s \end{array} \tag{11.9}$$

   Example: assign to each time step the time segment with the corresponding weekday and the corresponding time of day.
4. **Compute the average historical prices p_s for each time segment** using the formula:

$$p_s = \frac{1}{\sum_{\tau \in \tilde{T}_H} \mathbf{1}_{m(\tau)=s}} \sum_{\tau \in \tilde{T}_H} \mathbf{1}_{m(\tau)=s} \cdot p_\tau. \tag{11.10}$$

   The indicator function $\mathbf{1}_{m(\tau)=s}$ is thereby equal to one if and only if the mapping function $m(\tau)$ maps the time step τ to the time segment s, otherwise it is zero.
   Example: compute the average price in hour 8 on Mondays over the last three years

5. **Compute the average price $p_H(\tilde{T})$ for the considered future period $\tilde{T}$** based on historical prices, taking into account the occurrence frequency of the different time segment in the period $\tilde{T}$:

$$p_H(\tilde{T}) = \frac{1}{\mathrm{card}(\tilde{T})} \sum_s \left(\sum_{\tau \in \tilde{T}} \mathbf{1}_{m(\tau)=s} \right) \cdot p_s. \tag{11.11}$$

   Example: determine the average price for next year based on the frequency of the days of the week and hours of the day during next year and the previously computed prices for the time segments.
6. **Based on the price $p_H(\tilde{T})$ and the current future price $\tilde{F}(t, \tilde{T})$, the calibration factor for future hourly prices $g(t, \tilde{T})$ is determined** as follows:

$$g(t, \tilde{T}) = \frac{\tilde{F}(t, \tilde{T})}{p_H(\tilde{T})}. \tag{11.12}$$

   Example: if the current future price is 30 €/MWh and the average price based on historical values is 25 €/MWh, the calibration factor is 1.2.
7. **The calibration factor $g(t, \tilde{T})$ is used together with the mapping function to determine the hourly price $F(t, T)$** for each hour in the future from the historical average price for the corresponding time segment:

$$F(t, T) = g(t, \tilde{T}) \cdot \sum_s \mathbf{1}_{m(t)=s} \cdot p_s \tag{11.13}$$

   Example: with the factor computed previously, the future price for hour 8 on Mondays would be 1.2 times higher than the observed historical prices for this hour.

Note that a more detailed application example for this method is provided in Sect. 11.7. The adequacy of this method mainly hinges on two prerequisites:

- the appropriate selection of typical time segments and
- the absence of structural breaks between historical price structures and the expected future price structures.[2]

The first prerequisite implies a good balance between a sufficient distinction of different time segments and a sufficient number of observations per time segments to avoid substantial impacts from single outliers. Typically, one might choose every

[2] Yet all statistical and econometric methods rely in one way or another on the assumption of the absence of structural breaks.

weekday in each month as a separate typical day. But then, the question arises how public holidays should be treated: Are the Christmas holidays or Easter or regional holidays like All Saints to be treated as one single day type, or should there be a different day type for each of these holidays?

The second prerequisite leads to a preference for short historical periods, but again this has to be traded off against the limited number of observations in short periods.

A more fundamental inconvenience of this approach is that it only provides estimates of the expected hourly future prices but not the possible variability around that mean value. If this is searched for, the HPFC has to be complemented by a stochastic process describing the variations around that mean. This issue will be addressed in the following subsection.

11.3 Valuing Simple Options on a Stochastic Spot Price

Given the preceding discussion, we may now wonder what the value of a flexible generation (or demand side) option is considering future prices. To answer this question, we have to combine the elements outlined in the previous two subsections. Yet, a first terminological disambiguation is necessary: there are (at least) two meanings of the term "future prices" that we have to distinguish. The first meaning is "prices in the future", the second "prices of future contracts". To be more precise: when assessing the value of physical flexibility options in the electricity market, the key question is about "possible spot prices in the future" rather than on "current prices of future contracts". The focus is on spot prices since the physical options are to be used in the actual operation of the system – and spot prices (should) reflect the value of actual operations (cf. Sect. 7.2.3.2). The loose qualification of "possible" spot prices emphasises that the value of these physical flexibility options is related to the uncertainty surrounding operations and prices in the future.

Having this in mind, a standard recipe for valuing simple flexibility options may consist of five steps:

1. **Define the flexibility option under study**.
   In the simplest case, the flexibility option is fully characterised by its **variable cost** c^{var} in €/MWh at which it supplies additional electricity (or reduces demand) and its **capacity** K describing the achievable output rate in MW. Taking into account operational constraints or energy volume constraints (storage-type flexibilities) makes the valuation exercise more demanding (cf. below).
2. **Determine the expected spot price(s) for the valuation period**.
   Here, the method for constructing an HPFC described in Sect. 11.2 may be used.

3. **Describe the distribution of the spot price(s) around its expected value**. Here, stochastic price processes as discussed in Sect. 11.1 may be used. It is then essential to incorporate the time-varying mean as specified in step 2 into the formulation of the stochastic processes
4. **Determine the expected payoffs of the flexibility option at exercise time under the spot price distribution**. This requires a set of valuation formulas that are discussed subsequently for the case of a simple flexibility option.
5. **Obtain the current value of the flexibility option through discounting and aggregation**. The present value of the flexibility option is obtained by discounting the value at the time of delivery (so-called exercise time in finance slang). Moreover, the value may be aggregated over the relevant valuation period if it consists of more than one time step (hour).

Note that in step 2, the future prices obtained through the HPFC need in principle to be adjusted by the corresponding market risk premium to obtain expected spot prices (cf. Sect. 7.2.5.3). Conversely, the discount rate used in step 5 should in principle include not only the risk-free rate but also the risk premium. Yet practitioners tend to neglect the risk premium given the difficulty to obtain reliable estimates for it. From a theoretical perspective, one may argue that the effects in steps 2 and 5 at least partly cancel out each other, so the assumption of a zero risk premium is generally defendable.

Having clarified the preliminaries and prerequisites, we now turn towards the valuation of a simple flexibility option characterised by its variable cost c^{var} and capacity K (step 4). At given spot price S_T, the option will be used at full capacity if $S_T \geq c^{\text{var}}$, and it will not be used (by a profit-maximising operator) if $S_T < c_{\text{Var}}$. Under uncertain spot prices, the expected payoff of the option at exercise time is then given by the relationship:

$$\begin{aligned} V_{T|t}(T) &= K \cdot \int\limits_{-\infty}^{+\infty} \max(x - c^{\text{var}}; 0) f_{S_T|t}(x) \mathrm{d}x \\ &= K \cdot \int\limits_{c_{\text{Var}}}^{+\infty} (x - c_{\text{Var}}) f_{S_T|t}(x) \mathrm{d}x. \end{aligned} \tag{11.14}$$

The notation $V_{T|t}(T)$ emphasises that the option is exercised at time T (function argument T), and the value of the payoffs is also considered at time T (subscript T), yet based on the information available at time t (subscript $|t$). Note that this value is not dependent on the actual price process used for S_T, but only on the probability distribution for the prices at exercise time, here characterised by the probability density function $f_{S_{T|t}}$ and the corresponding cumulative distribution function $F_{S_{T|t}}$.

Explicit results for the option value may inter alia be obtained, if prices are normally distributed, i.e. $S_{T|t} \sim N\left(\mu_{T|t}, \sigma_{T|t}\right)$. This will notably be the case if prices result from a generalised Wiener process as given in Eq. (11.5) or of a mean-reversion price process as described in Eq. (11.6). Then, we obtain the following formula for the value:

$$\begin{aligned} V_{T|t}(T) &= K\left(\left(\mu_{T|t} - c^{\text{var}}\right)\left(1 - F_{S_T|t}(c^{\text{var}})\right) + \sigma_{T|t}{}^2 f_{S_T|t}(c^{\text{var}})\right) \\ &= K \cdot \sigma_{T|t} \cdot (\mathrm{d}\Phi(d) + \phi(d)) \end{aligned} \tag{11.15}$$

With $d = \frac{\mu_{T|t} - c^{\text{var}}}{\sigma_{T|t}}$.

Thereby Φ is the cumulative distribution function and ϕ the probability density function associated with the standard normal distribution. One may note that this result corresponds to the one obtained in finance for option values under the so-called Bachelier model (e.g. Schachermayer and Teichmann 2008). Furthermore, this total option value exceeds always the so-called **intrinsic value**, which is defined as

$$V_{T|t}^{\text{Intr}}(T) = K \cdot \max\left(\mu_{T|t} - c^{\text{var}}; 0\right). \tag{11.16}$$

This would be the option value if it were executed at the current expected price $\mu_{T|t}$. The difference between the total option value according to Eq. (11.15) and the intrinsic value is then labelled **extrinsic value** or **time value** – time value because it disappears as the exercise of the option gets closer, i.e. the uncertainty about future prices is reduced. Similar considerations have been established in finance for the Black–Scholes model that we discuss in the following section.

A small example may illustrate the point right here: Consider a flexibility option with variable costs c^{var}= 50 €/MWh, e.g. a combined cycle plant. With an expected price in the future $\mu_{T|t}$ = 60 €/MWh, the intrinsic value of the option is 10 €/MWh (cf. Eq. 11.16). If we consider a period T in the distant future, the uncertainty regarding the future price is large, e.g. the standard deviation reaches $\sigma_{T|t}$ = 20 €/MWh. Using Eq. (11.15), we then obtain the total value of the option as $V_{T|t}(T)$ = 13.96 €/MWh. This is almost 40% higher than the intrinsic value, and the extrinsic (or time) value equals 3.96 €/MWh. This value vanishes gradually if the price level remains constant while the price uncertainty decreases as the exercise time T approaches.

11.4 Analytical Approaches for Option Valuation: The Black–Scholes Model

The previously described valuation approach has the advantage that it combines rather standard methods and analytical tools of medium complexity. However, both practitioners and scientists in the field have in the past been more turned towards another option valuation approach, the famous Black–Scholes model (cf. Black and Scholes 1973, Merton 1973), respectively, its variant considering options on futures published by Black (1976).

The Black–Scholes model was originally developed for options on stocks and correspondingly, it does not consider normally distributed prices but a geometric Brownian motion as underlying stochastic price process (cf. Sect. 11.1, Eq. 11.7). Furthermore, its derivation is placed in the context of efficient, arbitrage-free markets and dynamic hedging and replication strategies (Schachermayer and Teichmann 2008). The objective of the model is to determine a "fair price" for so-called **European option**s on stocks or similar financial papers.[3] There are two types of European options (cf. Sect. 8.6):

Call options provide the holder the right (but not the obligation) to buy the underlying (the stock) at some point of time T in the future (called exercise or strike time) at a predefined price X, the so-called exercise or strike price.

Put options conversely provide the holder the right (but not the obligation) to sell the underlying at exercise time T in the future at the predefined price X.

It may be noted that the simple flexibility option discussed in Sect. 11.3 (e.g. a controllable power plant) with specified variable costs c^{var} is a real option analogy to a call option if all technical operation restrictions are disregarded. The (much less common) equivalent to a put option would be a pure flexible consumer willing to consume additional electricity below a specific price threshold – one may think of an electrolyser producing pure hydrogen and selling it at a given market price. But one has to be aware that electricity spot prices are usually not adequately modelled based on a geometric Brownian motion (cf. Sect. 11.1). Therefore, the Black–Scholes analysis is not directly transposable to flexibility options in the electricity system. Nevertheless, it is worthwhile to discuss the principles of financial option valuation based on the seminal Black–Scholes analysis.

This analysis focusses on the above-mentioned fair price, which is a price upon which sellers and buyers may agree. To be acceptable for both sides, such a price should be derived solely from objective market information and not depend on individual subjective preferences. By providing such a fair price, the Black–Scholes model has paved the way for a tremendous increase in financial derivatives trading

[3] A broad variety of options is traded on financial markets. The most standard options are labelled European and **American options**. European options may only be exercised at the exercise date, whereas American options may be exercised any time up to the exercise date. So for American options "early exercise", i.e. a use before the agreed exercise date is possible whereas it is not for European options. Real options involve a physical activity and hence obviously may not exercised in advance—they correspond to European options, or often rather to a sequence of European options (cf. Sect. 11.6).

in the four decades after its publication – until the global financial crisis in 2008 led to a deep questioning of many valuation practices. A major consequence for corporate and regulatory risk management concerning this and other similar models has been to take "model risk" seriously – and model risk arises notably from deviations between model assumptions and the real world.

This being said, the assumptions underlying the Black–Scholes model have to be scrutinised critically. On the other hand, the mathematical elegance and application simplicity of the Black–Scholes formula strongly hinge on these assumptions, which may be summarised as follows (cf. Hull 2018):

1. The price of the underlying asset follows a geometric Brownian motion.
2. Short selling of assets is possible, and there are no limitations to the use of corresponding revenues.
3. Transaction costs and taxes are negligible, and shares are infinitely divisible.
4. No dividend payment on the stock occurs [extension with dividends in Black (1976)].
5. There are no risk-free arbitrage opportunities.
6. Trading is done continuously.
7. The risk-free interest rate is constant and identical for all expiry dates.

Extensions of the Black–Scholes model aim to deal with less simplifying assumptions, yet we focus subsequently on the original model since it captures key features of option pricing. A complete mathematical treatment of the Black–Scholes model is out of scope for this book. We limit ourselves to sketching the key elements of the reasoning [for a more detailed but still accessible treatment cf. Hull (2018)]. The derivation of the valuation formula relies mainly on the three following elements:

1. Construction of a **risk-free portfolio** consisting of the option and the according underlying[4] in an appropriate ratio.
2. **No-arbitrage argument**: the risk-free portfolio will offer the same return rate as a risk-free bond.
3. **Risk-neutral evaluation**: the value of options on stocks is independent of the risk appetite of investors. Options can, therefore, be evaluated under the simplifying assumption of risk neutrality.

Considering the value $V(S,t)$ of the option as a function of the price of the underlying stock S and time t, the two first elements allow to derive the following stochastic partial differential equation, also known as the **Black–Scholes–Merton differential equation**:

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV. \tag{11.17}$$

[4] The term underlying is used in finance to designate the asset, which a derivative is based on, e.g. the shares of a particular company, cf. also Sect. 8.2.

A first important point to note on this equation is that it describes the changes in value V for all financial products[5] with the underlying S (e.g. also for forwards or complex options).[6] The differential equation has multiple solutions. These are obtained by adding specific boundary conditions to the equation. We will come back to that point later.

To provide some intuition, we take a closer look at the terms of the differential equation: the right-hand side describes the value change corresponding to interest payments based on the risk-free interest rate r. For a risk-free derivative, i.e. when both the first derivative $\frac{\partial V}{\partial S}$ and the second derivative $\frac{\partial^2 V}{\partial S^2}$ for S are zero, the interest payment corresponds to the value change over time $\frac{\partial V}{\partial t}$, as is to be expected in an arbitrage-free world. Another particular case arises for $\frac{\partial V}{\partial S} = 1$ and $\frac{\partial^2 V}{\partial S^2} = 0$. An obvious solution satisfying these boundary conditions is $V \equiv S$, i.e. the considered product is equal to the underlying (or at least always has the same value). Then, obviously $\frac{\partial V}{\partial t} = 0$, i.e. the (partial) derivative with respect to time at given asset price S is zero. While $\frac{\partial V}{\partial S}$ describes the direct dependency of the product value on the value of the underlying, the third term on the left side is less intuitive: its magnitude is determined by the variance σ^2 of the stochastic process, i.e. it is related to the stochasticity of prices. This term is labelled diffusion term. An intuitive understanding may be derived from considering the expected value change for a product with a positive second derivative $\frac{\partial^2 V}{\partial S^2} > 0$ in the presence of a discrete uncertainty for the underlying S (cf. Fig. 11.5).[7] If an up-movement $+\Delta S$ and a down-movement $-\Delta S$ of similar magnitude may occur with similar probability, the expected change in S is zero. Given the positive curvature of the value function, the expected change in V will be strictly positive, other things being equal.

With positive S and positive $\frac{\partial V}{\partial S}$ (as in Fig. 11.5) and typical magnitudes for these terms, a solution to the differential equation will then require $\frac{\partial V}{\partial t} < 0$, i.e. a product with positive second derivative with respect to S will lose value over time. This holds, other things being equal, notably for a given S. This value decrease corresponds to the loss in time value for an option. Explained differently: in the setting of Fig. 11.5, the likely up and down movements until expiry $\pm\Delta S$ decrease in size as the expiry date approaches. Then also the difference between the ex-ante expected value $\frac{(V(S_0 - \Delta S, t) + V(S_0 + \Delta S, t))}{2}$ and the realised value $V(S, t)$ shrinks – this is (a discretised version of) the loss in time value.

At the boundaries of the definition domain for the value function, boundary conditions have to be added, and these boundaries determine the specific solutions.

[5] These products are frequently subsumed under the term "**derivatives**" (cf. Chap. 8). Yet we avoid this nomenclature in the following to avoid confusion with the mathematical concept of derivatives of a function.

[6] Note that there are no indices $T|t$ or likewise to the value function V as in the previous subsection. In fact, we consider here always the value at time t evaluated with information at the same time t. Therefore, we drop these unnecessary, identical indices.

[7] Mathematically, it is a consequence of Ito's lemma, which is a fundamental theorem in stochastic calculus.

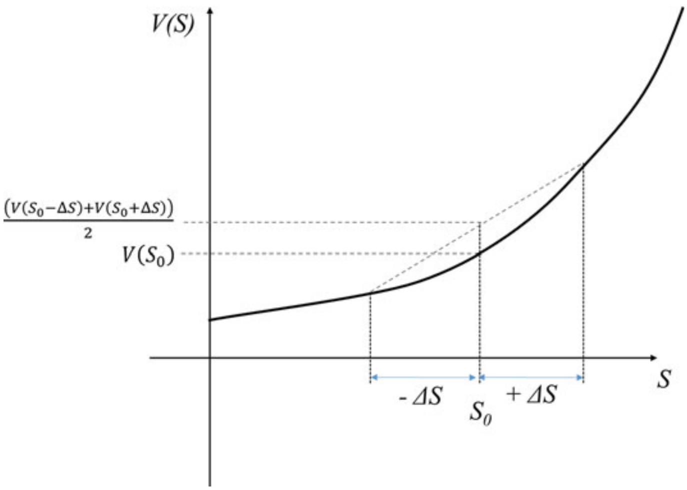


Fig. 11.5 Illustration of the diffusion term in the Black–Scholes–Merton differential equation

For the most common European options, the key boundary conditions are given by the **payoffs** at exercise time.

- for a **call option** (purchase option), this payoff may be written as follows:

$$V^{\text{Call}}(S,T) = \max(S - X, 0). \tag{11.18}$$

This condition summarises the definition of a European call option: the call option will be exercised at maturity T, if the price of the underlying S exceeds the strike price X. Then, the payoff will be equal to the positive difference $S - X$. At prices below the strike price, the option is not exercised and no payoff occurs. Additionally, the following boundary conditions are specified: $V^{\text{Call}}(0,t) = 0$ and $\lim_{S \to +\infty} \left(V^{\text{Call}}(S,t) - S \right) = 0$, i.e. the call option value is bounded by zero at low prices and by S at high prices.

- for a **put option** (sell option), the payoff at exercise time is

$$V^{\text{Put}}(S,T) = \max(X - S, 0). \tag{11.19}$$

Again, this condition describes mathematically the payoff of a European put option at maturity: it will provide a positive payoff if and only if the strike price exceeds the spot price at maturity, i.e. when it is more profitable to sell the underlying at the strike price to the option writer (seller of the option) than to the market at the current spot price. The payoff is in that case equal to the difference $X - S$.

Table 11.1 Limiting cases for option values according to the Black–Scholes formula

Limiting case		Implication	Value limit
Just before delivery of the option			
$t \to T$	$S > X$	$d_1 \to +\propto, d_2 \to +\propto$	$V^{Call} \to S(t) - X,\ V^{Put} \to 0$
	$S < X$	$d_1 \to -\propto, d_2 \to -\propto$	$V^{Call} \to 0,\ V^{Put} \to X - S(t)$
Current price far above exercise price			
$S \gg X$		$d_1 \to +\propto, d_2 \to +\propto$	$V^{Call} \to S(t) - Xe^{-r(T-t)},\ V^{Put} \to 0$
Current price far below exercise price			
$S \ll X$		$d_1 \to -\propto, d_2 \to -\propto$	$V^{Call} \to 0,\ V^{Put} \to Xe^{-r(T-t)} - S(t)$
Almost risk-free option			
$\sigma \to 0$	$S > Xe^{-r(T-t)}$	$d_1 \to +\propto, d_2 \to +\propto$	$V^{Call} \to S(t) - Xe^{-r(T-t)},\ V^{Put} \to 0$
	$S < Xe^{-r(T-t)}$	$d_1 \to -\propto, d_2 \to -\propto$	$V^{Call} \to 0,\ V^{Put} \to Xe^{-r(T-t)} - S(t)$

Further boundary conditions are again imposed—derived from limit case considerations: $V^{\text{Put}}(0,t) = X \cdot e^{-r(T-t)}$ and $\lim_{S \to +\infty} V^{\text{Put}}(S,t) = 0$. Note that the limiting value for an underlying price of zero considers the discount of the terminal payoff to the valuation time.

With these boundary conditions and under the assumptions above, Black and Scholes derive the following value formulas for European put and call options:

$$V^{\text{Call}}(S,t) = S \cdot \Phi(d_1) - X \cdot e^{-r(T-t)}\Phi(d_2) \tag{11.20}$$

and

$$V^{\text{Put}}(S,t) = X \cdot e^{-r(T-t)}\Phi(-d_2) - S \cdot \Phi(-d_1). \tag{11.21}$$

Thereby, the cumulative distribution function Φ of the standard normal distribution and the parameters given in the following formula are used.

$$\begin{aligned} d_1 &= \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= \frac{\ln\left(\frac{S}{X}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t} \end{aligned} \tag{11.22}$$

These formulas are best understood by considering various limiting cases, as summarised in Table 11.1. The first example given there is an option approaching expiry. As price uncertainty gets smaller and the boundary of the definition set is reached, the value approaches the final payoff for the option. Similarly, the reader is invited to consider the other cases listed there and to make use of Eqs. (11.20–11.22) to validate the results, cf. also Exercise 11.3.

11.5 Merits and Limits of the Black–Scholes Model for Electricity Market Analyses

The Black–Scholes model is generally considered as the reference model for valuing options in financial markets. Yet, there are multiple off-springs and alternatives to that standard model, too numerous to name. However, two are worth mentioning. Black (1976) discusses options on futures and includes a discussion of dividend-paying stocks whereas Margrabe (1978) generalises the valuation formula to options with two underlyings. The former is interesting for electricity markets (and more generally energy markets) since options therein are usually not written on the physical underlying but on futures. The latter provides a conceptual frame that allows dealing with thermal power plants as real options. We will come back to that in the next section.

In general, option valuation approaches derived from finance have found the following applications in the electricity industry and more generally the energy sector:

1. Valuation of financial options and similar products traded on the energy markets.
2. Valuation of optionalities embedded in contracts or complex products.
3. Support for hedging decisions for real options such as power plants.
4. Valuation of real options in medium to long-term perspective.

The first application field is rather straightforward yet it suffers in the case of electricity from a lack of liquidly traded options in most market places. For oil markets, this is, however, a typical usage of option price models. The second field encompasses a broad range of concrete applications – including, e.g. the evaluation of flexibility clauses in gas supply contracts. The third and fourth applications are most directly linked to the physical and system perspective on electricity markets: the applied model's assumptions must fit the actual market conditions to obtain reliable results. For the use of the Black–Scholes or similar formulas, two aspects are thereby critical:

- Given the non-storability of electricity, each spot delivery period corresponds to a separate product. For this product, price distribution parameters have to be assessed, and the corresponding real option is to be evaluated.
- Furthermore, it is questionable whether a geometric Brownian motion may adequately describe the price process for electricity spot prices. Notably, negative prices and prices of zero are not compatible with the assumption of a geometric Brownian motion process. Therefore, any application of Black–Scholes, Black (1976) or Margrabe formulas in the context of hedging or valuation of real options should be aware of the necessarily approximate nature of the results. In the following, we, therefore, follow a somewhat different route.

11.6 Thermal and Hydropower Plants as Real Options

From what we have discussed in the previous sections, five key elements may be distilled when it comes to conceptualising power plants as real options:

1. **Power plants** do not correspond to a single option on one underlying. Rather they correspond to a series of options – also called a "**strip of options**": a power plant provides production options for every delivery period of the spot market. A similar reasoning holds for demand-side flexibilities.
2. **Technical constraints** such as minimum operation times or start-up costs **limit the usage of these options**. They also prevent using simple analytical option formulas such as the ones discussed in Sects. 11.3 and 11.4.
3. If a **power plant** burns commercially traded fuels such as **hard coal or natural gas**, then it should be considered as an **option dependent on two underlyings**. Both the output electricity price and the input fuel price are time-varying and may be described by stochastic processes. If additionally emission certificates are to be used, then the option depends on three underlyings.[8]
4. **Storages** are a type of real option that does not have a common equivalent in financial options. They are usually assimilated to so-called **swing options**. Swing options describe the right to take more or less of a specified commodity over a time period.[9]
5. To value all these real options, an **adequate modelling of the price process** is vital. Assessing the value of flexibility options in the future electricity systems is particularly challenging since this requires an anticipation of the future prices, including their stochasticity.

These are key takeaways for anyone trying to link the challenging issue of valuing generation flexibilities in electricity systems to the broad literature stream of financial option valuation. By and large they are also applicable when it comes to valuing demand-side flexibilities. A few additional remarks may, however, be useful:

First, one should be aware that our treatment so far has focussed on analytical approaches to financial option valuation. Yet research in finance has also developed a broad range of numerical methods, cf. Hull (2018) for an overview. The most important classes are **Monte Carlo simulations**, (binomial) tree approaches, finite difference methods and the so-called **least-squares Monte Carlo** approach, cf. Longstaff and Schwartz (2001). Notably, the latter has emerged as a very flexible and computationally feasible method for evaluating path-dependent options such as storages or thermal power plants with operation restrictions.

[8] Pushing even further, a CHP plant with heat as second output besides electricity is dependent on four underlyings.

[9] Swing options have been introduced in the finance literature mostly to describe the characteristics of common gas contracts, which include minimum and maximum delivery quantities, cf. e.g. Jaillet et al. (2004).

Especially for storage valuation, numerical methods are crucial since there are no analytical valuation formulas readily available neither for swing options nor in general for storage plants. For thermal power plants, it may be quite useful to disregard operation restrictions and use analytical formula to obtain an upper bound to the flexibility value.

When the dependency of thermal power plant valuation on input factor prices is to be taken into account, then considering the spread between input factor costs and output prices is advantageous. For the Black–Scholes model, a corresponding generalisation has been developed by Margrabe (1978). He develops an analytical formula for an option dependent on the spread between two underlyings. Thereby, the option value is driven by the volatility of the price ratio of the two underlyings. There is then also not a specific strike price. Rather the exercise of the option depends on the ratio of the two commodity prices. The corresponding spread is called "spark spread" for gas-fired power plants, which corresponds to the gross margin at given commodity prices. For coal-fired power plants, the term "dark spread" is used. For an application to European power plants, an extension is required to include besides fuel also CO_2 certificates as input factor with separate price risks. This is then a "clean spark spread", respectively, a "clean dark spread". Yet such models are still based on several questionable assumptions, and therefore, we subsequently rather pursue a different approach – namely the application of the previously developed simple models to an actual flexibility valuation for a power plant.

11.7 Application: HPFC and Parsimonious Real Option Valuation for Thermal Power Plants

To assess the future value for a power plant, we have to first link the available market quotes for derivative products (in occurrence for quarter 3 of 2016) to hourly expected spot prices. This is done by establishing first an hourly price forward curve (cf. Sect. 11.2). Then, the flexibility value of an (idealised) CCGT plant for the considered period, here from July to September 2016, is determined based on historical data, in occurrence those available by the end of 2015. Thereby, the simple valuation approach described in Sect. 11.3 is used. The data used for the study as well as the corresponding spreadsheet *HPFC_Optvalue.xlsx* contained in the electronic appendix to this chapter.

For the construction of the HPFC, we apply the typical day method, with one typical day for each weekday. Yet as consumption and price patterns on Tuesdays to Thursdays are rather similar, they are aggregated to one typical day. For reasons of simplicity, we use only 2015 data to construct the HPFC. Following the procedure described in Sect. 11.2 above, we get for the corresponding steps (cf. also Table 11.2):

Table 11.2 Key elements for an HPFC for Q3 2016 based on price data of 2015

Row no.	Typical days s	Historical values Average prices p_s		Future frequencies in Q3 2016
		Base	**Peak**	**Number of days**
(1)	Monday	35.28	40.10	13
(2)	Tuesday–Thursday	35.74	39.07	39
(3)	Friday	35.92	38.31	14
(4)	Saturday	28.83		13
(5)	Sunday	22.33		13
		Future values for Q3 2016		
		Base	**Peak**	**Off-peak**
(6)	Number of hours	2208	792	1416
(7)	Weighted historical average $p_H(\tilde{T})$	32.83	39.11	29.32
(8)	Futures $\tilde{F}(t,\tilde{T})$ on Dec 30, 2015	27.94	33.80	24.69 (computed)
(9)	Calibration factor $g(t,\tilde{T})$	0.851	0.863	0.842

1. **Definition of the typical time segments:**

$$S = \{\text{'Mon } h1\text{'}, \text{'Mon } h2\text{'}, \ldots, \text{'Mon } h24\text{'}, \text{'Tue}-\text{Thu } h1\text{'},$$
$$\text{'Tue}-\text{Thu } h2\text{'}, \ldots, \text{'Fri } h1\text{'}, \ldots, \text{'Sun } h24\text{'}\}$$

   Hence, there are 5 typical days and 120 different typical time segments.

2. **Selection of the historical observation period $\tilde{T}_H$:** As proposed above, we only use 2015 data as historical observations, i.e., limiting ourselves to the summer months, we get

$$\tilde{T}_H = \{\text{'Jul 1 2015}, h1\text{'}, \text{'Jul 1 2015}, h2\text{'}, \ldots, \text{'Sep 30 2015}, h24\text{'}\}$$

3. **Definition of the mapping function $s = m(t)$:** we map each observation in the historical period $\tilde{T}_H$ onto the typical time segment with the corresponding weekday (respectively, the weekday aggregation Tue–Thu) and the same hour. The same is done for the future time period $\tilde{T}$.
   There is no concise mathematical description of the mapping function, yet it may be easily implemented in software code (cf. electronic supplement).

4. **Computation of the average historical prices p_s for each time segment:** The average price in hour 8 on Mondays over Q3 2015 is found to be 44.86 €/MWh. Prices averaged over base and peak periods and typical days are also indicated in Table 11.2, rows labelled (1) to (5).
5. **Computation of the average price $p_H(\tilde{T})$ for the considered future period $\tilde{T}$ based on historical prices:**
   The results are given in row (7) of Table 11.2, using the frequencies indicated in rows (1)–(5) in the right-hand column. Besides the average base and peak price, also an off-peak price is computed.
6. **Determination of the calibration factor $g(t, \tilde{T})$:**
   Based on the prices $p_H(\tilde{T})$ (row (7)) and the current future price $\tilde{F}(t, \tilde{T})$ (row (8)), the calibration factors $g(t, \tilde{T})$ are determined as indicated in row (9) of Table 11.2.
7. **Use of the calibration factors $g(t, \tilde{T})$:** to have a unique calibration factor for each time segment, we use the calibration factor obtained for peak hours for hours $h9$ to $h20$ on Mondays to Fridays. For all other time segments, the off-peak calibration factor is used. The base calibration factor is hence only given for information purposes.

The resulting prices for the typical time segments are shown graphically in Fig. 11.6. It is thereby evident that prices on Saturdays and especially Sundays are on average lower than during the week. In addition, the early Monday morning hours are more similar to weekend hours than to other weekdays.

With the hourly price forward curve, we may compute the intrinsic value for a thermal power plant. To determine the total option value along the approach developed in Sect. 11.3, including the time value, we have to estimate the standard deviation for the spot prices. A straightforward way to do so is to use the same data as for the estimation of the price forward curve.

We, therefore, compute for each hour of each typical day the standard deviation of the prices around the observed mean. They are then calibrated using the same factors as for the HPFC. The resulting standard deviations and expected prices (HPFC) are plotted for the typical day Tuesday–Thursday in Fig. 11.7.

In the same graph, we show the results from applying the option valuation formula derived in Sect. 11.3, namely Eq. (11.15). It becomes evident that the option value of the stylized power plant is close to zero during night hours when expected prices are far below variable cost and that the value increases to about 10 €/MWh during morning and evening hours. The option is said to be “deep in the money”, i.e. it is very unlikely that it is not used, and the value is close to the (positive) difference between expected price and variable cost, which is the **intrinsic value**. Comparing the option value for hours 9 and 20, the impact of time-varying volatility becomes obvious. Although the expected price (and correspondingly the intrinsic value) is slightly higher in hour 20, the option value is

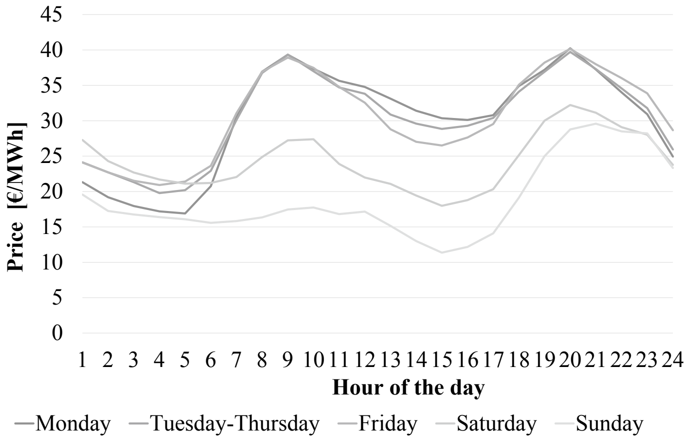


Fig. 11.6 Hourly price forward price curve (HPFC) for summer (Q3) 2016 based on price data of 2015

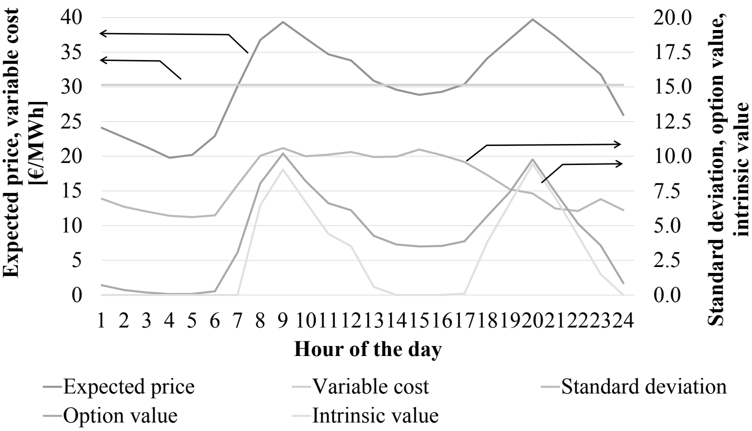


Fig. 11.7 Prices and option value for a power plant on an hourly basis

higher in hour 9 due to the higher price uncertainty. The highest difference between the total option value and the intrinsic value, i.e. the highest **extrinsic value**, occurs when the expected price is close to the variable cost, i.e. in hours 13–17.

The obtained values may be compared to the actual realisations of spot prices and option values during the period Q3 2016. For single hours, stochastic deviations may strongly influence the result. Therefore, we focus on the average values over the 2216 h of the period under question. The results are summarised in Table 11.3. It turns out that the ex-ante option value (left column) exceeds the realised option value if the variable cost as of the end of 2015 (30.29 €/MWh) is

Table 11.3 Backtesting of option values for a gas plant in summer (Q3) 2016

Ex-ante value end 2015	Ex-post value at constant variable cost	Ex-post value at actual variable cost
3.08 €/MW/h	2.09 €/MW/h	4.70 €/MW/h

used (middle column), cf. Fig. 11.7. In that comparison, the realised option value is lower by roughly one third. On the other hand, when taking the actual gas and CO_2 spot prices as a basis for the variable cost, the realised value (right column) exceeds the option value by roughly 50%. Hence, the model provides a first rough approximation, yet it needs to be enhanced to cope with fuel and CO_2 prices uncertainties for more accurate results.

11.8 Challenge: From Asset to System Perspective

We now come back to the question that served as a starting point of our discussion of flexibilities in the electricity system: What is the value of flexible assets in a future sustainable electricity system? One key issue has to be tackled: the endogeneity of market prices in bottom-up electricity system models. Put differently: the methods described in the previous sections, be it the Black–Scholes model or the Bachelier model, treat prices as exogenous (stochastic) input factors. From a system perspective, prices result from the interplay between supply and demand, including their respective rigidities and flexibilities. Therefore, prices and quantities are determined simultaneously in a stochastic equilibrium. And whenever some kind of storage is part of the flexibilities under consideration, this stochastic equilibrium will be one interlinking multiple periods in the year. Solving such an equilibrium in a detailed system modelling approach is challenging.

If we want to evaluate a single flexibility in the context of a prespecified electricity system, there is yet a possible way out: we can start with a stochastic process describing the fluctuations in residual load and then make use of a simple supply-stack model as described in Sect. 7.1.1 to transform the demand fluctuations into price variations.[10] Then, the flexibility may be valued against these prices using standard numerical approaches for option valuation, notably the **least-squares Monte Carlo** approach (cf. Longstaff and Schwartz 2001; Nadarajah et al. 2017, see also Sect. 8.6). Yet one must be aware that this approach breaks down as soon as larger quantities of this flexibility are introduced in the market – because then, the flexibility will start to influence prices in the market. And also the valuation of one flexibility (e.g. batteries) in the presence of another (e.g. pumped hydro storage) is only possible if the latter's operation and pricing strategy are approximated.

[10] The so-called ParFuM-model used by Kallabis et al. (2016) and Beran et al. (2019) is a somewhat more sophisticated version of a merit-order type model that may be applied in that context, cf. Pape (2018) for an application with more long-term focus.

Even more challenging would such an undertaking become if investments into the technologies are to be treated endogenously. In the context of fuel price uncertainty, a corresponding approach has been proposed in Weber (2005), yet this does not cover the full challenge of uncertain renewable power infeed. Hence, important research challenges are still ahead in that field.

11.9 Further Reading

Hull, J. (2021). Options, Futures and other Derivatives. 11th edition. Harlow et al.: Pearson.

This seminal textbook discusses the derivative markets and the various methods to value options on financial markets. It provides an introduction to the world of stochastic calculus applied in finance. Beyond that, it also includes a small chapter on energy and other commodity derivatives.

Burger, M., Schindlmayr, G., & Graeber, B. (2014). Managing Energy Risk. A Practical Guide for Risk Management in Power, Gas and other Energy Markets. 2nd edition. Chichester: Wiley.

The book provides an accessible mathematical treatment of energy trading and the corresponding risks, including the valuation of optionalities.

11.10 Self-check of Knowledge and Exercises

Self-check of Knowledge

1. What is the simple stochastic process in continuous time that serves as the basis for defining other, more complex stochastic processes? What are the key properties of this process?
2. Give the formulas of the following stochastic processes: generalised Wiener process, geometric Brownian motion and mean-reversion process. Indicate also key application areas for these processes.
3. What is an hourly price forward curve and what is it used for?
4. Why are power plants called real options?
5. Explain the basic principles that are used to derive the Black–Scholes option pricing formulas.
6. When is the time value of an option highest? What are the implications for the value of a flexible power plant – especially, when the difference between the expected price (from an hourly forward curve) and variable costs changes?

Exercise 11.1: Mean-Reversion Process

A mean-reversion process according to Eq. (11.6) applied to electricity spot prices p leads to the equation:

$$\mathrm{d}p = \kappa \cdot (\mu - p) \cdot \mathrm{d}t + \sigma \cdot \mathrm{d}z. \tag{11.23}$$

It can be shown that with given price $p(t_0)$, a solution of the stochastic differential equation may be written as

$$p(t) = \left(1 - e^{-\kappa(t-t_0)}\right) \cdot \mu + e^{-\kappa(t-t_0)} \cdot p(t_0) + \sigma\sqrt{\frac{1 - e^{-2\kappa(t-t_0)}}{2\kappa}}\varepsilon \tag{11.24}$$

with ε distributed according to a standard normal distribution, i.e. $\varepsilon \sim N(0{,}1)$. This may also be rewritten using the notation $\Delta p = p(t) - p(t_0)$ and $\Delta t = t - t_0$:

$$\Delta p = \left(1 - e^{-\kappa\Delta t}\right) \cdot \mu - \left(1 - e^{-\kappa\Delta t}\right) \cdot p(t_0) + \sigma\sqrt{\frac{1 - e^{-2\kappa\Delta t}}{2\kappa\Delta t}}\varepsilon. \tag{11.25}$$

1. Use the time series of daily average spot prices given below to estimate the parameters of the linear regression:

$$\Delta p_t = a + b \cdot p_{t-1} + \tilde{\varepsilon}. \tag{11.26}$$

2. Compare the terms in Eqs. (11.25) and (11.26) to derive formulas to compute the parameters κ, μ and σ of the mean-reversion process from the regression results.
3. Compute the estimated values $\hat{\kappa}$, $\hat{\mu}$ and $\hat{\sigma}$ from the regression parameters $\hat{a}$, $\hat{b}$ and $\hat{\sigma}_{\tilde{\varepsilon}}$, where $\hat{\sigma}_{\tilde{\varepsilon}}$ corresponds to the estimated standard deviation of $\tilde{\varepsilon}$.
In case, you have not solved part (2) of the exercise, you may use the relationships:

$$\hat{\kappa} = -\frac{1}{\Delta t}\ln(1+\hat{b}) \quad \hat{\mu} = -\frac{\hat{a}}{\hat{b}} \quad \hat{\sigma} = \hat{\sigma}_{\tilde{\varepsilon}}\sqrt{\frac{2\ln(1+\hat{b})}{(1+\hat{b})^2 - 1}}. \tag{11.27}$$

4. Compare the terms in Eq. (11.24) to those of a naïve discretisation of Eq. (11.23) obtained by simply replacing the infinitesimal differences d by discrete differences and using the property given in Eq. (11.3). Using a Taylor series expansion, you may demonstrate that the two converge when Δt tends towards zero.

Exercise 11.2: Hourly Price Forward Curve

The objective is to compute an hourly price forward curve for spot prices on Mondays in February 2021 based on historical observations from preceding years. Collect the historical data for all days in Februarys between, e.g., 2011 and 2020.

Table 11.4 Computation scheme for an HPFC for Feb 2021 based on information available on Nov. 19, 2020

Line no.	Typical days s	Historical values Average prices p_s		Future frequencies in Feb 2021
		Base	Peak	Number of days
1.	Monday			
2.	Tuesday–Thursday			
3.	Friday			
4.	Saturday			
5.	Sunday			
		Future values for February 2021		
		Base	Peak	Off-peak
6.	Number of hours			
7.	Weighted historical average $p_H(\tilde{T})$			
8.	Futures $\tilde{F}(t, \tilde{T})$ on Nov. 19, 2020			
9.	Calibration factor $g(t, \tilde{T})$			

The computation is to be based on the information available on November 19, 2020. On that day, the price quote for Germany at the EEX was 40.39 €/MWh for the product base Feb 2021 and the quote for the product peak Feb 2021 49.86 €/MWh.

1. You may perform the necessary computations using Excel and insert the intermediate results step-by-step into Table 11.4 (cf. also the similar Table 11.2).
2. Make a diagram showing both the average hourly historical prices for Mondays in February and the obtained HPFC for 2021. What are your key observations?
3. February 2021 is still amidst the COVID-19 pandemics that started to swipe over Europe in March 2020. What adjustments, if any, are advisable on the HPFC to reflect the ongoing pandemic situation?
4. Do you expect a lignite power plant with variable costs of 21 €/MWh will be in the money during all hours in February 2021? Why?

Exercise 11.3: Valuation of Financial Options

Evaluate a European call option on a financial stock using the Black–Scholes option pricing model.

The current underlying price is 41.72 €, and the annual volatility σ is estimated at 50%. The risk-free rate is assumed to be 3%. There are 262 trading days per year.

1. Evaluate the option with a time to maturity of 53 (trading) days and a strike price of 44 €. Thereby, you may use Excel and implement the Black–Scholes formulas for option pricing given in Eqs. (11.20–11.22).

2. Now evaluate the value of the corresponding put option with the same expiry date and same strike price. You may use the formulas again from above or make use of the so-called put-call parity:

$$V^{\mathrm{Call}}(S,t) - V^{\mathrm{Put}}(S,t) = S_t - Xe^{-rT} \tag{11.28}$$

3. What happens to the option values when you double the time to maturity? And what if the volatility is doubled?
4. Why is this valuation approach not appropriate when assessing the flexibility value of a power plant?

Exercise 11.4: Valuation of a Power Plant as a Real Option
We aim to determine the hourly value of a power plant with variable costs of 44 €/MWh for a Monday in February 2021 based on the information available on Nov. 19, 2020.

1. Use the HPFC determined in Exercise 11.2 and compute the intrinsic value of the power plant for each hour of this Monday in February.
2. Assume the price volatility for all Monday hours in February is 9.38 €/MWh. What is then the total option value in each hour based on the Bachelier model? You may use Eq. (11.15) to compute this value.
3. Compare the total option values obtained for the different hours of the day – both among themselves and with the corresponding intrinsic values computed in the previous step.
4. Compare the average of the hourly option values with the option value obtained for a financial option with rather similar parameters in Exercise 11.3. What drives the difference? You may also compute the option value using the Bachelier model for the average daily price to support your analysis.

References

Beran, P., Pape, C., & Weber, C. (2019). Modelling German electricity wholesale spot prices with a parsimonious fundamental model – Validation & application. *Utilities Policy, 58*, 27–39. https://doi.org/10.1016/j.jup.2019.01.008.

Bertoin, J. (1996). *Lévy processes*. Cambridge University Press.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics, 3*, 167–179.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*, 637–654.

Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics, 74*, 498–505.

Burger, M., Schindlmayr, G., & Graeber, B. (2014). *Managing energy risk. A practical guide for risk management in power, gas and other energy markets* (2nd ed.). Wiley.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies, 6*, 327–343. https://doi.org/10.1093/rfs/6.2.327.

Hull, J. (2021). *Options, futures, and other derivatives* (11th ed.). Pearson.

Jaillet, P., Ronn, E. I., & Tompaidis, S. (2004). Valuation of commodity-based swing options. *Management Science, 50*, 909–921.

Kallabis, T., Pape, C., & Weber, C. (2016). The plunge in German electricity futures prices – analysis using a parsimonious fundamental model. *Energy Policy, 95*, 280–290. https://doi.org/10.1016/j.enpol.2016.04.025.

Longstaff, F. A., & Schwartz, E. S. (2001). Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies, 14*, 113–147.

Margrabe, W. (1978). The value of an option to exchange one asset for another. *The Journal of Finance, 33*, 177–186.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science, 4*, 141–183.

Nadarajah, S., Margot, F., & Secomandi, N. (2017). Comparison of least squares Monte Carlo methods with applications to energy real options. *European Journal of Operational Research, 256*, 196–204.

Pape, C. (2018). The impact of intraday markets on the market value of flexibility – Decomposing effects on profile and the imbalance costs. *Energy Economics, 76*, 186–201. https://doi.org/10.1016/j.eneco.2018.10.004.

Schachermayer, W., & Teichmann, J. (2008). How close are the option pricing formulas of Bachelier and Black-Merton-Scholes? *Mathematical Finance, 18*, 155–170.

Weber, C. (2005). *Uncertainty in the electric power industry – methods and models for decision support*. Springer.

12 Moving Towards Sustainable Electricity Systems

Climate change is one of the largest challenges to humankind in the twenty-first century. Climate action is notably one of the 17 Sustainable Development Goals of the United Nations (UN 2015), see Sect. 12.5. It is strongly related to the patterns of use and production of energy as discussed in the previous chapters. Therefore, we want to distil key issues in this final chapter that have to be solved to move towards **sustainable electricity systems**. Thereby, we build on the models and concepts laid out in the previous chapters to highlight how decisions may benefit from the existing body of knowledge on power systems and their economics. Furthermore, we point at the needs for further research and development.

In view of a concise discussion, we focus on the greenhouse gas (GHG) emission reduction to be achieved to limit **global warming** to 1.5–2 °C as requested by the Paris Agreement. Consequently, we do not try to determine the optimal level of decarbonisation,[1] nor do we develop detailed transition pathways. The transition should occur as rapidly as possible, especially limiting global warming to 1.5 °C or less requires already drastic emission reductions until 2030 at a global scale according to climate scientists. But, this is not in focus here; instead, we look at the future system that meets the tight CO_2 emissions bounds defined above and is also sustainable in the sense that it could be perpetuated over hundreds or thousands of years.

Besides **decarbonisation**,[2] three development trends are currently key drivers for electricity systems worldwide (cf. inter alia Di Silvestre et al. 2018; EY 2019; Fulli et al. 2019):

[1] See Sect. 6.2 for a discussion why this is difficult.

[2] We use the technical term decarbonisation here being aware that the ultimate target is a defossilisation.

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2_12

- **Decentralisation**
- **Digitisation**
- **Acceptance** and **participation**.

The emphasis in energy strategy debates is frequently laid on the three Ds: decarbonisation, decentralisation and digitisation as game changers for future electricity systems. Yet, these have to be complemented by aspects of acceptance and participation, at least in liberal societies and democratic countries. New installations of many energy technologies, including but not limited to coal-fired and nuclear power plants, onshore wind energy and overhead power lines, face significant public opposition. This shapes energy strategies, although it is not necessarily specific to energy installations, as road and airport infrastructures and new industrial and commercial sites are as well subject to heavy public debates.

Also, digitisation is not a subject specific to the energy industry – rather, it transforms products, markets and supply chains in almost all economic sectors. The volume, velocity and variety of data being available or becoming so in the near future also enable new coordination mechanisms and markets in the energy field. Yet, these digitalisation technologies rather complement than substitute new generation, transmission and storage facilities, as we will see in the subsequent sections.

Decentralisation is a trend more specific to the energy industry, although it is not undisputed (cf. Leopoldina et al. 2020). It has many facets, and the term is also given a variety of significations. But, it is centred around the observation that renewable energy installations usually come with smaller unit sizes than conventional nuclear or fossil generation units. This contributes to the operational and design challenges for the energy infrastructure. Yet, it seems questionable whether this trend would be stable without the decarbonisation needs in the energy system.

Subsequently, we therefore focus on decarbonisation and its implications for the future electricity system. We first look at the general decision alternatives and challenges in decarbonisation (cf. Sect. 12.1). Then, we consider more specifically three domains where key challenges arise for decarbonised future sustainable electricity systems with high shares of intermittent renewables:

- Balancing supply and demand (Sect. 12.2).
- Grid operation and development (Sect. 12.3).
- Prosumer integration and network tariffication (Sect. 12.4).

To our understanding, these issues are closely linked, but they may be seen as a sequence of increasing complexity. The first one considers the electricity system collapsed to a single market. The second adds the challenges of spatial distribution and grid constraints, whereas the third also reflects the multiplicity of involved actors and the resulting regulatory and institutional challenges.

To provide a coherent overview, we subsequently identify subitems in some of the domains. We structure the presentation by discussing the **key challenges** in the field followed by the **technical solutions** available – or still to be developed. We then address what **regulatory and market-based concepts** have to be advanced and highlight **political and societal aspects** to be addressed beyond the scope of simple technoeconomic analyses. Against this setting, we discuss which **modelling approaches** may help to provide decision support – referring here (as in the earlier sections) back to the corresponding sections of the previous chapters. We conclude by proposing some key insights – not with the intention to provide definitive answers on the issues at stake, but rather to indicate starting points to the readers for their in-depth investigations.

Key Learning Objectives

After having gone through this chapter, you will be able to

- Describe major key trends in the energy sector and corresponding challenges.
- Explain consequences and challenges of the transformation of the energy system.
- Describe technical solutions and their combination to balance supply and demand in a future energy system with high shares of renewables.
- Explain the impact of the energy system transformation on frequency control and power system operation.
- Understand the needs and drivers of the adaptation of the transmission and distribution grid infrastructure.
- Explain the trade-off between investments in power networks and costs for congestion management.
- Explain the impact of the energy system transformation on ancillary services, especially reactive power provision.
- Explain the challenges and chances of a higher energy autonomy of prosumers for the energy system.

12.1 Challenges in Decarbonisation

Key challenges: according to the Fifth Assessment Report of the **IPCC** (IPCC 2014), limiting global warming to 1.5 °C or less is only possible in scenarios where global GHG emissions are reduced by 41–72% in 2050,[3] compared to 2010 levels and to near zero by 2100. Even achieving the 2 °C target requires emission reductions of 50% by 2050 and at least 75% by 2100. Developing countries and emerging economies mostly have lower per capita emission levels than

[3] The broad range of required emission reduction refers to the 10th and 90th percentiles of the scenarios of the IPCC (2014).

industrialised countries. Still, the development of these countries tends to increase per capita emissions to "Western" levels – although these should rather decrease. Industrialised countries will thus have to reduce their emissions by at least 80% by 2050. Even a decrease by 95% or more may be required to fulfil the 1.5 °C target. As electricity generation is more easily decarbonised than transport or industrial processes, many researchers even claim that we must achieve 100% carbon-free electricity to reach the Paris Agreement's climate goals. This may also help to decarbonise the transport, industry and heating sectors – with hydrogen and derived fuels as a complement for deep decarbonisation of branches like air and sea transport or steel making (cf. Ball and Wietschel 2010).

Technical solutions: decarbonising the electricity sector implies reducing carbon dioxide emissions resulting from the combustion of fossil fuels. Four approaches may be envisaged here:

- Massive use of technologies using **renewable energies** (see Sect. 4.2).
- Reduction of the electricity consumption, notably through **energy efficiency improvements** (see Sects. 2.4.1 and 3.1.4).
- Use **carbon capture and storage** technologies (see Sect. 6.2.2.3) to reduce the CO_2 emissions released into the atmosphere. If used on fossil-fired power plants, this may reduce emissions by up to 90%. If combined with biomass plants, even negative net emissions may be achieved since the biomass formation absorbs CO_2 from the atmosphere, then deposited through CCS in the underground (if managed sustainably).
- Extended use of **nuclear energy** (see Sect. 4.1.2). Nuclear energy does not lead to direct CO_2 emissions. Some GHG emissions occur in the fuel extraction and processing and the waste disposal cycle, yet these are rather limited. But, the use of nuclear energy raises several other environmental and safety concerns (see Sect. 4.1.2.2).

These solutions are not mutually exclusive, and different countries may strive for different mixes. The large-scale use of variable renewable sources like wind and solar as foreseen by the current EU strategy implies strong supply fluctuations, requiring a flexible complement, e.g. battery storage or gas turbines. Both nuclear and fossil-fuelled power plants with CCS have limitations in flexibility, notably when it comes to complete shutdowns and subsequent restarts.

Regulatory and market-based concepts: following the lines of mainstream economic thinking, straightforward first-best approaches to regulation for decarbonising electricity systems will be instrumental for achieving decarbonisation goals. As discussed in Sect. 6.2.3, these are a (Pigou) tax on GHG emissions or an emission certificate trading scheme with a cap on emissions. Strong theoretical arguments favour these instruments, but it is challenging to implement them globally in the short to medium run. Empirical evidence indicates that far-reaching policies are difficult to coordinate globally, and so far, more countries have put in place specific renewable support schemes than emission taxes or cap-and-trade systems.

Political and societal aspects: the practical relevance of so-called second-best instruments suggests that political and societal factors deserve a more careful consideration than done under standard assumptions in environmental and energy economic theory. In particular, the following simplistic assumptions about policymaking and regulation are not always fulfilled:

1. The idealistic conception of a benevolent, unique government.
2. The presumption that this government or the institutions it is emanating from (elections, parliament, etc.) has a clear and consistent preference ordering on which they base their decisions.
3. The misleading assumption that as distributional effects of policies can be compensated, they should not be a major concern in decision making.
4. The simplifying view that the government will base its decisions on the best available objective information.

Replacing these simplistic assumptions with more realistic conceptions (and not other simplifying assumptions) is likely to lead to a more processual view on policymaking, where the context of decision making and the stakeholders involved matter more. So-called second-best instruments such as support schemes for clean technologies may then be preferred as more feasible actions.

Modelling: the electricity market modelling approaches discussed in Chap. 7 are in principle well suited to model the implications of the policy instruments for decarbonisation. A CO_2 tax may be included in the variable costs of the different fossil generation technologies (as a markup on fuel prices based on tax level and the emissions factors given in Sect. 6.2.2).

A CO_2 emission cap can be implemented as an additional constraint in the optimisation programme by linking the production output to emissions via emission factors. When modelling the impact of the emission cap, one has yet to be thoughtful about the model scope compared to the scope of the emission trading system. This is true for the geographical and temporal scope as well as the sectoral scope. E.g. the EU ETS covers not only the electricity sector, but also a number of energy-intensive production sectors such as pulp and paper or iron-making. Hence, these sectors may also be incorporated into the model. Else, it has to be decided beforehand which share of the total certificates will be available for the electricity system. Similar considerations apply if only a subset of the relevant EU and associated countries is modelled. And, for short- to mid-term analyses, the transfer of emission credits between years deserves particular attention.

Quantity-based second-best instruments like renewable certificates may also be included in the market models by imposing an additional constraint. The impact of classical feed-in tariffs on renewable installations is yet beyond the scope of electricity market models as discussed in Chap. 7 since there is no link between electricity market outcomes and renewable investments and operation under this framework.

Moreover, coping with the political and societal aspects discussed above requires approaches going beyond the models outlined in Chap. 7. Those may, e.g., be used to assess distributional effects on different technology classes (by computing revenues and costs based on the model outcomes, including shadow prices). Multiple objectives or even inconsistent preferences may yet not directly be handled in the discussed models. In case all except one objective can be expressed through level constraints (like a CO_2 emission cap), the trade-offs, e.g. between costs and emissions, may however be illustrated by running the model repeatedly with varying levels for the constraint.

When it comes to climate change as a global environmental problem, any modelling effort has to acknowledge that all four assumptions mentioned above are violated in that case: there are multiple national governments involved (in contrast to assumption 1), the priority given to mitigating climate change evolves over time and according to context factors (assumption 2), resource-rich countries like OPEC members or Russia fear losing through tight climate agreements and the distribution of costs is in general highly disputed between countries (assumption 3), and finally the dispute over policy measures is also to quite some extent accompanied by a debate about the reliability of scientific measurements and model calculations on climate change (assumption 4). Hence, any modelling exercise based on welfare maximisation or cost minimisation must clearly be viewed as an idealising benchmark for identifying optimal solutions. It may provide guidance towards best achievable outcomes, yet actually implementing such policies will require additional efforts in negotiations and execution.

Challenges by substituting fossil raw materials with other (critical) raw materials: with the energy system transformation, the need for fossil raw materials will be replaced by other eventually critical raw materials. Wind and PV, supplemented by different storage technologies, e.g. batteries, will lead to new requirements concerning critical resources demanded and possibly to new environmental and social impacts. To analyse resource availability (see Sect. 2.3.1 regarding indicators for the availability of reserves and resources), inventories have to be set up showing which materials will most probably be used in future renewable and storage technologies. Even with today's technologies, renewable-based electricity systems lead to diverse resource necessities (cf. e.g. Giurco et al. 2019). To analyse possible future bottlenecks, a large quantity of information is needed (cf. Angerer et al. 2016, pp. 54–108): on the one hand, the future diffusion of different technologies in the electricity sectors worldwide over the next decades and the applied resources for these technologies have to be estimated. Furthermore, the critical resources may also be demanded in other sectors. Hence, the assessment has to be extended to these further sectors. On the other hand, to identify possible shortfalls, the demand for critical resources must be compared with the future supply of the different critical resources. Correspondingly, close scrutiny and the acknowledgement of relevant uncertainties are necessary when assessing critical resource requirements due to defossilisation strategies.

In this context, technoeconomic interdependencies are potentially relevant: as soon as a shortage is becoming apparent, prices for the corresponding materials will increase. This may trigger new activities like an intensification of the search for new deposits, more efficient production technologies, the substitution of critical resources by other materials and new ways of recycling to reuse the critical resource (cf. Wellmer and Dalheimer 2012).

Studies assessing the criticality of resources generally include indicators of economic importance and supply risk and sometimes also consider the environmental impact (cf. European Commission 2017). Most existing studies dealing with the availability of the critical resources needed for the transformation of electricity systems do not identify any insurmountable problem. Depending on the assumptions made, yet certain platinum-group metals and rare earth elements, as well as other metals like indium and tellurium, are found to be potentially exposed to critical supply risks (cf. e.g. Angerer et al. 2016; Moss et al. 2011, 2013; European Commission 2017; Viebahn et al. 2015).

Key insights: the preceding considerations highlight that different technological options may be used to progress towards a massive decarbonisation of the electricity system. Empirical evidence suggests that even within Europe, there are no consistent preferences regarding the different generation technologies. E. g. countries like France and Finland consider nuclear as a part of a sustainable electricity mix, whereas other countries like Germany, Denmark or Italy either have plans to phase out nuclear or have never installed nuclear reactors. Also, CCS is still envisaged as an option in some countries, whereas others are solely focussing on renewables. Coping with this heterogeneity in preferences, political choices are undoubtedly one challenge for market modelling in view of sustainable electricity systems. Independent from that, the European Union has a clear strategy to extend renewable energy sources in its member countries. Renewable energy sources play an essential role in future electricity systems of European member states and other countries worldwide. For electricity systems with a strong focus on renewables (as in Europe), the fluctuations in renewable supply are a key challenge that will be considered in the following sections. Furthermore, future technologies deserve additional assessments regarding the scarcity of materials and environmental and social impact.

12.2 Challenges in Balancing Supply and Demand

Market clearing and grid operations in European-style electricity markets are mostly unbundled (cf. Sect. 6.1 and Chap. 10). Consequently, the balancing between supply and demand has to be considered at two time scales. First, within the timeframe where markets are operating (cf. Sect. 12.2.1) and second in the period after "gate closure" where the grid operators have the responsibility of coping with any remaining imbalances through the use of reserves (cf. Sect. 12.2.2).

12.2.1 Balancing Energy Production and Demand

Key challenges: the use of CO_2-free electricity seems to be promising in sectors more challenging to decarbonise ("sector coupling"), and sector coupling technologies can increase flexibility in the electricity system. Hence, the relevance of the energy carrier electricity is likely to increase further in future. On the one hand, the industrial sector might use electricity directly or, on the other hand, use hydrogen produced via electrolysis or synthetic fuels produced via electrolysis and methanisation. Note that the latter transformation path requires an additional carbon source. In the transportation sector, electric mobility is gaining market shares in many countries. With more widespread realisation of such strategies for **sector coupling**, electricity demand will increase considerably – making energy-efficiency improvements an important complement in future. In any case, massive investments on the electricity supply side still have to be implemented. Renewable energy technologies, like wind and PV, will count for the bulk of these investments. To maintain supply adequacy, investments in secured generation, like (combined cycle) gas turbines (in the long-term fired with biogas or synthetic gases), storage technologies or demand-side flexibilities will have to complement, so that sufficient capacity is available to cover the load at any time (see Sect. 10.5). Another possibility is to import electricity, which requires, on the one hand, sufficient generation capacities in other regions during hours of scarcity in the importing region and, on the other hand, enough transmission capacities. Especially when there is hardly any feed-in from wind and PV, but a high electricity demand (so-called cold dark calm), the need for secured capacity will arise. To identify the optimal strategy to overcome **dark calms**, analyses of the frequency and duration of these situations are of utmost importance. For potential solutions like short-term storage technologies and shifting electricity demand, in-depth investigation may assess their contributions to cover periods of several days. Not only dark calms will challenge the electricity system, but also significant surplus of renewable feed-in will more often occur when political renewable extension objectives will be reached. In addition to the need to provide reliable power, the further expansion of renewable energies will in future also pose the no less important challenge that excess electricity must be integrated to a large extent. This requires not only technically feasible but also economically attractive solutions.

Also, the market mechanisms deserve increased attention: during hours with a lot of electricity feed-in by renewables, wholesale electricity prices will be low if not zero or negative (merit-order effect, see Sect. 7.1). Hours with a scarce electricity supply and wholesale prices exceeding short-run generation costs may become even rarer. Corresponding volatile revenues may discourage investments, and thus, further research efforts may be devoted to alternative market designs.

In an operational perspective, a supply side mainly based on wind and PV will result in strong short-term fluctuations of the electricity provided. Such short-term fluctuations originate, e.g. from variations of wind speeds in the case of wind turbines and the passing of clouds in the case of PV. To compensate for these fluctuations and to balance electricity supply and demand on a short-term basis, flexible technologies and electricity markets with short lead-times will play an important role.

Technical solutions: to balance the scheduled electricity supply and demand under these circumstances, different technical solutions can be used. These options are yet not mutually exclusive; rather, a combination of these flexibility options might be envisaged:

- "**Backup system**": a backup system consisting of power plants that can be operated in a very flexible way is held available (cf. Sect. 4.3.1). The backup power plants have, on the one hand, to bridge the cold dark calm. On the other hand, they should have the technical capability to quickly ramp up and down and often compensate for the short-term fluctuations mentioned above. These backup power plants will be used only for a limited number of full-load operating hours during the year, so typically peak load capacities like open-cycle gas turbines (OCGT) or combined cycle gas turbines (CCGT) come into consideration – possibly run in the future with (green) hydrogen. Furthermore, it might be attractive to use generation capacities already available but typically used for other purposes. Small-sized decentralised generation, like combined heat and power plants using internal combustion engines, which have so far been installed with the intention to avoid high electricity (and heat) end consumer prices, e.g. in industrial companies, could be used to make a more significant contribution to supply adequacy. This could require strategies to use the additional produced heat. As the backup system might consist mainly of generation units, this option must be complemented by a strategy to avoid surplus situations. Often, curtailment is used to prevent surplus electricity, meaning that wind and solar resources are curtailed. While this solves the technical problem, the economic and ecological effects are negative, as available installations to produce electricity without (CO_2) emissions and hardly any variable costs are not used.
- "**Storage**": the balance between electricity supply and demand can also be ensured by the installation of storage units (see Sect. 5.2). Long-term storage systems, e.g. via the use of the gas grid with the help of power-to-gas (PtG), can help bridge a cold dark calm. Stationary batteries, which become more and more attractive from an economic point of view, e.g. at the household level, as consumers want to avoid high end user electricity prices, have a limited storage capacity and are widely distributed, so they seem to be more appropriate to compensate short-term fluctuations on a daily basis, e.g. of PV installations.
- "**Demand-side flexibility**": the needed flexibility to balance electricity supply and demand can also be provided by the shifting of loads on the demand side, so-called flexible loads. In hours with a surplus of electricity, appliances are switched on, or more energy-intense operations are pushed forward to increase electricity consumption. In hours with scarcity, just the opposite is done: loads are shifted away to reduce the electricity demand. Flexible loads can be found in all demand sectors, e.g. in households, the shifting of the operation time of heat pumps or tumble dryers. In transportation, the electricity demand for electric vehicles is suitable for load shifting, as, due to the highly long parking time, the

charging of the cars offers considerable potential. Yet demand-side flexibilities generally are only suitable to compensate short-term fluctuations, since demand shifts over more than 24 hours are hardly feasible.

- "**Imports/exports**": if the feed-in from fluctuating renewables exceeds the country's electricity demand, this "surplus" electricity may be exported to neighbouring countries. And contrariwise in scarcity situations, electricity can be imported from neighbouring countries. Such a solution requires, on the one hand, sufficient cross-border transmission capacities. On the other hand, this strategy becomes less effective if the corresponding import/export countries have a similar electricity mix, also based on fluctuating renewables. Then, the surplus and scarcity situations in the different countries will be at least partly correlated.

Market-based and regulatory concepts: balancing supply and demand implies coordination between a multitude of stakeholders in a more decentralised future energy system. Market design and regulatory framework conditions have here a key role to play. Appropriate incentives can make supply-side and demand-side capacities and storage units available to secure supply adequacy. The incentives and regulatory concepts especially need further development for demand-side flexibilities provided by a multitude of electric devices with different owners (e.g. domestic appliances), so that capacities are available when needed. It is still an open question whether a market design solely based on an energy-only market is sufficient to set appropriate incentives for investments in an adequate number of controllable units, especially in a power system predominantly based on fluctuating renewables (see Sect. 10.5). This even raises the question, who is responsible for the provision of an adequate level of supply adequacy in a liberalised market, where market participants base their investment decisions on their expectations of the corresponding profitability.

Trading electricity close to physical delivery enables reactions to short-term fluctuations of wind and PV. As weather forecasts become more and more precise, the shorter the forecast horizon is, the shorter the lead-time of electricity trade is, the smaller the difference between expected and real electricity generation will be. Consequently, harmonisation efforts in European electricity markets with shorter time scales (e.g. intraday markets) are currently of high importance.

Furthermore, as batteries are mainly used to reduce industrial or household customers' electricity supply from the grid by optimising self-consumption, new incentive schemes are required to use these batteries for short-term flexibility. In addition, there is the need for incentives aligned with wholesale market prices to obtain electricity feed-in from decentralised generation units in industry (and households) during hours of scarcity and to deviate from the firm-specific production profiles. For the realisation of a reliable shifting of loads (see Sects. 3.1.5 and 10.5), a system based on prices reflecting the scarcity or surplus situation and voluntary reactions of the customers may lead to a higher acceptance by the customers at the expense of relatively limited reliability. The alternative is the direct load control, e.g. by the energy supplier, associated with discounts on the corresponding energy tariff.

Political and societal aspects: without the participation of market players, the technical solutions will not be implemented. Hence, the right incentives are crucial as is acceptance. This is true both for the construction of new generation, storage and transmission capacities and for a modified operational mode of existing installations available on the demand side. It may be economically efficient to use private installations like decentralised heat and power plants available in the household sector (e.g. in multi-family houses) or flexible loads in industry to maximise the feed-in or to minimise the demand during hours of scarcity. Yet, this raises concerns among the customers affected. A market-based, extreme form of load shedding would be to totally cut off the electricity supply for those consumers, who accept this when rewarded with a low electricity tariff. This could effectively be the basis for a complete privatisation of supply adequacy.

Modelling: energy and electricity system models can determine the optimal mix of the different flexibility options ensuring an exogenously given level of supply adequacy. But to consider short-term fluctuations and long-term investment decisions, the complexity of the model is relatively high (cf. e.g. Seljom and Tomasgard 2015). Furthermore, uncertainties have to be considered, e.g. concerning the availability of different technologies. This can be done by examining feed-in profiles of renewables in different weather years and stochastic outages for conventional power plants, e.g. using a Monte Carlo simulation. Agent-based simulation (ABS) approaches seem to be a suitable type of energy model to analyse the question whether sufficient investments are realised to ensure supply adequacy. This form of simulation allows modelling the investment decisions of different market players with their specific investment behaviour (according to their preferences). Such a kind of analysis yet requires comprehensive data. ABS-based energy models may not include a restriction to satisfy the energy demand unconditionally. Rather, the available capacity emerges from the independently taken investment decisions of the different market players. This might lead to a situation where the exchange or system operator cannot clear the market due to an inadequate supply adequacy level as a result of the given market design (cf. Fraunholz et al. 2021a). And the challenge of increasing surplus of renewable energy has to be tackled, not only in a technical way, but also how economical solutions can be implemented.

Short-term market equilibrium models can be used to analyse the effects of short-term fluctuations of electricity production based on renewable energies. Again, a high temporal resolution and the consideration of uncertainties are indispensable in an energy system based on fluctuating renewables.

Key insights: this subchapter shows that existing technical solutions and especially their combination must be used to balance electricity supply and demand. It will be important to adapt the market design to the new framework conditions. This is likely to result in a market segment for capacity remuneration and will definitely lead to markets with gate closure close to physical delivery. The short-term production fluctuations, inevitably occurring with wind and PV, are challenging. Still, with the help of continuously improving forecast methods, supply-side storage and demand-side technologies, increased levels of ramping flexibility and energy markets with steadily reducing lead-times, a renewable energy system with shares of renewables close to 100% seems to be manageable from a technical point of view. The main challenge in this field might be rather the bridging of the cold dark calm. To bridge this gap, the provision of capacity is needed, which will only be operated for a very limited time, resulting in economic challenges due to the coverage of the fixed costs. Furthermore, acceptance is needed for the construction of new generation, storage and transmission capacities.

12.2.2 Balancing Short-Term Fluctuations

Key challenges: the expansion of wind and PV results in a higher dependency of electricity generation on weather conditions. Network operation and especially the balancing of the system have to cope with more significant uncertainties. These uncertainties are mainly driven by uncertain weather forecasts, which induce deviations between expected and actual power generations. These deviations may concern different time scales, from less than seconds up to hours or even days. This section addresses only the short-term response to deviations, i.e. primarily frequency control (see Sect. 5.1.4.2). For this task, resources with different technical characteristics are used at different time scales. These resources are deployed in order of response speed, from milliseconds (usually inertia) to a few seconds, minutes or even hours – based on corresponding reserve products ranging from frequency containment reserve up to replacement reserve (see Sects. 5.1.4.2 and 10.3). This section covers both inertia and reserves for frequency control.

Excess demand (or supply) results in a drop (or increase) of rotational frequency. However, at the shortest time scale, the inertia[4] of generators prevents a much stronger deviation. The **inertia** of power systems decreases as more and more inverter-based generation and storage units (PV, most wind turbines and batteries) as well as loads are connected to the system. This results in a power system with a lower frequency stability which raises concerns among many grid operators. In such a low(er) inertia grid, there is a risk of experiencing an excessive frequency change after a contingency, such as an outage of a power plant or a transmission line. This may result in cascade effects by tripping other generators due to a high frequency deviation.

While inertia is relevant to cushion frequency drops immediately after a disturbance, **frequency containment reserves** are automatically activated and aim at stabilising the frequency after an incident (resulting in an imbalance) within the synchronous area.

[4] *Inertia* is a property of (rotating) masses such as large synchronous generators in conventional power plants. It may be understood as a resistance to any change in velocity. As such it limits the short-term impact of imbalances between power supply and demand for electric power systems. This is similar to the effect of the mass of a car (or other vehicle) when the brakes are activated: the inertia associated with the mass of the car prevents the car from stopping immediately under the action of the brakes – and ditto the speeding up under the effect of the accelerator is limited.

Technical solutions: technical solutions for power grids with low rotational inertia (and thus faster frequency dynamics) are the usage of faster reserves (e.g. labelled fast frequency reserve by the Finnish TSO) as well as the provision of synthetic inertia, also known as virtual inertia or inertia mimicking, provided by special converter control equipment in wind and photovoltaic generation units or storage units. Moreover, battery storage systems may play a growing role in future energy systems due to their fast response behaviour resulting in synthetic inertia. The power electronics in the inverter of a battery storage system can measure system frequency similar to speed sensors in conventional generators and respond accordingly.

Market-based and regulatory concepts: while technical solutions are primordial for the provision of inertia, there is also an important regulatory aspect: should the provision of virtual inertia become a mandatory connection requirement for inverter-based storage and generation units? Or should there be a remuneration for this service, either based on the installed capabilities or based on the short-term contracts? These questions become more pressing in future electricity systems when periods with only renewable-based generation become more frequent. For (fast) reserves, market-based procurement and the corresponding regulatory concepts are expected to play an important role. Key issues are the **dimensioning of reserves** and the duration of reserve provision: forecast errors of renewable feed-in depend on their actual power feed-in. For example during night time, the forecast error of photovoltaic is zero as no feed-in can occur. On a winter day, feed-in from photovoltaic units may be highly variable and difficult to predict, as (unforeseen) snow or fog at a sunny day may lead to an heavily reduced electricity generation (in comparison to the forecast). Depending on the weather situation, wind energy can also be subjected to large forecast errors. As a consequence, demand for short-term balancing is expected to be more dynamic in future. Today, the dimensioning of reserves is rather static in Europe, but is this adequate for the future? In this context, static means that the reserve demand is kept constant over multiple days or weeks, while dynamic means an adaptation of the actual reserve demand to the respective situation. A dynamic dimensioning of the reserves can contribute to a more efficient procurement of reserves.[5] Besides the dynamic dimensioning, the duration of reserve provision also has an impact on potential technologies, which can provide frequency services. Notably, storage technologies are typically not able to provide reserves over longer periods. Consequently, the lead-times of the balancing tenders and the duration of reserve provision should be adapted to the new requirements.

Political and societal aspects: on a European level, the guideline on electricity balancing (Commission Regulation (EU) 2017/2195 of 23 November 2017) is a first achievement towards harmonising European control reserve markets. The regulation establishes an EU-wide set of technical, operational and market rules to govern the functioning of control reserve markets. It sets out rules for the

[5] The interested reader is referred to Zipf (2021, Chap. 6), where the topic of reserve dimensioning is quantitatively assessed.

procurement of reserve capacity, the activation and pricing of reserve energy and the financial settlement of balance responsible parties. Furthermore, the regulation requires developing harmonised methodologies, i.e. detailed binding rules that transmission system operators (TSOs) shall apply. As a further step, common European platforms for operating the imbalance netting process and enabling the exchange of balancing energy are likely to be established. Besides these regulatory aspects, a widespread generation mix across Europe contributes to a well-balanced portfolio.

Modelling: the field of low rotational inertia is already addressed in some articles. Ulbig et al. (2014) provide a good starting basis for interested readers. The authors investigate the impact of low rotational inertia on power system stability and operation, and they show possible mitigation options.

There are several modelling approaches in the fields of dimensioning of reserves, contract duration or harmonisation of capacity procurement. Subsequently, only two exemplary articles are mentioned for this research field. Dallinger et al. (2016) focus on the procurement of upwards and downwards reserve products and how the joint procurement of frequency restoration reserves across countries influences wholesale electricity market clearings. They conclude that in addition to asymmetric procurement of upwards and downwards balancing capacity, common procurement has significant advantages in terms of cost reductions. Bucksteeg et al. (2016) address the topic of dynamic reserve dimensioning. They propose an improved dynamic reserve sizing method using non-parametric distributions as a forecast error description and show the economic benefits.

Key insights: low rotational inertia leads to faster frequency dynamics in power systems. This makes frequency control and power system operation more challenging. Wind, photovoltaic and battery systems have to be adjusted to provide synthetic rotational inertia. Especially, battery storage systems may play an increasingly important role in future energy systems due to their fast response behaviour, which enables the provision of synthetic inertia.

Besides the continued need for inertia, reserve requirements will increase in future in a system with higher feed-in variability and unpredictability. Consequently, currently applied static methods for dimensioning reserves as well as lead-times and the duration of reserve provision are under investigation. Adjustments of control reserve mechanisms and harmonisation efforts of the European regulators can be expected during the transition towards sustainable electricity systems.

12.3 Challenges for Grid Operation and Development

Photovoltaics and wind energy are playing a crucial role in the transformation of the electricity system. Both technologies have characteristics that affect grid development and operation: 1. feed-in of electricity (at least partially) takes place in distribution grids, especially photovoltaics and a large share of onshore wind energy are connected to distribution grids – while offshore wind energy is typically connected to the transmission grid, 2. the amount of inverters and electronic equipment is increasing in combination with a displacement of inertia in the system and 3. plants will primarily be built in regions with favourable weather conditions, which will often not coincide with the consumption centres. Especially for offshore wind energy, but also for onshore, this may result in a discrepancy between the location of production and the place of consumption, resulting in the need for transporting the produced electricity over long distances. This challenge can be divided into two aspects: one is that the electricity must be transported, and another is that the transport infrastructure might not yet exist where it is needed.[6]

These issues imply new tasks for strategic and operational planning of both transmission and distribution grids and require a strengthened coordination between transmission and distribution grid operators.

12.3.1 Grid Extension and Reinforcement Needs

Key challenges: the transformation of the electricity system requires the adaptation of the infrastructure, both of transmission and distribution grids. Having a look at the history of power grids, developments over the last century show that power grids have been adapted to developments both on the supply and demand sides. And, the currently ongoing transformation of the electricity system also results in a need for adaptation of the grid infrastructure. Especially, the expansion of wind energy results in increasing transport distances in the power grid. However, larger network transports also arise in general for decentralised electricity supply when varying weather conditions and corresponding volatile feed-ins lead to time-varying regional surpluses.

Technical solutions: the extension of networks is one solution to cope with the higher transport amounts. This can be realised by investments in new transmission lines or reinforcements of existing lines. The extension of existing lines can be realised by investments in new cables (e.g. double overhead line system) or by operative measures to increase the transport capacity on existing cables. This can be reached e.g. by **dynamic line rating**, also known as real-time thermal rating (RTTR). Dependent on the environmental conditions (primarily, the cooling of the line by outside weather conditions), the maximum load of the line is increased without compromising safety.

[6] Of course, also so-called “renewable pull” – meaning that the availability of renewable power attracts new industries (e.g. the high renewable energy availability in Brandenburg is sometimes referred as one important factor for the location selection of the Tesla factory, however, energy is likely to be just a minor factor in this decision) – can result in an increased demand close to attractive renewable sites. Although, renewable pull is an interesting research field from various research perspectives, the effects on the energy system are limited, so that they are not further elaborated here.

Market-based and regulatory concepts: when optimising the costs of infrastructure, there is generally a trade-off between the costs of grid extension and the costs of congestion management. The more power lines are installed, coming close to a perfect transport system ("copper plate"), which means that there will be sufficient grids for transporting all energy, the higher the investment for installing the necessary capacities. However, without adaptation of power grids to the situation (hence, no investments in power grids), heavily congested grids may occur inducing high congestion management costs for transmission grid operators. Congestion management costs are currently driven by two cost factors in European markets: curtailment of renewables and redispatch of conventional capacities. When optimising costs for infrastructure, the degree of grid extension impacts costs of congestion management. As costs of grid extension increase but congestion management costs decrease with the level of new grid investments, there is, in consequence, a minimum at an optimal degree of grid extension (see Fig. 12.1). However, as supply and demand are changing over time, this optimal degree is not stable. Moreover, there is a difference in time horizons and reversibility: grid investments are large, long-term investments that have to be amortised over many decades. Congestion management costs are purely operational costs. This difference in temporal scope also comes along with uncertainties relating to demand and decentralised generation development, which make this trade-off even more challenging.

Political and societal aspects: an urgent question regarding the successful transition of electricity systems is how social acceptance for this infrastructure transformation can be achieved. This is not only relevant for transmission lines but also for generation and storage units. As generation units based on renewable energies, e.g. wind power plants, will be widely distributed, most citizens are facing generation and transportation units, which is less the case in traditional electricity systems with less numerous but much larger power plants typically located nearby load centres. According to Wüstenhagen et al. (2007), social acceptance can be differentiated into socio-political, community and market acceptance. Using this differentiation, the central challenge seems to be related to the community acceptance, as primarily affected citizens rise against new electricity generation, storage and transportation units in their neighbourhood. The techno-economic feasibility of sustainable electricity systems will not be sufficient for a successful transition. Rather, the support of local stakeholders will be of huge importance. It is thus indispensable to address this issue from the very beginning of the planning process for a new project (cf. e.g. Perlaviciute et al. 2018; Perras 2015) and to develop possibilities to let local players (communities, residents, local companies, etc.) participate and benefit from the project (cf. e.g. Local Energy Consulting 2020, pp. 23–29).

The lack of community acceptance of new infrastructures, notably of additional grid capacities required to reduce congestions, is also often referred to as NIMBY phenomenon ("Not in my backyard[7]"). Transmission system operators in Europe are currently facing lengthy authorisation processes and strong public opposition to their transmission line projects. These often result in substantial project delays; e.g. only a limited number of new transmission lines have been built in Europe in the last years. In Europe, the Ten-year Network Development Plan (TYNDP) has been set in place, which should help to prioritize the most relevant energy infrastructure projects. Furthermore, these plans have been specified on a more detailed level in some European countries (see e.g. Grid extension plan in Germany). Additionally, grid expansion acceleration laws have been implemented in some European countries (e.g. Germany) to support a fast(er) grid extension.

Modelling: several modelling approaches address the topic of grid extension in Europe on different time scales (e.g. from 2030 to 2050) and different levels of regional aggregation (e.g. European vs. national focus). While a complete analysis of the European power grid is generally realised with transhipment models (cf. Sect. 7.2.2) and thus an approximation of the transfer capacities, national analyses of power grids are generally more detailed regarding the modelled grid infrastructure using mainly DC load flow models (cf. Sect. 7.3). An example of such an analysis of the extension of the European power grid at a quiet highly aggregated level can be found in Müller and Gunkel (2013). In contrast, Gunkel (2020) analyses the factors influencing the transmission grid extension from a more detailed perspective using a DC approximation applied to the German energy system.

Key insights: the transition towards higher shares of renewable energies comes along with an adaptation of the current electricity transmission and distribution system. Spatial and temporal imbalances between demand and supply imply significant grid extensions at transmission and distribution levels. The transformation towards a renewable-based energy system thus requires successfully adapting the transmission and distribution grid infrastructure to the new transport and distribution needs. Lagging grid development may threaten the further installation of renewables and could thus put the transformation of the energy system in jeopardy. Consequently, grid extension will remain a high priority, which is also shown by the network development plans at the European and national levels.

The trade-off between grid investments and congestions is especially relevant for transmission grids; however, this trade-off also exists for distribution grids between grid extension costs and curtailment.

[7] NIMBY characterizes an opposition or resistance by residents against proposed developments in their local area.

12.3.2 Congestion Management and Market Design

Key challenges: network extension and building up a congestion-free network are not always economical, especially when the full capacity is only required in a few hours of a year. Consequently, congestions will occur in sustainable electricity systems. To realise an efficient operation, a market-based allocation of scarce transmission capacity is advantageous. Thereby, the physical characteristics of power flows have to be taken into account to manage congestions at lowest cost. This has to be complemented by an adequate grid expansion in the long-run perspective as discussed in the previous section.

Technical solutions: in a first step, topology optimisation is applied to reduce congestions in the power grid. However, given the physical characteristics of the power grid, these measures are not always sufficient. In the European electricity system, **redispatch** is used as a next step to avoid line overloading (see Sect. 10.6.2). Redispatch is an intervention to adapt the power feed-in of power plants at the request of the transmission system operator to avoid or eliminate regional overloads of individual equipment in the transmission system. **Curtailment** is often closely related to redispatch (see Sect. 10.6.2) and means that energy delivery from a renewable generator is reduced below the feasible output level.[8]

Market-based and regulatory concepts: congestion management raises two regulatory issues: one is the appropriate delimitation of market zones to handle congestion effectively, and the other is achieving an economically efficient trade-off between grid expansion and remaining congestions (see Sect. 12.3.1).

As power markets in Europe are organised as spot markets with larger market zones, congestions may not necessarily occur between two market areas but can also arise within one market area (see Sect. 10.6). From an economic point of view, nodal pricing is often seen as a favourable approach to organize electricity markets. A nodal pricing approach considers all infrastructure data – supply, demand and the electricity network – to determine local prices (see Sect. 7.3). Consequently, congestions directly affect local market prices, and economists argue that the invisible hand of this market provides the best (locational) price signal to steer dispatch and investment decisions locally. The related issues have been discussed in Sects. 7.3, 10.6 and 10.8, and further elements may be found inter alia in Neuhoff et al. (2011). As an alternative that requires less fundamental changes in the European market design, splitting large market zones into smaller ones can be envisaged. Such a market split (or more generally a “bidding zone reconfiguration”) may lower congestion management costs. However, it has to be orientated towards physical constraints of the electricity network, and market zones have to be regularly adapted to the physical necessities, which means that they may not be stable over time (cf. e.g. Felling et al. 2019 and Fraunholz et al. 2021b).

[8] Not only redispatch related to grid necessities can result in curtailment, also a market-based curtailment may happen, if e.g. electricity prices are below zero.

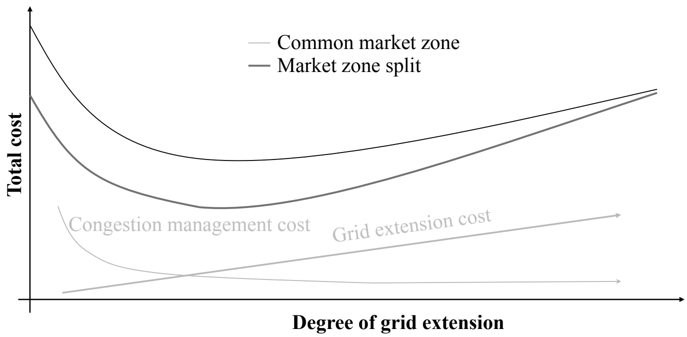


Fig. 12.1 Illustrative trade-off between grid extension, congestion management and market splitting

In the hypothetical case of very limited congestions (i.e. a "copper plate"), the method of congestion management would have little impact. Market results would be relatively similar for a common market zone, a split one or within a nodal pricing approach. Yet in the presence of substantial bottlenecks, congestion management matters.

But at the same time, long-term measures to alleviate congestions may be advisable, notably an expansion of the electricity grid. This leads to a trade-off between grid extension costs, congestion management costs and market splitting, as schematically depicted in Fig. 12.1. From the perspective of a risk-averse decision-maker, an optimal degree of grid extension should tend to be more on the side of overinvesting (right of the minimum point in the diagram) as costs for congestion management in the case of underinvestment are more steeply increasing. However, this statement strongly depends on the situation and the individual cost curves in the electricity system.[9]

Redispatch within distribution grids is currently less relevant. However, more and more generation capacity is connected to the distribution grids. In this context, an effective cooperation between transmission and distribution system operators is getting more and more important. With the transformation of the energy system, the cooperation between transmission and distribution system operators (TSOs and DSOs) becomes essential for grid operation. This is especially true for all kinds of ancillary services (cf. Sects. 10.3 and 10.4), starting from the provision of reserve power, via the provision of redispatch, voltage control and finally the availability of black start capability. This implies that interfaces for data exchange between TSO and DSO as well as access to relevant flexibility measures have to be developed. Furthermore, the regulation in most European countries is to be adapted to these new kinds of ancillary services from distribution grids.

[9] For example in Germany, congestions in the grids especially occur for the transport of energy from north (high wind capacity) to south. Consequently, grid extensions are planned (documented in the grid development plans) and are to be realized to a large extent by new HVDC corridors from north to south (see also Sects. 5.1.1.2 and 5.1.3.2).

Political and societal aspects: high amounts of congestions management costs in transmission grids result in higher grid fees, which have to be borne by the final consumers. The introduction of market zones or a nodal pricing system – as a market-based solution for reducing congestion management – is furthermore likely to face substantial political and societal resistance.

At a more fundamental level, the current market design with a central spot market is well designed for technologies with different marginal costs. The different marginal costs of the technologies determine the dispatch of generation. However, in a future market with significant shares of technologies with zero marginal costs, the dispatch decision cannot be derived from the difference of marginal costs – at least in absence of substantial storage capacities. Regionality, especially a surplus or lack of electricity provision in a region, will be decisive for power plant dispatch. This makes a strong argument to place a lot more emphasis on locational aspects in a future market design since the use of power plants will be primarily determined by the regional availability of renewable resources. While nodal pricing systems will bring several advantages for a system with scattered feed-in with zero marginal costs, policymakers and society are likely to sustain a single-price zone per country due to the vital path dependencies.

Modelling: to analyse congestion management in zonal markets, a two-step approach is necessary in general. A fundamental market model is used to determine market results, and in a second step, a load flow model determines congestion volumes. Examples of such approaches are described in Kunz (2012) and Fraunholz et al. (2021b). In Kunz (2012), the European approach of managing congestion in transmission grids is analysed and benefits of an integrated congestion management regime are quantified. A closer cooperation of national TSOs should be aimed at as benefits can be achieved. Furthermore, Kunz (2012) investigates the German congestion management regime in more detail and determines the impact of higher renewable generation on congestion management costs. He shows that joint generation and transmission infrastructure development is essential for congestion management as otherwise substantial increases in congestion management costs can be expected. In Fraunholz et al. (2021b), the effect of market splitting on redispatch costs is in the focus of the analysis. The authors show that a German market splitting substantially impacts day-ahead electricity prices, the investment planning of generation companies, required congestion management and, ultimately, system costs and social welfare. Furthermore, the authors show that the optimal zonal configuration for 2020 is becoming inefficient over time due to dynamic effects such as grid extension, renewable expansion and investments in new power plants.

Key insights: due to the economic characteristics of transmission investments, a completely congestion-free grid is not economical. In consequence, there is a trade-off between investments in grids and costs for congestion management. The challenge is to combine an efficient operation of the grid in the short run with adequate expansion in the long-run perspective. Different congestion management methods have been developed to allocate transmission capacity. The interested reader is referred to Neuhoff et al. (2011), where a detailed overview of congestion management methods and criteria to assess these is given.

12.3.3 Voltage Control and Reactive Power Management

Key challenges: with the energy system transformation, generation is shifted from the transmission to the distribution grid. These changes affect the voltage profiles in the different grid levels and the reactive power behaviour of the individual grid components. Reactive power flexibility is required to counteract voltage deviations and keep the operating voltage within the established limits everywhere in the grid (see Sect. 5.1.4.2). Conventional power plants in the transmission grid have historically been an essential source of flexible, reactive power. The availability of conventional sources to provide reactive power is reduced due to an increasing electricity feed-in from renewable energy sources. Voltage stability and reactive power management are important issues in energy systems with higher shares of electricity from renewable resources. The decrease of large power plants (as well as their reduced dispatch) leads to diminishing reactive power potentials in the transmission grid.

Technical solutions: besides the investment in hardware (e.g. inductors, capacitors or static VAR compensators[10] – see Sect. 5.1.3), power units in the distribution grids may support voltage stability and reactive power management by a controlled reactive power exchange with the transmission grid (cf. Hinz and Möst 2018). Decentralised generation units (such as recently installed wind turbines and PV parks) are technically capable of providing reactive power. However, this requires information and communication systems connecting the decentralised units with the control centre. Coordination of reactive power provision in the distribution grid requires strong(er) cooperation between TSOs and DSOs. Digitalisation is one enabler supporting the provision of reactive power from smaller units within distribution grids.

Market-based and regulatory concepts: while dedicated markets for (active) power exist in most European countries (see Chap. 10), reactive power markets and even incentivising remuneration schemes are very scarce. Two reasons can be given for that: 1. reactive power cannot be transported over longer distances, and accordingly, reactive power has to be provided locally. 2. In current energy systems with significant capacities of conventional power plants, the generation of reactive power usually does not cause any or at least only low variable costs. Generators can supply a certain range of reactive power even at full load. As reactive power is required locally to ensure voltage stability, remuneration mechanisms should ideally allow regional differentiation. In general, regulatory and market-based concepts can be distinguished. A comprehensive overview of remuneration concepts for reactive power can be found in Hinz (2017).

[10] Static VAR compensators are electrical devices for providing fast-acting reactive power on high-voltage electricity transmission networks. VAR indicates that reactive power is provided as it just refers to the unit Volt-Ampere Reactive. The term static expresses that the compensation takes place without the use of rotating machines such as three-phase synchronous machines in phase shifter operation.

Political and societal aspects: several technical solutions for providing reactive power in power systems with high renewable shares exist. Currently, reactive power services are provided by conventional generators generally without compensation. As future energy systems might lack these services, the question arises who can provide these services and how they should be incentivised. Switzerland has implemented an interesting remuneration scheme, which seems straightforward and quite effective and could thus serve as a good example or even benchmark.[11] The provision of reactive power both from generators and distribution system operators has been effective in Switzerland since 2011 based on a voltage-based incentive. The transmission system operators generate a day-ahead and, if necessary, an intraday voltage schedule for every node in the electricity grid which serves as guideline for the corresponding generators and loads. If the provision of reactive power from a source conforms to the requirements in at least 80% of all cases, a premium is paid for every provision. Conversely, a penalty has to be paid for all non-conform reactive power provisions. If reactive power in at least 30% of all cases cannot be delivered for two months in a row, the participant will receive an inactive status.

Modelling: several approaches are addressing the topic of reactive power provision from distributed sources. However, only a few modelling approaches exist that address the subject from a system perspective. One example with a system perspective is given in Hinz and Möst (2018) assessing the provision of reactive power from the 110 kV grid to support the transmission grid. The authors show that a controllable and situation-dependent feed-in of reactive power from distributed sources (mainly renewable sources) depending on the individual situation could not only reduce grid losses in the distribution grid but also facilitate a flexible, reactive power exchange with the transmission grid to support the voltage stability of the system. A more methodological contribution is provided by Larscheid et al. (2018). The authors discuss a simplifying modelling approach for reactive power which uses functional correlations between the reactive power demand and aggregated grid parameters. These correlations are determined using simulations of generic distribution grids corresponding to structurally different regions in Germany.

Key insights: reactive power provision is a topic with increasing relevance, as reactive power is currently provided by conventional power plants, which are being phased out over the coming decades with the transformation of the energy system. In general, different technical solutions are available to provide reactive power. Currently, incentive mechanisms are yet missing to stimulate these new solutions. As a consequence, new mechanisms will most probably be put in place in the years to come. It is currently not foreseeable whether market-based or regulatory concepts will prevail. Given the characteristics of reactive power, regulatory concepts are yet

[11] Switzerland has a very high share of hydropower generation, including numerous small-scale plants. Further studies may investigate to what extent the Swiss approach is thus transferable to other countries in Europe and beyond.

advantageous to provide adequate incentives while also being easier to implement. A good example of a regulation-based remuneration mechanism for reactive power is implemented in Switzerland.

Besides the provision of reactive power from renewable units, a further ancillary service – the provision of black start capability for re-establishment of supply – is challenged by the transformation of the energy system. Research is also ongoing here to identify new concepts and solutions.

12.4 Challenges in Prosumer Integration and Network Tariffication

Key challenges: with the energy system's decarbonisation, distributed energy generation is gaining importance (cf. Sect. 12.1). Given the drastic cost decreases and a high acceptance, rooftop solar PV is already gaining importance even in countries like Germany with only moderate irradiation levels. This raises technical challenges regarding the coordinated operation and control of literally millions of producers. Related to the operational challenges are the challenges of billing, e.g. concerning balancing energy. Important regulation and market design challenges arise, notably concerning efficient grid usage and cost allocation. Besides congestion management (cf. Sect. 12.3.2), the incentives for prosumers are primordial. These are notably dependent on the grid tariffication schemes – in a future sustainable electricity system, network charges are a key element when it comes to investment in and operation of distributed generation and storage resources.

Technical solutions: for the operation and control of multiple devices, digitisation provides numerous opportunities. Digitisation is not new to the energy industry, and both the transmission grids and the conventional generation units have been increasingly equipped with digital devices for measurement and control over the last four to five decades. Also, the operational procedures have been more automated, especially in normal operation but also for the handling of disturbances or unit starts. A state-of-the-art coal-fired power plant of 1000 MW needs fewer operational staff than a 300 MW unit built 50 years ago. But with cheap sensors, large computing power and fast communication, also the operation of lower voltage grids and corresponding devices may be effectively digitalised.

One application case is so-called smart meters which increasingly replace conventional electro-mechanical meters. These new electronic meters enable automated meter reading also for households and other small customers, whereas they used to be in place only for consumers with 100,000 kWh or more of annual consumption. Going beyond meter reading, advanced meter infrastructures even allow the grid operator or suppliers to remotely control devices in the household. With automated, frequent meter reading, the usage of the grid capacity may be monitored regularly. This provides the basis for modified regulatory and market-based concepts. It also provides opportunities for new business concepts independent from the infrastructure and thus independent from classic utilities and network operators. It has yet to be mentioned that the pace of smart meter roll-out diverges considerably among European countries and is very slow in some countries.

Market-based and regulatory concepts: since balancing demand and supply as well as congestion management are key concerns in future sustainable electricity systems, it is of utmost importance that also small distributed resources efficiently contribute to these tasks. At the same time, grid operators need to recover their grid costs. Otherwise, they will not survive in the longer run. And in this case, the electricity system will not be economically sustainable, even if it is ecologically. Hence, there are two competing objectives for grid tariffication in future sustainable electricity systems:

1. Efficient price signals to grid users.
2. Cost recovery for grid operators.

In Sect. 6.1.4, the issue of adequate network charges to fulfil these general requirements has been discussed. It has been shown that the usual recipe of marginal cost pricing does not fulfil the second condition as marginal costs in a network are below the average cost – this is why electricity grids are considered a so-called monopolistic bottleneck and need to be regulated. The second-best economic solution (after marginal cost pricing as first best) has been identified in Sect. 6.1.4.1 as the so-called Ramsey pricing. Ramsey pricing describes the optimal way to meet the aforementioned two competing objectives.

Yet, it has to be carefully adapted to apply to the case of prosumers. Notably, not only consumers should be viewed as grid users but also producers as they may also be a major driver for grid expansion. At the same time, it is essential to carefully design the different elements of the grid tariffs, notably base, capacity and energy charges (cf. Sect. 6.1.4.2). Thomsen and Weber (2021) highlight that energy charges applied asymmetrically to generation and consumption may provide substantial incentives for self-consumption and installation of storage which may not be used efficiently in a system-oriented way. In order to better align tariff schemes with system benefits, Linvill et al. (2017) present different rate design principles.

Political and societal aspects: the preceding considerations suggest that the integration of prosumers will only be successful if another condition is met: adequate incentives for investment and operation of distributed generation and flexibilities. This is important for the economic efficiency of prosumer investments in a system perspective, yet this includes also fairness aspects. E.g. renters have frequently fewer opportunities to install and operate a rooftop solar system than owner-occupiers. They may then consider it unfair if the (in general wealthier) owner-occupiers can more easily reap additional benefits they are excluded from.

Another issue that may be raised is non-discrimination. Again, this could be invoked as a norm derived from fairness considerations. Yet, it may also be seen as a consequence of the information asymmetry between the grid operators and both the regulator and the customers (cf. Sect. 6.1): to limit the grid operators' strategic behaviour, the regulator requests them to apply a unique and transparent grid tariff.

Modelling: the heterogeneity of grid users in a future smart grid with multiple generation, storage and consumption opportunities makes it challenging to identify "efficient and cost-reflective price signals". This is particularly true in the presence of prosumers who both produce and consume electricity and may have some flexibility in their generation and consumption pattern, e.g. through the use of storage possibilities. This point has been discussed on an exemplary low-voltage grid in Sect. 6.1.4.4, and some (short- to mid-term) implications have been derived. For a comprehensive answer, a modelling approach has to be developed that combines Ramsey pricing with an endogenous computation of the market prices and a consistent derivation of the price elasticity for the different grid users. Given the temporal variability of feed-in, load and resulting prices, such an analysis will also encompass a detailed bottom-up market model (cf. Chap. 7) run for many hours if not an entire year.

Key insights: overall, the definition of grid tariffs for prosumers and other grid users that meet all the objectives identified above turns out to be puzzling. A perfect fit to all four requirements (efficiency, cost recovery, incentive adequacy for prosumers and non-discrimination) is unlikely. Hence, a context-specific compromise is likely to be the best achievable result.

An additional dimension of complexity is added when non-grid-related levies and surcharges apply to retail electricity consumption, e.g. levies for renewable financing ("renewable levy"). Such levies are typically defined on a per energy basis. This induces similar distortions like an energy-based grid fee. These distortions are significant in the case of prosumers with internal storage or other flexibilities. The levies are in general imposed on consumed electricity and not refunded for electricity fed into the grid. I.e. there is no longer a unique energy price for electricity; rather, the procurement price exceeds the sales price for the prosumer. Then, the most cost-efficient use of flexibilities is to use them to increase the degree of self-consumption. This simultaneously prevents and partly contradicts an efficient use of the flexibilities for grid purposes (cf. also Thomsen and Weber 2021).

One may conclude by referring back to the beginning of this chapter: a coherent and comprehensive pricing of CO_2 instead of these specific levies may alleviate many of the identified incentive problems, as it reduces the price gap between purchase and sales prices. Simultaneously, time-variable wholesale market prices will reflect the actual scarcity of carbon-free electricity generation. This then provides more appropriate signals to prosumers both for actual operation and for investments.

12.5 Prospects for Sustainable Energy Systems

The mitigation of climate change is one of the top priorities in global politics. The risks of inaction have become apparent in the last years through extreme weather events like unprecedented heat waves, long-lasting droughts or heavy rainfalls and floodings. Single events might be attributed to local weather variations, but taken together and combined with the observations about rising global average temperatures and model simulations, the scientific evidence about climate change is now overwhelming. And, strong political forces and an increasing proportion of economic players are pushing towards a deep decarbonisation almost everywhere on the globe – and particularly in industrialised countries in Europe and North America.

Such a combat of climate change requires an in-depth transformation of the worldwide energy systems – which have to rely more on electricity and on decarbonised power systems. The preceding chapters of this book have given an introduction to the technological and economic solutions available to achieve this goal. As a major transformation process, this development yet has to involve multiple stakeholders and to combine innovations across many domains. This offers bright opportunities for ambitious, well-trained students with interdisciplinary skills as well as for innovative professionals of all ages with open minds and a clear view on the opportunities and challenges of the transition to come.

The fields of action are multiple and intertwined, including notably:

- Renewable technologies in general, their components and their embedding in the overarching system context.
- Storage solutions including batteries, demand-side solutions and hydrogen.
- Decentralised energy systems and digital energy management strategies.
- New concepts for sustainable mobility.
- Green buildings and concepts for refurbishment and decarbonisation of the existing building stock.
- Innovative business models in combination with digitalisation.
- Regulatory concepts, market designs and green financing mechanisms to foster the energy transition.
- Technologies for full climate neutrality including direct air capture (DAC) for carbon dioxide or carbon capture and sequestration (CCS) from bioenergy use.

In a broader perspective, the aforementioned action fields and the energy system transformation as a whole are embedded in the global striving for **sustainable development**. The UN has defined 17 **Sustainable Development Goals** (SDGs), cf. (UN 2015). Climate action is one of these goals (SDG 13), and affordable and clean energy is another (SDG 7). Also, the energy transition is related to two further goals, namely sustainable cities and communities (SDG 11), which also includes sustainable transport, and responsible consumption and production (SDG 12).

At the same time, the transition towards low-carbon energy systems may sometimes also require a balanced approach keeping an eye on other SDGs. E.g. in the case of bioenergy, the goals no hunger (SDG 2) and life on land (SDG 15) are very relevant and potentially competing objectives. But also, co-benefits with other goals may arise from a decarbonisation of the energy system. These include notably a better air quality by a reduction of conventional pollutants such as SO_2 and NO_x (see Sect. 2.3.2). Also, relying on domestic renewable resources will reduce the dependence on energy imports – which may be a strategic geopolitical benefit in times of global tensions. But in general, limiting climate change will rather contribute to foster global cooperation – since this is a truly global issue, it can only be addressed successfully through joint efforts.

We hope that this book contributes to build a common ground of knowledge for these undertakings.

12.6 Further Reading

Möst, D., Schreiber, S., Herbst, A., Jakob, M., Martino, A., & Poganietz, W. (Eds.) (2021). The Future European Energy System—Renewable Energy, Flexibility Options and Technological Progress. Cham: Springer.

This collective volume investigates the transition towards a low-carbon energy system in Europe focusing on the provision of flexibility and the role of technological advances. The industry, residential, tertiary and transport sector are covered along with the heating and electricity sector. This enables a cross-sectoral analysis and an integrated assessment of flexibility requirements in an energy system with high shares of renewable energies. The contributions furthermore examine the impacts of available technologies and the expansion of renewables on the energy system and climate change mitigation.

12.7 Self-check of Knowledge

Self-check of Knowledge

1. What are the three Ds and how do they impact the transformation of the energy system?
2. What are technical solutions to decarbonize the energy system?
3. What are regulatory and market-based concepts for decarbonisation?
4. How do intermittent renewables affect the balancing of production and demand and what is a cold dark calm?
5. What are solutions coping with intermittent renewable generation in a system with high shares of renewable resources?
6. Explain the NIMBY phenomenon.
7. Explain the challenges resulting from low inertia system and how this challenge can be addressed.
8. Why are grid extensions and reinforcement needs in transmission and distribution networks required?
9. Explain why a congestion-free network is not aimed at and which relationship exists between network extension and congestion management.

10. Why is reactive power provision a topic with increasing relevance and how are corresponding challenges addressed?
11. Is a higher energy autonomy attractive from a household's perspective and if so, why?
12. What is the impact of a higher energy autonomy of prosumers on the recovery of network costs?

References

Angerer, G., Buchholz, P., Gutzmer, J., Hagelüken, C., Herzig, P., Littke, R., Thauer, R., & Wellmer, F. (2016). *Rohstoffe für die Energieversorgung der Zukunft: Geologie – Märkte – Umwelteinflüsse*. Schriftenreihe Energiesysteme der Zukunft. Available at: https://energiesysteme-zukunft.de/fileadmin/user_upload/Publikationen/PDFs/ESYS_Analyse_Rohstoffe_fuer_die_Energieversorgung.pdf [Accessed May 26, 2022].

Ball, M., & Wietschel, M. (2010). *The hydrogen economy: Opportunities and challenges*. Cambridge University Press.

Bucksteeg, M., Niesen, L., & Weber, C. (2016). Impacts of dynamic probabilistic reserve sizing techniques on reserve requirements and system costs. *IEEE Transactions on Sustainable Energy*, 1408–1420.

Dallinger, B., Auer, H., & Lettner, G. (2016). Impact of harmonised common balancing capacity procurement in selected Central European electricity balancing markets. *Applied Energy, 222*, 351–368.

Di Silvestre, M. L., Favuzza, S., Riva Sanseverino, E., & Zizzo, G. (2018). How decarbonization, digitalization and decentralization are changing key power infrastructures. *Renewable and Sustainable Energy Reviews, 93*, 483–498.

EC. (2017a). *Commission Regulation (EU) 2017/2195 establishing a guideline on electricity balancing*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2017.312.01.0006.01.ENG&toc=OJ:L:2017:312:TOC [Accessed January 11, 2021].

EC. (2017b). *Study on the review of the list of critical raw materials*. European Commission. Available at: https://op.europa.eu/s/oNqH [Accessed June 16, 2022].

EY. (2019). *Decarbonization, digitization and decentralization are accelerating the countdown to a new energy world faster than expected*. Available at: https://www.ey.com/en_gl/news/2019/07/decarbonization-digitization-and-decentralization-are-accelerating-the-countdown-to-a-new-energy-world-faster-than-expected [Accessed February 7, 2020].

Felling, T., Leisen, R., Podewski, C., & Weber, C. (2019). Evaluation of risks for electricity generation companies through reconfiguration of bidding zones in extended Central Western Europe. *The Energy Journal, 40*(SI), 81–104.

Fraunholz, C., Bublitz, A., Keles, D., & Fichtner, W. (2021a). Impact of electricity market designs on investments in flexibility options. In D. Möst, S. Schreiber, A. Herbst, M. Jakob, A. Martino, & W. Poganietz (Eds.), *The future European energy system—Renewable energy, flexibility options and technological progress* (pp. 199–218). Springer.

Fraunholz, C., Hladik, D., Keles, D., Möst, D., & Fichtner, W. (2021b). On the long-term efficiency of market splitting in Germany. *Energy Policy*, art. 111833.

Fulli, G., Masera, M., Spisto, A., & Vitiello, S. (2019). A change is coming: How regulation and innovation are reshaping the European Union's electricity markets. *IEEE Power and Energy Magazine, 17*(1), 53–66.

Giurco, D., Dominish, E., Florin, N., Watari, T., & McLellan, B. (2019). Requirements for minerals and metals for 100% renewable scenarios. In S. Teske (Ed.), *Achieving the Paris climate agreement goals* (pp. 437–457). Springer.

Gunkel, D. (2020). *Einflussfaktoren auf das Übertragungsnetz im Jahr 2030 für Deutschland - Eine technoökonomische Analyse der Wechselwirkungen auf den Umfang des Netzausbaus, die Systemkosten und die Integration erneuerbarer Energien* (Vol. 17). Series of the Chair of Energy Economics. Available at: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-716026 [Accessed May 26, 2022].

Hinz, F. (2017). *Voltage stability and reactive power provision in a decentralizing energy system* (Vol. 12). Series of the Chair of Energy Economics and Dissertation: Technische Universität Dresden. Available at: http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-229585 [Accessed May 27, 2022].

Hinz, F., & Möst, D. (2018). Techno-economic evaluation of 110 kV grid reactive power support for the transmission grid. *IEEE Transactions on Power Systems, 33*, 4809–4818.

IPCC. (2014). *Climate change 2014: Synthesis report*. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Available at: https://www.ipcc.ch/site/assets/uploads/2018/02/SYR_AR5_FINAL_full.pdf [Accessed February 8, 2020].

Joint Research Centre, Institute for Energy and Transport, Tercero, E., Bryson, R., Chapman, A., et al. (2014). *Critical metals in the path towards the decarbonisation of the EU energy sector: Assessing rare metals as supply-chain bottlenecks in low-carbon energy technologies*. Publications Office. Available at: https://data.europa.eu/doi/10.2790/46338 [Accessed June 16, 2022].

Kunz, F. (2012). *Managing congestion and intermittent renewable generation in liberalized electricity markets* (Vol. 1). Series of the Chair of Energy Economics and Dissertation: Technische Universität Dresden, TU Dresden. Available at: http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-108793 [Accessed May 27, 2022].

Larscheid, P., Taylor, K., van Leeuwen, T., Moser, A., & Hermes, R. (2018). Modelling reactive power demand of distribution grids subjected to renewable energy sources. In *IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*.

Leopoldina, acatech & Union der Akademien der Wissenschaften. (2020). *Zentrale und dezentrale Elemente im Energiesystem: Der richtige Mix für eine stabile und nachhaltige Versorgung* (115 p). Nationale Akademie der Wissenschaften Leopoldina, acatech – Deutsche Akademie der Technikwissenschaften und Union der deutschen Akademien der Wissenschaften. ISBN: 978-3-8047-4060-0.

Linvill, C., & Lazar, C. (2018). Smart non-residential rate design: Aligning rates with system value. *The Electricity Journal, 31*(8), 1–8. https://doi.org/10.1016/j.tej.2018.09.011

Local Energy Consulting. (2020). *Akzeptanz und lokale Teilhabe in der Energiewende. Handlungsempfehlungen für eine umfassende Akzeptanzpolitik*. im Auftrag von Agora Energiewende. Agora Energiewende. Available at: https://static.agora-energiewende.de/fileadmin/Projekte/2020/2020_07_EE-Akzeptanz/182_A-EW_Akzeptanz-Energiewende_WEB.pdf [Accessed May 27, 2022].

Moss, R., Tzimas, E., Kara, H., Willis, P., & Kooroshy, J. (2011). *Critical metals in strategic energy technologies*. JRC Scientific and Technical Reports. Available at: https://setis.ec.europa.eu/document/download/7b6d43a5-a227-4283-b345-ccb0c3f6a0a5_en?filename=CriticalMetalsinStrategicEnergyTechnologies-def.pdf [Accessed May 27, 2022].

Möst, D., Schreiber, S., Herbst, A., Jakob, M., Martino, A., & Poganietz, W. (Eds.). (2021). *The future European energy system—Renewable energy, flexibility options and technological progress*. Springer.

Müller, T., Gunkel, D., & Möst, D. (2013). How does renewable curtailment influence the need of transmission and storage capacities in Europe? In *IAEE: Proceedings of the 13th European IAEE Conference*.

Neuhoff, K., Hobbs, B., & Newbery, D. (2011). *Congestion management in European power networks: Criteria to assess the available options*. Smart Power Market Project. Available at: https://climatepolicyinitiative.org/wp-content/uploads/2011/12/Congestion-Management-Paper.pdf [Accessed May 26, 2022].

Perlaviciute, G., Schuitema, G., Devine-Wright, P., & Ram, B. (2018). At the heart of a sustainable energy transition: The public acceptability of energy projects. *IEEE Power and Energy Magazine, 16*(1), 49–55.

Perras, S. (2015). *Electricity transmission line planning: Success factors for transmission system operators to reduce public opposition*. Series of the chair of Energy Economics and Dissertation: Technische Universität Dresden. Available at: http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-161770 [Accessed May 26, 2022].

Seljom, P., & Tomasgard, A. (2015). Short-term uncertainty in long-term energy system models—A case study of wind power in Denmark. *Energy Economics, 49*, 157–167.

Thomsen, J., & Weber, C. (2021). Getting prices for prosumers right? Incentivizing investment and operation of small-scale PV-battery storage systems through prices, charges and levies. *Zeitschrift für Energiewirtschaft, 45*, 35–59.

Ulbig, A., Borsche, T. S., & Andersson, G. (2014). Impact of low rotational inertia on power system stability and operation. *IFAC Proceedings Volumes, 47*, 7290–7297.

UN. (2015). *Paris agreement*. United Nations. Available at: https://unfccc.int/sites/default/files/english_paris_agreement.pdf [Accessed February 15, 2021].

Viebahn, P., et al. (2015). Assessing the need for critical minerals to shift the German energy system towards a high proportion of renewables. *Renewable and Sustainable Energy Reviews, 49*, 655–671.

Wellmer, F., & Dalheimer, M. (2012). The feedback control cycle as regulator of past and future mineral supply. *Mineralium Deposita, 47*, 713–729.

Wüstenhagen, R., Wolsink, M., & Bürer, M. J. (2007). Social acceptance of renewable energy innovation: An introduction to the concept. *Energy Policy, 35*, 2683–2691.

Zipf, M. (2020). *Economics of ancillary services for electricity—Managing uncertain power generation and grid operation in the distribution network* (Vol. 20). Series of the Chair of Energy Economics. Available at: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-751856 [Accessed May 27, 2022].

Index

C. Weber et al., *Economics of Power Systems*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-97770-2