# Chapter 7
# Electricity Generation and Wholesale Markets

**Carlos Batlle**

*The market is not an invention of capitalism. It has existed for centuries. It is an invention of civilization.*—Mikhail Gorbachev

The generalized worldwide process of restructuring and liberalization of the electric power sector has primarily concerned the generation activity. Although liberalization and restructuring of the power industry has not been universally adopted, it is undoubtedly the prevalent framework in most countries.

In *The Wealth of Nations*, [31] contended that a free market economy is more productive and more beneficial to society. Smith noted that individuals, in the pursuit of their individual self-interests, interact on the market place guided by an "invisible hand" that inadvertently leads them to reach in the end socially optimum results. In a way, the market price acts as this "invisible hand" that drives the activity and ensures the efficient allocation of resources.

It is well known that, wherever possible, competition is beneficial because it places pressure on individuals to act more efficiently. In the context of electricity systems, this competition is not only expected to make suppliers to reduce costs but also help to naturally send sound economic signals to consumers. That is, consumers are made aware of the costs incurred to meet their demands.

However, the introduction of a competitive framework in electricity systems is not as straightforward as in other economic activities. The particular characteristics of the underlying commodity and the large diversity of typologies in electricity systems worldwide have led to the implementation of an enormous variety of

The author wants to thank Pablo Rodilla, Michel Rivier, and Ignacio Pérez-Arriaga for their support in the development of this chapter.

C. Batlle (✉)
Institute for Research in Technology, Comillas Pontifical University, Sta. Cruz de Marcenado 26, 28015 Madrid, Spain
e-mail: Carlos.Batlle@iit.upcomillas.es

C. Batlle
MIT Energy Initiative, MIT, Cambridge, MA, USA

C. Batlle
Florence School of Regulation, European University Institute, Florence, Italy

I. J. Pérez-Arriaga (ed.), *Regulation of the Power Sector*, Power Systems,
DOI: 10.1007/978-1-4471-5034-3_7, 

alternative wholesale market designs. The objective of this chapter is to introduce and shed some light on this complex topic.

Whatever the reason that motivated the regulatory change, the success or failure of the new regulatory framework should be judged on the grounds of the overall efficiency achieved for the electric power industry.

## 7.1 The Necessary Steps Toward Competitive Electricity Markets

The key steps in electricity industry liberalization and restructuring include[1]:

(1) Privatization to enhance performance and reduce the ability of the state to use these companies as a mean to achieve costly political agendas.
(2) Unbundling: separation of the electricity businesses that can be conducted competitively (generation and retail) from the natural monopolies (transmission and distribution), which must be regulated.
(3) Horizontal restructuring to ensure competition (otherwise market power may put in danger the whole scheme, see Sect. 7.5).
(4) Designation of an Independent System Operator (ISO). This ISO would be responsible to maintain network stability and should ensure open entry to the wholesale market and full access to the transmission network.
(5) Establishment of a wholesale market where generators compete to supply electricity on an hourly, daily, weekly, monthly, and annual basis, or longer. This wholesale market also has to suitably integrate market-based mechanisms aimed to acquire operational reserves services (see Sect. 7.4).
(6) Unbundling of retail tariffs and rules to enable access to the distribution networks in order to promote competition at retail level. Open access to the retail market, so that all consumers can choose their supplier of electricity.[2]

This chapter focuses on step two and, in particular, step five.[3] The study of the wholesale market is essential to understand the workings of the electricity industry in a regulatory environment open to competition.

These steps are necessary but not sufficient to ensure an efficient electricity market. In certain circumstances markets may be unable to deliver optimum resource allocation, and this must be taken into account in electricity market design. Under such circumstances, normally termed market failures, regulatory intervention is required to maximize social welfare. Intervention mechanisms must be carefully analyzed in the wholesale market design phase, with a view to

---

[1] See also the textbook conditions proposed by Joskow [18].

[2] This is a necessary condition for the full liberalization, but it is not required for the implementation of a market mechanism at generation level.

[3] The last step, the liberalization of the retail business, is analyzed in Chap. 9.

distorting market operation as little as possible by regulation and verifying that the intended objectives are achieved.

### 7.1.1 The Multiple Dimensions of the Generation Business

The design of a stable regulatory framework for the efficient and reliable delivery of electric power at present and in the future is one of the major concerns of electricity market regulation policies. The type of regulation chosen for electricity production must wed sound economic criteria with the technological aspects of electricity generation, in a format that is compatible with the industrial structure and the legal context prevailing in each country. It must also cover all the considered timescales, from investment decisions made several years before plants come on stream to decisions made in real time to select the unit that must respond to an imminent change in demand.

Among many other things, the regulatory framework identifies the decision maker in each case. This is one of the key factors that differentiates the regulatory designs in place in different electric power systems.

In the present discussion, the generation activity is not addressed as a whole, but decomposed into stages in which the discriminating variable is the distance to real time. This leads to a division of three major phases in the electric power generation business:

- The first timescale involves decisions concerning the installation of additional generation capacity to replace obsolete facilities and cover demand growth. Depending on the type of regulatory framework, decisions in this timescale may be centralized in a government agency, incumbent upon vertically integrated utilities under the supervision of the regulatory authority, or left to the initiative of private investors.
- The second timescale revolves around the scheduling of existing generation capacity (a short- to medium-term issue, i.e., generator maintenance management, fuel supply contracts, hydro reservoir management, start-up schedules, and close to real-time economic dispatch of generation units). Here again, decisions may be made centrally or left to the initiative of producers and consumers, either in some organized fashion or bilaterally.
- The third timescale involves generators' short-term responses oriented to keep the generation-load balance (a short-term issue, i.e., the use by the System Operator of the various types of so-called operating reserves). In this phase of the generation business, maintaining security is the overriding concern and System Operators everywhere are in charge of centrally managing the system, which must necessarily also include the grid. In most competitive markets, the boundary between this phase and the preceding one is known as "gate closure". Thus, the timescale involved by this third phase includes all decisions between gate closure and real time, when electric power actually flows to the end user.

Regulation must strike a balance among three traditional objectives of electricity supply: economic efficiency, reliability of supply, and environmental impact. Each of these objectives can also logically be decomposed into the three timescales. All are interrelated with one another and across timescales.

A few general comments are in order here. In some countries the power industry has been restructured and liberalized fairly abruptly. This was typically the case in government-owned electricity utilities, such as in the UK or Argentina. By contrast, where utilities were primarily investor-owned, such as in the USA or Spain, the process was more gradual. Globally viewed, the traditional regulation paradigm has been transformed very slowly, step by step, starting from the longer timescale (free entry for small independent generation facilities) and evolving toward real time and ancillary service markets.

While the vast majority of the technical literature on electricity systems describes regulatory frameworks for markets fully open to competition in a rather matter of fact fashion, many of the electricity market designs in place around the world are somewhere in between a traditional centralized and a fully competitive model. As any number of experiences show, deregulation has not only been incomplete but in certain cases regressive, for some recently implemented regulatory mechanisms indisputably constitute a partial return to the centralized paradigm.

## 7.2 Expansion of Generation Capacity

As generation investment costs typically account for the largest share of the total cost of electricity, this is the area where the potential benefit from regulatory reform is greatest. Moreover, a country's economy is impacted very differently depending on whether investment in generation is made with public resources or by private investors. Privatization of formerly state-owned generation assets often results in much needed income for the national treasury. This is why regulatory change frequently goes hand-in-hand with change in the ownership and control of electric utilities, via privatization or corporatization[4] and the entry of private investors. However, changes in ownership must be distinguished from regulatory change per se.

### *7.2.1 A Gradual Process: From Independent Power Production to Unrestricted Entry and Access*

The gradual exposure of traditional vertically integrated utilities to competition was initiated in response to concerns about energy security and independence from

[4] Corporatization refers to the transformation of state assets or agencies into state-owned corporations in order to introduce corporate management techniques to their administration. Corporatization is sometimes a precursor to partial or full privatization (see more in Wikipedia).

foreign energy sources as a result of the 1970s oil crisis. This and other concerns led the United States and other developed countries to enact rules that favored the development of renewable energy and cogeneration. The most significant was the 1978 Public Utilities Regulatory Policies Act (PURPA) in the USA. PURPA enabled a certain kind of non-utility electric power producers (so-called "qualifying facilities", essentially small renewable generators and cogeneration) to become alternative electricity suppliers by requiring electric utilities to buy power from these facilities at the "avoided cost" rate, i.e., the cost the electric utility would incur were it to generate or purchase from another source.

The precise definition of "avoided cost" was left to the discretion of each individual state. The resulting spectacular development of qualifying facilities in some states led to the need to establish mechanisms for the competitive selection of more efficient producers and establish conditions to govern their funding. The competition mechanism, which began to be used for this purpose in Maine in 1984, soon spread to other states.

Although the volume of generation from qualifying facilities was very small compared to generation from vertically integrated utilities, these new participants introduced a completely new scenario in the electricity industry. What had been a closed shop for many years, opened its doors to independent power production, albeit with very specific characteristics.

Other factors favored the advent of independent power production in the USA and elsewhere, which was not restricted to renewables or cogeneration only. Electricity prices under traditional regulation had consistently declined for many years, aided by technological developments and the fact that economies of scale had not yet been fully exploited. That also changed in the late 1970s and early 1980s, leading to more stringent regulatory reviews of utility costs. Delays in the regulatory process ("regulatory lags") and the occasional recognition by regulatory authorities of only part of companies' generation investment costs ("prudent and justified expenditures"), which were examined retroactively, raised vertically integrated utilities' risk, encouraging them to seek alternatives to cover the investment needed in generation.

When new investment was needed, these utilities began to buy the production required from third parties under a variety of instruments, subject to the approval of the regulatory authorities. The costs involved were then passed on to consumers with no risk to the incumbent utility. The conditions governing these arrangements were not established by PURPA or imposed by regulators, but were the result of inter-party negotiations and laid down in what were called "power purchase agreements" (PPA). Trading between the independent power producers (IPP) and any other parties except the vertically integrated utility were banned or severely limited.

Such independent power production became standard practice in the USA. Ad hoc methods to select suppliers were replaced by extremely detailed "competitive bidding" rules to determine the lowest cost generation portfolio. Finally, the 1992

Electricity Act allowed independent power producers[5] to trade freely in the power system and sell wholesale power anywhere to any vertically integrated utility or distribution company. This entailed open access to the transmission network. Selling to end consumers or "retail competition" was not allowed, however. This area of the market was liberalized years later at the individual state level. Other countries, such as Chile, UK, Norway, and Argentina got off to a later start but carried liberalization through to completion much more swiftly.

Similar transformations in developing countries were driven by a number of factors. After years of huge investment projects and subsidized rates, which were often insufficient to recover costs, state-owned electricity companies in many countries lacked the resources to continue investing. As a result, they resorted to independent producers, private companies keen on entering the generation business. Plant construction and operation was often tendered. The agreement stemming from the award was usually either a build-operate-transfer (BOT) or a build-operate-own (BOO) arrangement. Under the first, the plant had to be transferred to the incumbent company after a given term, while under the second the investor could retain ownership of the facility. Several electric power systems are presently organized around these arrangements today, Mexico being a paradigmatic example [19].

The massive entry of independent power producers in the electricity industry in the 1990s and the early twenty-first century was favored by an environment of declining interest rates, controlled inflation, liberalization of capital movements, and development of financial markets. The present prevalence of private investors in the electricity industry, hitherto mostly controlled by State companies in many countries, has brought fundamental change to the perception of risk and investment priorities.

#### 7.2.1.1 Power Purchase Agreement Contracts[6]

A power purchase agreement (PPA) affords potential investors and financial institutions a sound legal guarantee. The lack of adequate legal safeguards increases risk, resulting in higher capital costs.

In traditionally regulated systems, generators' remuneration is governed by a set of laws. Generally known as a "regulatory contract", this arrangement is based on the Government's implicit commitment not to change the rules if the change is detrimental to business. A regulatory guarantee is by definition skewed, however, since regulators have their own objectives that may not necessarily include protection for producers. In the present context of frequent regulatory change the world over, international financial institutions do not generally consider this so-called "regulatory contract" a sufficient guarantee to grant funding under the most favorable conditions. The regulatory contract is based on a relationship of "trust"

[5] They were called "exempt wholesale generators" in the Act.

[6] This section is based on [8].

between the regulator and business agents, which does not typically exist with new entrants (especially if they are foreign companies).

But contracts are not even a full guarantee that problems will not arise. Political interference can nullify legal guarantees, albeit often with great difficulty. For instance, when short-term energy prices decline steadily for whatever reason, regulatory authorities and governments in particular are often tempted to renegotiate any existing long-term contracts, initially regarded as advantageous, but overly expensive for consumers in the new scenario.

PPAs are often only the most visible part of a much more complex network of agreements among manufacturers, equipment suppliers, maintenance contractors, fuel suppliers, financial institutions, insurance companies, consumers, and a long list of entities involved in the project. Ideally, each type of risk should be allocated to the parties best able to control it. The main objective of a PPA is inter-party risk sharing. The risks involved in generating electric power include variations in fuel prices, costs of labor and materials, unexpected plant failures, the uncertainty as to whether a contract counterparty will be found, and regulatory amendments.

At the very least, a PPA should specify a price per MW and a price per MWh. The former may consist of an annual charge to remunerate the generation plant for its nominal capacity plus its annual fixed operation and maintenance costs and is usually indexed to its availability record. The price per MWh is typically the cost of fuel times a plant-specific efficiency factor, plus a variable operation and maintenance item. The cost of fuel may be indexed to an international market reference price and the other costs to the consumer price index (CPI) or similar. The contract must include the definition of *force majeure*, i.e., situations beyond the control of the parties to excuse compliance with certain stipulations.

Because of their very nature, PPAs are usually very long-term contracts (at least ten or fifteen years), typically closely linked to the physical aspects of the project and inflexible in structure. They can be viewed as an extension of the traditional regulatory scheme, suitable for cases where the objective is not a drastic transformation of the existing regulatory framework, but only to enable private investors to enter the business to meet the need for new capacity.

### 7.2.2 Liberalization of Generation Investment: Opening the Door to Wholesale Markets

Chile made major progress toward the liberalization of generation expansion in the early 1980s. Prior to the Chilean initiative, the only experiences were the qualifying facilities in the USA and elsewhere, the incipient entry of independent generators as the result of auctions or negotiations with the incumbent company and the regulator, and PPAs. In 1981 a group of Chilean economists, inspired by the Chicago School of Economics, proposed and implemented radical reform of the Chilean power industry, with a view to introducing economic rationality and attracting foreign investment.

In essence, the reform established freedom of installation of generation facilities and a remuneration scheme based on market prices. The specifics of the Chilean regulation include features such as some manner of government intervention where private investment fails to meet needs and a mechanism for determining market prices outside actual market conditions at any given time. But for the first time in many years, it established an electric system in which private investors could freely install new plants and sell the electricity generated at market prices.

The first principle underlying this revolutionary reform, later adopted by many other countries, was deregulation of generation investment. In these systems, the free entry principle ensures that any investor is allowed to install new generation capacity, subject only to the customary legal obligations in connection with land use or environmental impact. Beyond that, all that is required is a licence or authorization that the authorities are obliged to grant indiscriminately to ensure fair competition among potential agents. Under such arrangements, the central planner is replaced by an unspecified number of decentralized planners. One possible objection that may be raised is that the optimal mix of generation technologies is more difficult to establish under this approach. Experience in other industries has shown, however, that efficient markets in which participants are sent appropriate economic signals typically reach suitable solutions, with less risk of committing the glaring errors often made by central planners. At least the possibility and the seriousness of error making is widely distributed.

In a competitive environment, companies decide to invest only when it makes economic sense, therefore eliminating vested political, industrial, or private interests that may be present in centralized planning. Here also, Adam Smith's "invisible hand" aligns the interests of the agents making decentralized decisions with the interests of overall economic efficiency, as discussed below.

The second basic principle of Chilean reform was to replace cost-of-service remuneration with market prices for the generated electricity.[7] This obviously raised the question of whether this regulatory scheme would be able to attract the necessary investment. For this scheme to work, the expected revenues from electricity sales and (to a lesser extent) ancillary services (see Sect. 7.4) would have to cover investment and operating costs, and provide for a reasonable rate of return on the capital invested.

This liberalized scheme for generation expansion planning was later adopted in a substantial number of electric power systems around the world. England and Wales, New Zealand, Norway, Argentina, and Colombia were among the first to follow suit. Many others took the same route in the late 1990s, including Spain, California, and New South Wales. Free entry to the generation business and market-based pricing were the two features common to all these systems.[8]

---

[7] In the Chilean scheme generators also participated in a capacity market, whose results were similar to the capacity payments later adopted in a number of Latin American countries.

[8] See Batlle et al. [5] for a description of the expansion of electricity systems in Latin America.

***Marginal pricing and investment cost recovery***
The marginal cost of production is a concept that plays a fundamental role in the analysis of competitive markets. As explained in Chap. 2, the use of marginal costs to compute market prices has its justification in microeconomic theory. Specifically, theory has it that on a perfectly competitive market, prices should equal marginal production costs.

In principle, ideally marginal energy pricing is the most suitable remuneration mechanism [32]. As microeconomic theory has shown, the short-term marginal energy price, defined as the production cost of responding to a unit change in energy demand, is the appropriate signal for attracting new investors. If the margin between installed available capacity and peak demand narrows, the price rises, therefore ideally providing the incentive for the entry of new investors.

However, the translation from costs to prices is not immediate, since the cost structure of a generator is far more complicated than just the addition of fixed investment costs plus variable costs proportional to production: plant operation is subject to a large number of technical constraints, limited energy plants (hydro reservoirs, and also storage facilities, as for instance batteries or pumping units) are difficult to manage, certain types of generation are highly unpredictable in the short term, such as run-of-the-river (non-impounded) and wind generation, or the output from CHP plants and finally, demand-side modeling is still only scantly developed.

This does not mean that marginal pricing is not the right choice in the case of electricity generation business. Pérez-Arriaga and Meseguer [27] demonstrate, by investigating the optimal economic signals that generators and consumers must receive in a competitive market under diverse circumstances, that even in the presence of a variety of planning and operation constraints for the generators, the approach is consistent with the goal of a correct regulatory policy: the maximization of global net social benefit. See Chap. 2 for a brief demonstration that, besides leading the system to the maximization of the operation efficiency, marginal prices allow investors to recover their investment costs (obviously only for those cases in which the investment is economically rational).

## 7.3 Generation Management and Scheduling

### *7.3.1 From Central Dispatch to Wholesale Markets*

This section reviews the schemes in place for efficient generation unit dispatching, i.e., efficient demand coverage. Each approach to generation investment is characterized by a specific scheme for generator management and remuneration. These schemes range from the centralized optimization of generation dispatch under traditional regulation to arrangements provided for in the latest regulatory systems, in which no market price is calculated, and plant operation is the result of bilateral agreements based on supply and demand.

In this section, “generation management and scheduling” covers the following activities and decision making:

- long-term decisions (years) on plant transformation and overhauling, long-term fuel purchases and power sale contracts, plant maintenance programming, multi-annual reservoir management and nuclear fuel cycle management;
- medium-term (months to days) decisions concerning fuel, annual reservoir and pumping management, futures contracts for fuel and power.

Short-term (same-day) decisions to connect steam or hydroelectric generating units, overnight shut down management, hourly generator scheduling and operating reserves; decision making for shorter timescales, incumbent upon the System Operator, are considered in Sect. 7.5.

Regardless of whether this suite of decisions is made by a vertically integrated utility under traditional regulation or decentralized market agents, the reasonable underlying assumption is that, when shorter term decisions have to be made, the longer term decisions are already in place. For instance, when weekly decisions are made whether to connect or not some steam plants based on expected system behavior, the information on generating capacity, maintenance or the status of long-term fuel contracts is a given and not subject to change.

#### 7.3.1.1 The Traditional Unit Commitment

For the sake of simplicity, this section focuses only on the so-called unit commitment problem, i.e., the determination of which units should be in operation and which should remain disconnected at any given time, to ensure that demand is met at the lowest possible cost, subject to any existing constraints. Given the highly detailed nature of such decisions, unit commitment is only applied within a timescale of about one day to one week before real time, depending on system characteristics.

More specifically, the unit commitment problem consists of supplying the estimated demand profile in a one day or longer horizon at the lowest cost, given each power plant’s technical characteristics and cost functions. What has to be determined is which generators should be simultaneously connected to the system at any given time, i.e., when each should start up and shut down, and the distribution of total production among the units connected to the system in each time interval, usually hours or half hours.

When working so close to real time, system details are very important, and account must be taken of aspects such as steam plant generating unit start-up and shut down timing and costs, the hydrological restrictions in place in river basins, stations in tandem arrangement, demand chronology profiles, and the generating capacity to be held in reserve to respond immediately to fortuitous equipment failure. Other considerations weighing in these decisions include long-term hydroelectric management, system reserve capacity requirements, and network constraints that may render an economically efficient solution unfeasible.

Roughly speaking, assuming that demand is inelastic and that the marginal utility of demand is constant and hereafter referred to as the cost of the non-served energy, $C_{\mathrm{nse}}$, [€/MWh], the problem to be solved can be expressed as the minimization of the cost of generation plus the total cost associated with the non-served energy nse, [MWh]:

$$\min(C_G + C_{\mathrm{nse}} \cdot \mathrm{nse})$$

where $C_G$, [€/MWh], is generation cost (obviously depending on the amount of energy generated).

For instance, an extremely simplified version of this minimization problem may be expressed as:

$$\min \sum_{i,t} \left(C_{i,t}^{\mathrm{start-up}} + C_i^{\mathrm{fuel}} \cdot g_{i,t}\right) + \sum_t C_t^{\mathrm{nse}} \cdot \mathrm{nse}_t$$

$$\text{subject to} \qquad \sum_i g_{i,t} + \mathrm{nse}_t = \mathrm{d}_t$$

$$\text{Operational constraints} \begin{cases} g_{i,t} \geq \underline{g}_i \\ g_{i,t} \leq \bar{g}_i \\ \sum_t g_{i,t} = \bar{e}_i \\ \text{etc.} \end{cases} \quad \forall t, i$$

where for every unit $i$ and time interval $t$, $C_{i,t}^{\mathrm{start-up}}$, [€], denotes the start-up cost and $C_i^{\mathrm{fuel}}$, [€/MWh], the fuel cost of each generating unit producing a given amount of energy, $g_{i,t}$ [MWh]; $\mathrm{d}_t$ is the demand, $\underline{g}_i$ and $\bar{g}_i$ are the minimum and maximum output of the unit and $\bar{e}_i$ represents the maximum energy available for the whole time horizon (e.g., the day) for the case of an energy limited plant (e.g., a hydro unit).

Under traditional regulation, centralized optimization is based on the cost structures of each generating unit, which are either supervised by the regulator or included in the PPAs. Where generating plants are remunerated under cost-of-service arrangements, each plant receives its annual fixed costs plus the variable costs deriving from the time it is actually operating. Plants with PPAs are paid according to the terms of their agreements and their production records.

#### 7.3.1.2 The First Step: Wholesale "Markets" Based on Audited Costs

The electric power system reform implemented in Chile changed the former paradigm. Investment decisions are no longer incumbent upon the regulator and generators are remunerated on the grounds of system marginal cost. Fixed remuneration no longer exists.

The establishment of fair, transparent, and effective access to the transmission network and generating unit scheduling for all players was one of the primary hurdles to introducing competition in this activity. The two new entities created in the Latin American context at the start of the deregulation process for this purpose are defined below.[9]

- The System Operator (SO), which must be independent of generators and marketers, is responsible for managing the transmission network, essential to ensure effective competition, and the security functions, which in general also involve the generation plants.[10]
- The Market Operator (MO) is responsible for operation decisions that are based only on the economic data provided by generating units (the audited generation costs). In all the Latin American countries that adopted the Chilean model (all the countries that deregulated the industry, with the possible exception of Argentina, whose design is slightly different), this new institution, often called a market agent committee, consists of market agents' representatives (normally generators, retailers, distributors,[11] large consumers, the SO, and the regulator). It is responsible for operating the system and obtaining the unit commitment on the basis of the generators' declared costs/bids and calculating the system marginal cost for each time block (not necessarily hours, in some cases days or even weeks, as for instance in the Brazilian market, whose MO calculates weekly marginal prices).

This scheduling and operating model does not mean, however, that generating units are free to bid any "opportunity cost" whatsoever at which they are willing to generate electricity. The MO calculates the optimal schedule in accordance with the criteria described in the preceding section, but introducing two major changes: the cost structure of the conventional thermal plants assumed in the minimization problem is not the same as the structure laid down in long-term contracts; rather, they are what are termed as "audited costs". Moreover, hydro plants declare their inflows and reservoir levels and it is up to the MO to decide how the plants are to be operated, on the grounds of medium- or long-term optimization criteria (the same criteria that were in place prior to system reform).

As stated, in addition to scheduling, the MO calculates the time-blocks system marginal costs used as the (marginal) prices that remunerate all the generating

[9] In the schemes more open to competition that are presented later, the roles of the SO and MO differ from what is presented here.

[10] For instance, the European Commission [13] defines the tasks of the Independent System Operator stating that 'is responsible for granting and managing third-party access, including the collection of access charges, congestion charges, and payments under the inter-TSO compensation mechanism (…) is also responsible for operating, maintaining, and developing the transmission system. (…) has full responsibility for ensuring the long-term ability of the system to meet reasonable demand through investment planning (…) is responsible for planning, including obtaining the necessary authorizations and for the construction and commissioning of new infrastructure'.

[11] In their role of retailers of the consumers under regulated tariffs.

units in the system. Roughly speaking, these marginal prices are expected to be close to the cost of fuel for the marginal units (the most expensive units) committed in each time block. However, when binary (such as start-up costs), timing (e.g., the optimal operation of limited energy plants, such as hydro plants) or network constraints are taken into consideration, calculating this price is always complex and subject to considerable controversy.

This pricing mechanism, based on minimizing operation on the grounds of units' audited costs constitutes a major drawback. Since generating units cannot bid their opportunity costs (which roughly speaking would be the "avoided cost", i.e., the cost of the next more expensive unit), peak-time units would only be dispatched at a higher price than their marginal costs (to enable them to recover their investment costs) in the event of scarcity. In such a scenario, the system marginal price should be the aforementioned cost of non-served energy. But this is not actually the case. The actual value used is always capped by the regulator at a much lower level than would be required to enable the peak-time units to recover their investment costs.

To remedy this shortcoming, the Chilean model added a supplementary mechanism to the system marginal calculation, the so-called "capacity market" (which involves a demand-side obligation to hedge expected peak consumption for two years in advance under an agreement with a generating unit), in an attempt to provide additional remuneration. In Argentina, a capacity payment was devised instead of a capacity market (see Chap. 13 for a brief description of the design initially implemented in Argentina).

As stated, this scheme, based on the audited costs of generating units (and hydro inflows and reservoir levels), is currently implemented in most Latin American countries that reformed their electricity industries. Colombia is possibly the only one where scheduling is based on daily bids (and where generators submit a single bid for the entire day).

The full deregulation of the generation business, however, called for an even bolder move. This was introduced with the compulsory pool model discussed in the following section, in which scheduling follows the same rules, but generators are free to establish their own opportunity costs and therefore to bid the price at which they are willing to be committed.

#### 7.3.1.3 The Second Step: From the Cost-Based Prices to the Bid-Based Prices

***Pioneers***

Ten years after the Chilean reform, in the early 1990s (1991–92), England and Wales and Norway took the system one step farther. These two markets differed in a number of ways, initially with respect to the nature of the underlying electric power systems: while the E&W generating units are almost all steam facilities (which means that unit commitment is subject to many technical constraints), the Norwegian system is based almost entirely on hydro production.

Out of the many aspects of these innovative designs that might be highlighted, the most prominent features, common to both, are considered here. To begin with, the system marginal price was based not on costs but on bids freely submitted by the generating units (freely does not, of course, imply lack of regulator supervision).

*England and Wales*

The market in England and Wales [36] was initially a compulsory day-ahead market (electricity could only be sold on the pool) whose clearing mechanism was the same optimization algorithm that was used prior to system liberalization to dispatch generating units for the following day, at half-hour intervals.[12] Under the former arrangements, the model inputs were the data furnished by the generating units on their technical constraints (such as technical minima and ramps) and production costs (such a start-up costs and fuel costs expressed as a piece-wise linear function). The calculation was based on a classic unit commitment and the system marginal price was computed for each half hour. Network constraints, however, were ignored by the model and subsequently treated in a simplified manner.

***Price calculation in the England and Wales Electricity Pool***[13]

In the England and Wales Electricity Pool all generators were paid the same price every half hour, which was computed *ex ante*. As the unit commitment model that was used was deterministic, an additional term was necessary to account for the possibility that available generation could not meet demand at each half hour of the next day. Therefore:

$$\text{Marginal price } = \text{SMP} \times (1 - \text{LOLP}) + \text{VOLL} \times \text{LOLP}$$

where SMP is the system marginal price computed with the deterministic GOAL model; VOLL is the value of lost load for the consumers, i.e., the value established by the regulator for the system marginal price when there is non-served energy; LOLP (separately estimated with another model) is the probability that available generation cannot meet all the demand during the considered half hour and 1-LOLP is, obviously, the probability that all demand is met. Thus, the equation above computes the expected value of the marginal price for the system for the considered half hour when calculated one day in advance.

SMP is computed as “the highest genset price”. Once the unit commitment has been determined for the next day, for each generator a marginal price is computed at each half hour of the day that allows the generator to recover its operating cost, where any start-up costs and no-load costs (i.e. all nonlinearities in the generator’s cost function) are included. For instance, the start-up costs

[12] It was a heuristic model named GOAL, later replaced by another one that was based on Lagrangian relaxation.

[13] This description was written by Prof. Pérez-Arriaga.

are distributed during all the non off-peak demand hours. Then SMP at each half hour equals the highest of all the generator prices for that half hour. In this way it is made sure that all generators recover their operating costs.

The computation of LOLP happened to be clearly biased and systematically resulted in high values that did not correspond to reality. During the years that this scheme of remuneration was used, there never was a case of non-served energy due to lack of available generation. However, generators obtained substantial income from the VOLL $\times$ LOLP term. Moreover, some companies apparently manipulated this mechanism, by declaring unavailable units at critical times and causing the algorithm to result in artificially high values of LOLP [21]. Many people have classified the VOLL $\times$ LOLP term as a capacity mechanism. Conceptually it is only one necessary element in the correct computation of the ex ante marginal price. However, when misused, as it was the case in the UK, it became a systematic extra remuneration for generators, which was linked to situations of system stress, and its de facto became a capacity instrument.

On top of SMP, consumers had to pay an uplift charge, which covered the generation costs incurred because of transmission constraints (which were ignored by GOAL) and the payments of some of the ancillary services. As it has been seen, the England and Wales pool, although a competitive market at the core, contained a fair amount of regulation.

On March 2001, the Electricity Pool model of England and Wales was replaced by a fully bilateral model, the New Electricity Trading Arrangements (NETA), which in 2005 became BETTA (British Electricity Trading Transmission Arrangements), with the integration of Scotland. The electricity trading arrangements in England and Wales are now managed by ELEXON, the Balancing and Settlement Code Company.

As for example described by Green [15] 'the guiding principle of NETA's (and hence BETTA's) design was that electricity should be treated as much like a "normal" commodity as possible, while still recognizing the physical characteristics of electricity. This means that there is a balancing mechanism run by the system operator, National Grid, to ensure that demand and generation are kept in balance and transmission constraints are respected, but no other market was centrally organized. Instead, most electricity is traded bilaterally (or internally, for integrated firms), with some trading on electronic exchanges to aid transparency'.

*Norway*

The market established in Norway in 1993 (Statnett Marked AS[41]) differed in a number of meaningful ways. First, it was implemented as a voluntary and purely

[14] Currently NordPool (www.nordpool.com), owned by the national grid companies Fingrid, Energinet.dk, Statnett, Svenska Kraftnät, provides a marketplace for trading both physical and financial contracts in Finland, Sweden, Denmark, and Norway.

financial market on which generating units can trade their energy in different timescales, up to the day-ahead spot market, on which financial positions are cleared. As it is later explained in more detail, bids are not fully simple (mainly quantity/price bids only, unrelated to any technical constraint, since, as later described, a number of semi-complex conditions have been progressively allowed).

These two approaches, the E&W and the Norway wholesale markets, represent not only the pioneering experiences but also probably the two opposite poles regarding the extent to which the resulting wholesale spot market design resembles the traditional centralized paradigm.

These spot (or day-ahead) markets based on an organized daily auction (plus in some cases a free bilateral market in which market agents agree to match their production/consumption programs) play a central role in the liberalized context for they determine the expected generation schedule for the day after.

***Electricity wholesale market sequence***

The electricity wholesale market is composed of all the commercial transactions of buying and selling of energy and also other services related to the supply of electricity (the so-called ancillary services, which are essential for this to occur in adequate conditions of security and quality, see Sect. 7.4). These transactions are organized around a sequence of successive markets where first market agents (supply and demand) trade energy, and then, the SO acquires from these agents (mainly from the supply) the above-mentioned ancillary services products related to the supply of electricity in periods closer to real time.

The overall trading timetable covers a number of timescales: months or years before a trade is to be implemented; "gate closure"; real time when the transaction takes place; and post-transaction settlement.

The generation and load parties must notify the SO of their expected physical schedules at real time by market gate closure (one day, one hour, or possibly less before real time). One of the many ways of splitting this sequence of markets and transactions is into the following categories:

- long-term markets,
- day-ahead markets (DAM), and
- intraday plus balancing markets (in the EU) or real-time (in the US) markets.

Additionally, the SO acquires operating reserves (for example, secondary reserves or 10-min spinning reserves, see description on ancillary services later on) in different timescales, sometimes in the long term (e.g., with two years in advance) or once the energy market closes.

When the market structure is competitive and opened enough (what naturally leads to significant levels of volatility and liquidity), a financial long-term market arises. The primary purpose of these long-term markets is to provide hedging mechanisms for producer and consumer entities. But they can also be used by arbitrageurs and speculators, who critically contribute to market liquidity.

These long-term markets function prior to the day-ahead auction. The morning of the day before (D-1) those agents that have not previously committed their supply in a bilateral contract, submit their offers and bids to the market operator, who clears the auction and gets first preliminary schedule results for the day after.

When the MO does not run an extensive nodal auction (i.e., when the transmission constraints in all detail are not considered when clearing the auction, see description below), the SO checks that the schedule resulting from the declared bilateral contracts plus the DAM is feasible. If there is any transmission constraint the SO solves the constraint at the least possible cost and gets the final feasible schedule.

Once the DAM feasible schedule is known, additional mechanisms have to be implemented to allow market agents themselves or the SO fix any deviations from this program that might occur, either because the schedule resulting from the DAM is not feasible for a generating plant[15] (only in the case that simple bids are considered, see explanation in the next section) or because for some reason the plant cannot perform as expected. This rebalance can be done right after the DAM schedule is finished, but others might need to be done later in the day D-1 or a few hours before real time. For instance, the weather forecast might happen to be inaccurate, so wind producers might be interested in selling more or less than what it was estimated at the time the DAM closed on D-1.

In power system operation, once the market is closed, the SO must ensure that supply matches demand in the real time. This task calls for ceasing any further market transaction (at some point in time at which the SO considers there is no time enough for agents to react efficiently) and leaving all the power system control to the SO. Thus, the regulator together with the SO needs to determine the point in time at which this economic trading should be over; a time called "gate closure". Until gate closure the market agents are allowed to balance their positions and correct their deviations without any type of intervention of the SO. After gate closure, the final production schedule is determined for all participants and only the SO can act to adjust any deviation.

- In some cases, (in the EU) subsequent trading is allowed within the day (centralized in the so-called intraday markets organized by the power exchange) and the gate closure is set very few hours ahead of real time. At gate closure, market agents are supposed to have submitted their balancing bids (upwards and downwards) for the so-called balancing market run by the SO. This auction determines the least-cost resources for the SO to fix the potential imbalances.
- In other cases (namely in the US markets), once the day-ahead market clears in the afternoon of the day before, generating units have to submit their bids for both the so-called real-time energy market (similar to the balancing bids in the

[15] These deviations may be due to several reasons. For instance, a thermal unit could have bid expecting to produce in four consecutive hours the day after, but in the final DAM schedule it has been committed in eight hours. If the unit just counts on fuel for the expected four hours, the generator will require a way to readjust the schedule, in order to purchase the committed supply in the four hours in excess.

EU model) and the regulation market (similar to the reserves market in the EU model, providing AGC services, see description of Ancillary Services below in the chapter). These are the tools for the SO to correct the imbalances and to maintain the system stability. Instead of clearing the market agents bids (expressed as just selling and buying quantity-price pairs), the SO runs an optimization tool considering all the technical constraints of the different units (the so-called Security Constrained Economic Dispatch) and calculates prices for each five-minute interval. Gate closure is in some cases set the day before (thus, unless exceptionally justified, generating units cannot trade or modify their bids after the market closes) as at the time of this writing it is the case in PJM or seventy-five minutes before real time, as in NYISO.

Finally, in both models, in most systems the SO acquires in the long-term other ancillary services, namely very short-termed reserves that might be needed to respond to very specific contingencies. These complementary reserves are described later in more detail (Fig. 7.1).

The market implemented in East Australia works in a slightly different way. As taken from the introductory report from the Australian Energy Market Operator, AEMO [2], daily bids are submitted before 12:30 pm on the day before supply is required, and are reflected in pre-dispatch forecasts. Generators may submit rebids up until approximately five minutes prior to dispatch. In doing so, they can change the volume of electricity from what it was in the original offer, but they cannot change the offer price. AEMO issues dispatch instructions to generators at five-

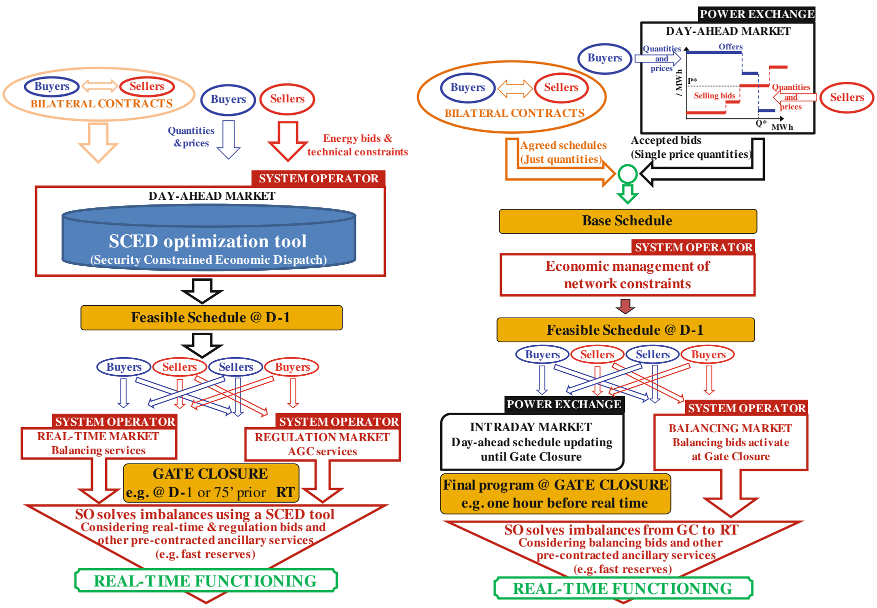


**Fig. 7.1** Market sequence in the US (*left*) and in the EU (*right*) models

minute intervals throughout each day based on the offers that generators have submitted in the bidding process. In this way, there are 288 dispatch intervals every day.

Next we review in larger detail the main design characteristics of the components of these energy markets.

### *7.3.2 Long-Term Markets: Over-the-Counter and Futures Markets*

Prior to the market gate closure, generating units, suppliers, and qualified consumers can always trade freely their future supply needs. Long-term (in most cases one-year and typically not more than two years) contracts—in all their diverse formats—are the dominant form of transaction in wholesale electricity markets. Most market participants do not want to be subject to the uncertainty of the short-term electricity prices in liberalized markets, and long-term contracts provide hedging against this risk. In addition, speculators without any interest in physically buying or selling electricity may wish to bet against the uncertainty of the short-term electricity prices, and participate in organized markets selling all sorts of financial products with the short-term price of electricity as the underlying reference price.

Long-term contracts can adopt two basic formats: either physical or purely financial contracts. Physical contracts entail physical and cash delivery on expiry. The delivery point is the high voltage grid in general or some prescribed node in it—in some cases some important node or "hub" that is chosen as reference; in other cases the node where the buyer is located. Note that the selling party in a physical bilateral contract does not necessarily have to produce electricity (although this is the most frequent case); it suffices with purchasing electricity from other participants or in a short-term market and making sure it can reach the agreed delivery point. In some power systems it is possible to purchase a transmission right to transport the scheduled volume of power from the generation node to the agreed delivery point, when bottlenecks between the two nodes are expected, so as to reduce the risk that the agreed quantity may not be supplied by the contracting generator. Ideally, (as this depends on the specific regulation of the power system or systems involved) physical contracts provide the guarantee that the power will be delivered at the consumption point if the generator is producing and (if this is the case) the contracted physical transmission link is available, regardless of any situation of scarcity that may happen in the concerned power system(s). In the short-term the involved system operators must approve the previously declared schedule of the transaction, in order to prevent any violation of operating constraints. The system operators may put in place some mechanism for the settlement and management of any real-time imbalances between the declared and the actual transaction schedule.

On the other hand, a purely financial contract only entails cash delivery on expiry, when the differences between the specified reference index and the contractual price are settled. For instance, in the contract termed “contract for differences” the buyer pays the contractual price to the seller for the contracted volume, and the seller pays the reference index to the buyer for the same volume. The hourly or half-hourly spot prices of the day-ahead market at the power exchange are normally selected as the reference index. Market liquidity is important to provide a reliable index. These contracts are not related to the physical dispatch of the plants or to any actual consumption; in fact these financial contracts can be signed by any legal entity without any physical relation to the power sector. See below a case example that shows the implications of this type of contract.

Trading may be conducted essentially under one of two schemes: bilateral over-the-counter (OTC) contracts, and organized futures markets.

***Over-the-Counter markets***

In the OTC market model, each pair of counterparties reaches an agreement and concludes their trades independently. Generators and suppliers negotiate their contract terms and electric power is physically transmitted: these *forward contracts* are physical contracts, they influence actual dispatching, and they take place outside any organized market.

Brokers, organizations that bring buyers and sellers together (centralizing purchase and sale offers), also operate on these markets, but do not centralize default risk.[16]

There is no “single price”, but organizations such as Dow Jones and Platt’s attempt to compile information on the closing prices in OTC transactions and publish indices that players can use to establish the price of their operations.

One of the advantages of bilateral contracts is that they accommodate customized formats to match counterparty requirements, therefore reducing the basis risk.[17]

However, this advantage also entails drawbacks, because the more highly customized the contract, the higher the cost of possible assignment to a third party. To facilitate such transactions, the European Federation of Energy Traders (EFET) has setup a standard master agreement for the delivery and acceptance of electricity (the General Agreement) [9] in an attempt “to improve the conditions of energy trading in Europe and… promote the development of a sustainable and liquid European wholesale market”.

16 Default or credit risk is the risk that a counterparty will be unable to meet its obligations, i.e., the risk that the counterparty will default on its contract over the life of the obligation.

17 Basis risk in finance is the risk associated with imperfect hedging using futures. It may arise due to the difference between the price of the asset that is to be hedged and the asset underlying the derivative, or to a mismatch between the futures expiration date and the price of the actual selling date of the asset (www.wikipedia.com).

**Table 7.1** PJM Western Hub Peak Calendar-Month Real-Time LMP Swap Future

| Contract Unit | 80 Megawatt hours (MWh) (5 MW per peak hour) |
|---|---|
| Price quotation | The contract quantity shall be 80 Megawatt Hours (MWH) and is based on 5 megawatts for peak daily hours. Transaction sizes for trading in any delivery month shall be restricted to whole number multiples of the number of peak days in the contract month |
| Minimum fluctuation | $0.05 |
| Floating price | The Floating Price for each contract month will be equal to the arithmetic average of the PJM Western Hub Real-Time LMP for peak hours provided by PJM Interconnection, LLC (PJM) for the contract month |
| Termination of trading | Trading shall cease on the last business day of the contract month |
| Peak days | "Peak day" shall mean a Monday through Friday, excluding North American Electric Reliability Corporation holidays |
| Peak hours | From Hour Ending (HE) 0800 Eastern Prevailing Time (EPT) through HE 2300 EPT |

http://www.cmegroup.com/trading/energy/

***Futures markets***

Derivatives (*futures* and *options*) contracts are traded on a commodity exchange where the delivery date, location, quality, and quantity have been standardized. A derivative is a standardized contract where all terms associated with the transaction have been defined in advance, leaving price as the only remaining point of negotiation [34]. Standardized contracts allow participants to benefit from market liquidity and transparency, trading anonymously. *Future contracts* are standardized forward contracts traded in organized exchanges that, contrary to forward contracts, are typically of a purely financial nature. *Option contracts* confer the buyer, against the payment of a fee, the right, but not the obligation, to purchase ("call options") or to sell ("put options") a specified quantity of electricity at a specified time in the future, at a predetermined price (the strike price). Option contracts are purely financial contracts.

For example, The New York Mercantile Exchange (NYMEX) offers financially settled monthly futures contracts for on-peak and off-peak electricity transactions based on the daily floating price for each peak day of the month at the PJM western hub. In Table 7.1 an extract of the specification of one of the contracts traded in this market is shown.

On these organized exchanges, transactions are concluded by settling the difference between some contracted strike price and an index price (typically the day-ahead spot market price). The exchange takes the position of central counterparty to all operations it has registered, guaranteeing the fulfilment of obligation of both parties. Once an operation is registered the exchange manages the resulting positions, through its interposition as (central) counterparty of the operations, becoming the buyer in relation to a seller and a seller in relation to a buyer, and

**Table 7.2** Characterization of over-the-counter and organized contracts

| Properties | Trading method | |
|---|---|---|
| | Over-the-counter | Power exchange |
| Anonymity of trading | No | Yes |
| Counterparty | Bilateral | Central counterparty |
| Counterparty risk | Yes, unless cleared | No |
| Trading method | Continuous trading | Either continuous or central auction |

therefore allowing contract fungibility and eliminating counterpart credit risk.[18] The exchange requires counterparties to advance an initial amount of cash, the margin. Every day the contract is "marked to market": the exchange's clearing house records the price to value the contract, in order to reflect its current market value rather than its book value. If the current market value causes the margin account to fall below its required level, the trader will be faced with a margin call. The exchange will draw money out of one party's margin account and put it into the other's so that each party has the appropriate daily loss or profit.

Exchange derivatives prices are widely and instantaneously disseminated. It follows from the latter that the index price must be credible, reflecting the price at which power may actually be bought and sold on the market (which should be "liquid") at any given time (Table 7.2).

***A case example: Contracts for Differences (CfDs)***

CfDs are the best-known example of a risk hedging financial instrument. CfDs are two-way financial contracts that can be traded bilaterally or in an organized power exchange. CfDs specify the amount $q_c$ of contracted energy (either with a flat profile or with a prescribed pattern during the contract period) and the contract price (strike price) $P_c$. The reference price $P_m$ is the hourly or half-hourly spot price of the power exchange. There is no option fee or risk premium in this type of contract, since both participants agree on the strike price.

The monetary outcome of the contract is that the party with the consumer role receives from the party with the generator role the amount $(P_m - P_c) \cdot q_c$ This amount is positive when $P_m > P_c$, i.e., when the market price $P_m$ is higher than the contract price $P_c$ and it is negative in the opposite case.

A CfD has interesting consequences for both parties. Let us assume first that one of the parties is a true consumer with an expected flat demand $q_c$. Then, if the actual demand q of this consumer is precisely $q_c$, this consumer will be fully hedged with an acquisition price $P_c$ against any volatility in the market price, since, if it happens that the spot market price $P_m$ is higher than the contract price $P_c$, so that the consumer will have to pay $q_c \cdot P_m$ in the market, the CfD will make the generator to pay $(P_m - P_c) \cdot q_c$ to the

[18] See for instance www.omip.pt.

consumer, with a net value of $P_c \cdot q_c$ regardless of the market price. A similar reasoning can be made if $P_m < P_c$, with the net result that the consumer is fully hedged *if its consumption is precisely the contracted amount* $q_c$. What happens if the consumer deviates from the scheduled demand $q_c$, for instance with an actual $q > q_c$? Since the CfD only covers the amount $q_c$, any extra consumption $q - q_c$ will have to be paid at the current spot market price $P_m$. And, if $q < q_c$, the consumer will pay the actual consumed energy $q$ at the market price $Pm$, although the CfD will hedge this purchase—so that the net charge will be $q \cdot Pc$—and the remainder of the CfD will result in a payment by the generator to the consumer of the amount $(P_m - P_c) \cdot (q_c - q)$. This amount may be positive or negative, depending on the actual values of $P_m$ and $P_c$. Note that, if the market price happens to be very high ($P_m \gg P_c$), the consumer has the incentive to close all non-essential loads (e.g., shut down an industrial production facility) and reduce its actual load q as much as possible, of course depending of its utility function. On the other hand, if the market price is low ($P_m < P_c$) the consumer has the incentive to increase $q$ as much as possible, exactly as it would be done in the absence of the CfD.

A parallel discussion can be expounded for a true generator that has signed the CfD. If the generator produces the contracted amount $q_c$, the CfD will provide a complete hedge against any variability of the market price $P_m$ and $q_c$ will be sold at the price $P_c$. However, any production above $q_c$ will be priced at the market price $P_m$ and, if the production $q$ is lower than $q_c$, the actual production $q$ will be hedged at the price $P_c$ but the generator will be financially exposed to the risk of having to pay to the consumer the amount $(P_m - P_c) \cdot (q_c - q)$, which could be positive or negative. The decision-making process for the generator requires to include also the variable cost of production VC into the picture. The net income for the generator—including the market remuneration, the CfD, and the variable production cost—is:

$$q \cdot P_m - (P_m - P_c) \cdot q_c - q \cdot \mathrm{VC} = q_c \cdot P_c + (q - q_c) \cdot P_m - q \cdot \mathrm{VC}$$

meaning that the generator is subject to the market price $P_m$ for any deviations with respect to the contracted value $q_c$. Since any production has per unit cost VC, the generator must produce as much as possible if $P_m > \mathrm{VC}$ and shut down otherwise. The economic incentives are the same as in the absence of the CfD, but now the generator has a hedge for a production equal to $q_c$.

In summary, both the consumer and the generator are perfectly hedged against the volatility of the market price $P_m$, but only for the contracted amount $q_c$; any deviations are valued at the current spot market price. Therefore, a purely financial CfD has the property of hedging against the market price for the contracted amount, without impairing the incentive of the market agents to react to the real-time value of the spot market prices.

The situation is very different for speculators who have signed a CfD, perhaps because they thought that they had some a priori knowledge that the future market price will be higher or lower than the strike price $P_c$, or for any other reason. When $P_m > P_c$ the speculator in the generator role will have to pay $(P_m - P_c) \cdot q_c$, while an actual generator that produces $q_c$ will receive $q_c \cdot P_m$ from the market to compensate that loss. The same happens with the consumer when $P_m < P_c$, so that the consumption of electricity ends up being priced at $P_c$. In principle, speculators have no intrinsic hedge and they are subject to the full risk of the CfD.

Now, we present an alternative and much simpler view of the same situation. Since the CfD happens in parallel with the spot market, the economic implications of the market and the CfD can be examined separately. In the spot market the true consumer has to pay $P_m \cdot q$, whatever the value of $q$, while the CfD will provide an amount $(P_m - \mathrm{P_c}) \cdot q_c$, positive or negative, *which the consumer cannot modify* once the CfD has been signed. It is therefore straightforward to see that the consumer is always subject to the incentive of the spot market price, with the consumer being able to respond by decreasing or increasing consumption, regardless of the economic outcome of the CfD, which is independent of the consumer's actions. The same reasoning applies to the generator. This simple and direct way of analyzing this situation is one of the countless applications of Coase's Theorem. A folk version is presented in Annex B of this book, under the name of "Grandma's inheritance theorem". It is the experience of the authors of this book that the need for application of this theorem appears very frequently in regulatory decisions, rendering them much easier than when grandma's advice is ignored. Numerous regulatory authorities have adopted wrong decisions by ignorance of this basic economic principle.

### 7.3.3 Day-Ahead Markets

The core activity of the wholesale electricity markets is the day-ahead market (DAM), where trading takes place on one day for the delivery of electricity the next day. Market members submit their orders electronically, after which supply and demand are compared and the market price is calculated for each hour of the following day.

The design of DAMs has been permanently evolving over time. Different countries have developed a variety of market models, in such a way that it is difficult to classify them all into a small number of categories. Next, instead of trying to propose a classification of the known-to-date designs, we will review the different elements and features that characterize the design of a DAM.

#### 7.3.3.1 DAM Organization: Role in the Wholesale Market

Attending to the type of implementation, wholesale markets have traditionally been classified into two major groups, corresponding to two different conceptions on how full liberalization of the generation activity can be carried out (originally represented by the two pioneering models previously described): the so-called Electricity Pool model and the one we will denote as Power Exchange (PX) model.

***The Electricity Pool model***

The Electricity Pool model (as used here) means a highly centralized market (originally although not necessarily compulsory[19]) run either by the System Operator or by a Market Operator. At this stage, one of the key differences with the PX model is that these institutions operate on a cost recovery basis, recovering their operating costs through fees (administratively approved by the regulator) paid by market participants. PXs also charge fees, but in principle they are just subject to the regulator supervision and no cost recovery is guaranteed.

As previously mentioned, this alternative was the one originally implemented in the UK market back in the early 1990s. At the time of this writing, some of the main examples that can be included under this category are the wholesale power markets in North America (both in the US and Canada), Australia (e.g., the Australian Energy Market Operator), New Zealand, and in Europe the Irish market and to some extent the Iberian one, since, although most of its features are closer to the PX model, its fees are still determined by the regulators.

***The PX model***

Power Exchanges operate in an open trade context, in which the generating units' scheduling is decentralized, so market agents can either bilaterally engage into any type of agreement for the delivery of energy (in the so-called bilateral market) and then declare their production/consumption schedule directly to the System Operator at the market gate closure or can submit bids for buying and selling power to the trading platform of a PX. These organized markets are optional and anonymous and accessible to all participants satisfying admission requirements in exchange of a fee.

Ideally, the main objective of power exchanges is to ensure a transparent and reliable wholesale price formation mechanism on the power market, by matching supply and demand at a fair price, and to guarantee that the trades done at the exchange are finally delivered and paid.[20]

---

[19] As pointed out by Boisseleau [7], this model has been mainly the result of a public initiative and the participation has been usually mandatory (or highly encouraged). In some cases these markets are not strictly but de facto compulsory, since generators with capacity obligations are required to submit bids to the day-ahead market. At the time of this writing (end of 2011), some of the main compulsory pools left were the ones implemented in Ireland, Alberta, Australia, and New Zealand.

[20] See for instance the web page of the European Power Exchange, www.epexspot.com.

### 7.3.3.2 Bidding, Clearing and Pricing

The manner in which the different market mechanisms are cleared and the products are priced typically differs from one electricity spot market to another.

Roughly speaking, in organized short-term electricity markets the day-ahead market prices are, in principle, determined by matching generators offers and consumers bids. However, this can be achieved in a number of different ways.

We find three major features that characterize short-term electricity auctions:

- Whether they use complex bidding or simple bidding;
- Whether the pricing rule is discriminatory or non-discriminatory;
- Whether single, zonal, or nodal prices are computed.

A number of other aspects could also be distinguished [4]: the trading intervals used (hourly, half hourly, or even every five minutes), if portfolio bidding is allowed or not (i.e., if no link is required between bids and units or on the contrary each bid must refer to a particular unit), if is there a limited number of bids for each portfolio or unit per time interval, if price caps are implemented, if negative prices are allowed, etc. However, next we will focus on discussing the three ones previously highlighted as most relevant.[21]

***Complex versus simple auction***

Since electricity is a very complex commodity, and its production is subject both to inter-temporal constraints and to the existence of a number of non-convex costs, the format of the generators' offers can range from the so-called simple one (a series of quantity-price pairs per time interval) to a gray scale of more complex alternatives, in which inter-temporal constraints and/or multidimensional cost structures can be declared. We build our brief review of the main alternatives around the two extremes (complex and simple auctions), and then we introduce the hybrid alternatives implemented to amend these latter simple designs.

*Complex auctions*

In a complex auction generation agents submit offers, representing the parameters and costs which define best their generating units' characteristics (fuel cost, start-up cost, ramp up limit, etc.). For example, in the case of PJM, some of the technical parameters a generating unit declares are [28]: Turn Down Ratio,[22] Minimum Down Time, Minimum Run Time, Maximum Daily Starts, Maximum Weekly Starts, Hot Start Notification Time, Warm Start Notification Time, and Cold Start Notification Time.

With all these data, the market operator clears the market using an optimization-based algorithm that maximizes the net social benefit. This optimization

[21] The discussion that follows is based on Rodilla et al. [29].

[22] Turn Down Ratio is defined as the ratio of economic maximum MW to economic minimum MW.

algorithm shares most of the characteristics of the traditional unit commitment (see formulation in Sect. 7.3.1.1), but with the only difference that the data considered are market agents bids instead of costs.

Usually, market prices are obtained as a by-product of the complex optimization-based algorithm. In the next section, the way these prices and the agents remuneration are computed is introduced.

*Simple auctions*

The downside of the complex-auction approach is the associated complexity of the market clearing process. This factor has been the key argument held by (mainly) generators to move toward a much simpler auction, where the efficiency of the economic dispatch that results from the market clearing is sacrificed in favor of the transparency of the price computation process.

In the so-called simple auction scheme, the format of the offers does not explicitly reflect the generation cost structure (e.g., an offer component for the start-up cost) or imply any inter-temporal constraint. Instead, market agents submit simple offers/bids, which exclusively consist of price-quantity pairs representing the willingness to sell/buy the underlying product (one MWh in a certain time period of the day, e.g., an hour).

In this type of auction no optimization algorithms are needed to clear the market in such a way that net social benefit is maximized. Matching the market and obtaining the volume of electricity that is traded in each time period of the day is straightforward when offers and bids are simple: generation's offers are sorted in order of increasing prices and the demand's bids are sorted in order of descending prices. By finding independently for each time period the interception point between both aggregate curves (demand and generation) the operator directly determines the total volumes sold. Thus, a generator's offer will have to be accepted if and only if the market price in that particular hour equals or falls above the price offered, and analogously, a demand's bid will have to be accepted if and only if the market price equals or falls below the price bid.

This simple clearing procedure ensures that the net social benefit is maximized both in each particular hour and also along the whole day.

Summarizing, if complex conditions are disregarded, the pricing algorithm is as follows.

- Bids and offers for each delivery period are submitted by a specified deadline.
- A merit order is established.
  - Bids are ranked by price in descending order.
  - Offers are ranked by price in ascending order.
- The (equilibrium) market outcome is defined by the equilibrium market price (EP).

- The EP is the bid price for the amount of energy that corresponds to the cumulative amount of energy demanded.
- Bids specifying a price not lower than the EP are accepted.
- Offers specifying a price not higher than the EP are accepted.

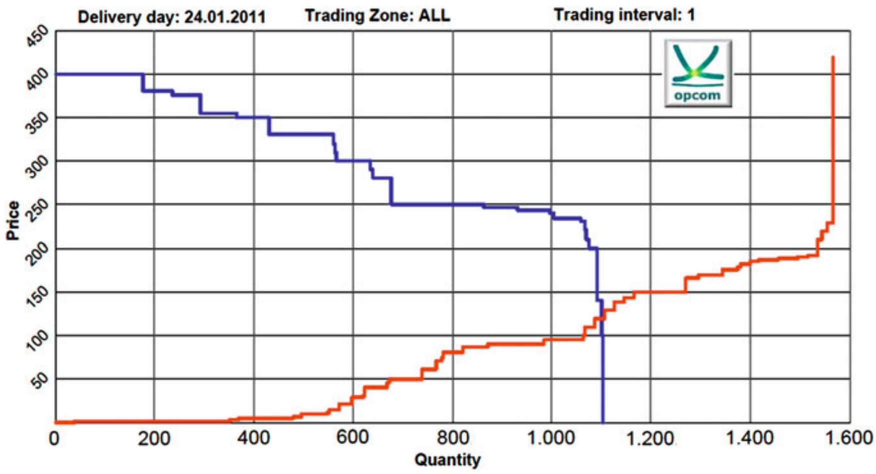


**Fig. 7.2** Aggregated curves in Opcom. *Source* www.opcom.ro

Figure 7.2 represents the aggregated curves (demand and supply curves) for a Romanian electricity market auction back in year 2011. As explained above, offers made at a price above the equilibrium price (the system marginal price, SMP) are accepted and are by the generators bidding at prices lower than the EP.

Fully simple offers/bids do not imply any inter-temporal constraint. This means that, for instance, the offers of one thermal generating unit in the day-ahead market could be accepted in the third, fifth, and seventh periods, leading to a resulting unit schedule which could be highly uneconomical or simply infeasible from the technical perspective. As we later further discuss, the main drawback of this approach is that it entails that to some extent generators have to anticipate (based on conjectures) the dispatch so as they properly internalize all cost in the hourly price component.

Due to this need to internalization that has just been introduced, simple offers lead to a significant lack of offering transparency. This fact severely complicates the market monitoring task, since it is difficult for the regulator to determine whether the offers represent actual costs or not, for there are many cost components represented in a simple hourly price. Thus, paradoxically, simple auctions offer a more transparent and replicable clearing process (the interception of two curves) based on not-so-transparent generating offers.

***Hybrid or semi-complex auctions***

In principle, the previous inconvenience could be partially fixed either by means of subsequent secondary trading (in the so-called intraday markets, in the EU context, or in the real-time market, e.g. in the US, see below) or closer to real time later in the balancing mechanisms/markets managed in most cases by the System Operator. However, in an attempt to combine the advantages of the complex and the simple auction design, EU PXs have opted for implementing hybrid alternatives, allowing linking semi-complex conditions to their offers.

The common idea behind the design of these semi-complex designs is simply to introduce as few complex constraints as possible in the auction, so as to not to complicate the matching process in excess while at the same time removing the huge risk at which agents are exposed in the simple auction context. Obviously, there is a whole continuum, between the extreme of including all potential constraints and the extreme of including none of them. The larger the number of constraints allowed, the closer the offers can represent the cost functions of the generating units.

In practice, this trade-off has been achieved either by introducing some of the most relevant (most difficult to be internalized) constraints, as it is the case with the ramp-up constraint (used in the Iberian day-ahead market) or by allowing some heuristic-based inter-temporal constraints in the offers format, in most cases not necessarily representing actual constraints or cost components, but rather a mixed effect of many of them. This is the case of the EU PXs semi-complex block-bids. These block-bids, as understood in the EU PXs context (for the term is also used with other meanings in some North American markets), allow agents to submit on one hand a certain interval of consecutive hours where they are willing to produce, and on the other hand the average price they require to be committed in that very period.

Some of the complex conditions and offers used in semi-complex auctions are for example user-defined block-bids (implemented, among others in the Nordpool, EPEX Germany, and EPEX France), meaning that a market agent can offer/bid a price/quantity pair for a set of consecutive hours (three as a minimum), flexible hourly bids (Nordpool and EPEX France), i.e., price/quantity pairs with no pre-defined hourly period assignment or the so-called minimum income condition implemented in OMIE, enabling a generating unit to include a minimum income condition expressed as a fix (expressed in euros) and variable term (in euros per MWh) associated with the whole set of hourly bids corresponding to one particular unit.

Figure 7.3 illustrates the impact of these semi-complex conditions for the case of the Iberian day-ahead market[23]: the resulting market price turns to be higher than the equilibrium price that would have resulted if no conditions would have activated.

---

[23] It can also be observed how in this market at that time a 180 EUR/MWh price cap was in force.

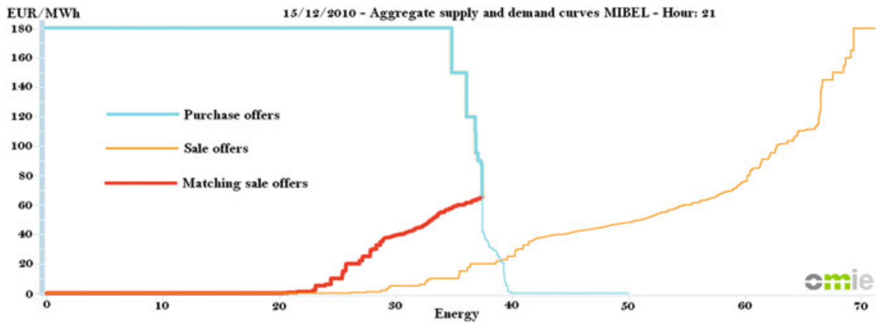


**Fig. 7.3** Impact of the semi-complex conditions on the resulting market price

***Pricing rules: discriminatory versus non-discriminatory payments***

The computation of market prices as well as the related determination of the generating units' remuneration is a quite controversial and still open issue in the context of complex auctions. We can classify these approaches into two large groups:

- nonlinear pricing rules (also known as discriminatory pricing schemes), according to which, on top of the hourly prices, some additional side-payments are provided on a differentiated per unit basis;
- linear (or non-discriminatory) pricing rules, according to which the same hourly price is used to remunerate all the hourly production and individual no side-payments exist.

It is straightforward to observe that total charges to consumers and payments to generators will be different under the two rule systems, and individual generators will get differentiated treatment under the first rule.

*Nonlinear pricing*

In the context of complex auctions, nonlinear (or discriminatory) pricing is undoubtedly the most extended pricing rule (especially in the US markets). This mechanism translates into each generator having a remuneration consisting of:

- first, a set of (nondiscriminatory, i.e., common[24]) prices which serve to remunerate all production in each time period,
- and then, some additional—individual—side-payments (in practice computed as a lump-sum daily payment) which are calculated on a per unit basis.[25]

[24] The term "discriminatory" should not have here any derogatory interpretation; it just means that the pricing rule results in different charges or payments for the several agents.

[25] A very closed-related technique is the alternative one proposed by O'Neill et al. [26]. This approach entails considering start (and other non-convex) costs as additional commodities different from energy, and thus, needing to be priced independently from this latter.

The set of non-discriminatory prices are computed as the dual variables (shadow prices) associated to the generation-demand balance constraint of the linear optimization problem that results when the commitment decisions have been fixed.[26] As a consequence of the method used to compute the marginal prices, these prices do not include the effect of non-convex costs (e.g., start-up or no-load costs). This is the reason why additional payments are considered on an individual basis in order to ensure (if necessary) that every unit fully recovers its operating costs. Roughly speaking, the price determination could be understood as a two-step process in which:

- First, the unit commitment optimization problem is solved considering all the convex and non-convex costs. The solution of this problem provides, among other results, the binary variables (e.g., start-up decisions) that result from the economic dispatch.
- Second, the same unit commitment problem is resolved, but in this case fixing those binary variables. The dual variables associated to the generation-demand balance constraint are taken as the non-discriminatory price.

In practice, there is not a single way in which this two-step problem can be addressed and solved. The most common approach consists of solving a mixed integer problem (MIP) by means of the branch and bound technique.

*Linear pricing*

Although the nonlinear pricing approach is the most extended alternative in the context of complex auctions, linear pricing is also a possibility, see e.g. Ref. [3]. Linear pricing in this type of auctions entails computing non-discriminatory hourly prices in such a way that all generating units fully recover their operation costs (thus avoiding the need for discriminatory side-payments of any kind), so in each time period (e.g., hour) every MWh produced is remunerated with the same hourly price.

Finally it is important to remark that we have just focused on the complex auction context. The reason is that the linear versus the nonlinear pricing discussion has been less relevant in the context of simple auction. This is mainly because the question on whether or not the single price should internalize the effect of non-convex costs (such as the start-up cost or the no-load cost) makes no sense in the simple auction scheme. In the simple auction context generators have to internalize all types of costs in their price-quantity pairs offers. Once submitted, there is no way for the market operator to make distinction on which part of the price corresponds to convex and which part of the price corresponds to non-convex costs.

[26] These prices in the vast majority of cases correspond to the variable fuel cost of the most expensive unit committed in each time interval.

***Prices and transmission constraints: nodal, zonal, and single pricing***

A number of regulatory options are open to deal with the allocation of limited transmission capacity for transactions among players under normal market conditions. One way to differentiate the main categories of options is to gather them in two main groups: those pricing algorithms that involve a detailed representation of the transmission network, and those other which consider a simplified one.

*Nodal pricing*

Nodal pricing applies security constrained economic dispatch to derive a bus by bus Locational Marginal Prices (LMP), the prices paid for the energy consumed or generated at a given transmission node, as used for instance in PJM, ISO-NE, or ERCOT.

LMPs reflect the value of energy at a specific location at the time that it is delivered. If the lowest-priced electricity can reach all locations, prices are the same across the entire grid, except for the losses effect. When there is transmission congestion (heavy use of the transmission system in an area), energy cannot flow freely to certain locations. In that case, more expensive and advantageously located electricity is ordered to meet that demand. As a result, the LMPs are higher in those locations.

LMPs result from the application of a linear programming process, which minimizes total energy costs for the entire interconnected power system, subject to a set of constraints reflecting physical limitations of the power system. The process yields the three components of LMPs: LMP [$/MW] = Energy component + Loss component + Congestion component. The energy component is the same for all locations. The loss component reflects the marginal cost of system losses specific to each location, while the congestion component represents the individual location's marginal transmission congestion cost.

Nodal energy pricing provides an accurate description of the technical and economic effects of the grid on the cost of electricity. They implicitly include the effect of grid losses and transmission congestion, internalizing both effects in a single value (monetary unit per kWh) that varies at each system node (Fig. 7.4).

*Zonal pricing*

Zonal pricing consists of using a single market price except where significant grid constraints arise frequently between a limited number of sufficiently well-defined zones of the power system. Once the most frequent points of congestion are identified, the grid nodes affected by internodal congestion are grouped into areas or zones. As defined by O'Neill et al. [25], in this context "a zone is a set of nodes in geographical/electrical proximity whose prices are similar and are positively correlated over time". This pricing mechanism distinguishes energy prices by zone in lieu of by nodes, and the same price prevails at all nodes within a given zone. Figure 7.5 shows an example of the application of zonal pricing in the NORDEL market.

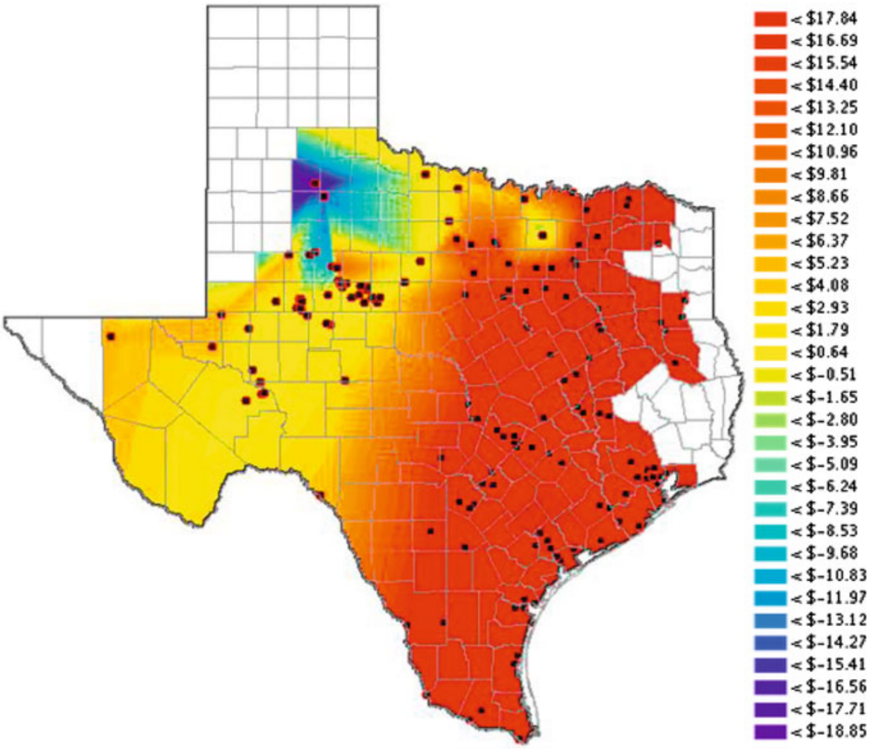


**Fig. 7.4** Locational marginal prices (LMPs) in ERCOT. *Source* www.ercot.com

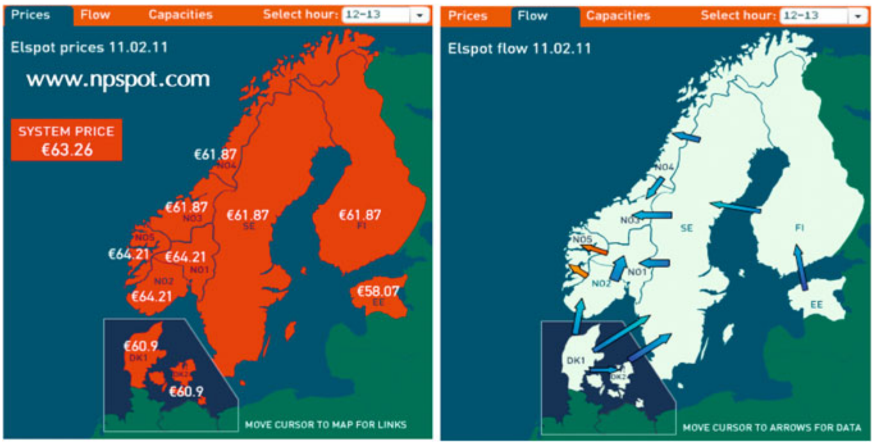


**Fig. 7.5** Zonal prices of the Elspot market in the Nordic region in Europe

At the other extreme of nodal pricing we have the so-called single pricing model, whereby any transmission congestions are fully ignored when the electricity market is cleared. This alternative is implemented in those markets where supposedly no systematic or structural congestions occur.

Thus, the market is first cleared in the day-ahead PX considering the simplified representation of the network (e.g., taking into consideration predefined theoretical interconnection capacities between the zones or directly ignoring transmission congestion in the case of single pricing).

In (supposedly) few cases in which grid constraints are detected, the System Operator re-dispatches the system, determining which players must withdraw from the system and which are to be included. Energy removed to solve the network constraint may be paid at the respective agent's bid price (if a specific bid related to the constraint solving mechanism is in place), at the opportunity price (energy market price less the price of the agent's bid), or not at all. When additional energy is requested, it is normally paid at the respective agent's bid price.

## *7.3.4 Adjustment Markets and Tools*

Once DAM closes, additional shorter termed tools have to be implemented to enable participants or the SO in their behalf to improve the day-ahead market results where sub-optimal, and afford the opportunity to assimilate new information (e.g., an unplanned outage). As previously introduced, two main models can be distinguished: the most common one in the EU context, built around intraday markets organized by the Power Exchange followed by balancing markets run by the System Operator, and the other one implemented in the US markets, centralized in the so-called Real-Time Markets.

### 7.3.4.1 Real-Time Markets

In the US model, real-time energy markets balance deviations between the day-ahead scheduled quantities of electricity required and the actual real-time load needs. These markets are spot markets in which current Locational Marginal Prices are calculated at five-minute intervals based on actual grid operating conditions. Security Constrained Economic Dispatch (SCED) is the real-time market evaluation of offers to produce a least-cost dispatch of online resources.

As a way of example, let us take the brief description of the real-time market operated by PJM [28]:

> The real-time energy market is based on actual real-time operations. Generators that are PJM capacity resources and Demand Resources that are available but not selected in the day-ahead scheduling may alter their bids for use in the real-time energy market during the Generation Rebidding Period from 4:00 PM to 6:00 PM (otherwise the original bids remain in effect for the balancing market). Real-time LMPs are calculated based on actual system operating conditions as described by the PJM state estimator. Load Serving Entities (LSEs) will pay the Real-time LMPs for any demand that exceeds their day-ahead scheduled quantities (and will receive revenue for demand deviations below their scheduled quantities). In the energy market, generators are paid the Real-time LMPs for any generation that exceeds their day-ahead scheduled quantities (and will pay for generation deviations below their scheduled quantities).

#### 7.3.4.2 Intraday Plus Balancing Markets

***Intraday markets***

Power exchanges generally hinge on a day-ahead market, where electricity is traded the day before the day of delivery. Most also provide adjustment markets (so-called intraday markets) where additional trading can take place when supply or demand situations change with respect to the estimates cleared on the day-ahead market. Market participants can modify the schedules defined in the DAM by submitting additional supply offers or demand bids.

As a way of example, a couple of these intraday markets are outlined below.

- Nord Pool: Elbas is an intraday market for trading power operated by Nord Pool Spot. Elbas supplements Elspot and covers the Nordic region, Germany and Estonia. At 14:00 CET, capacities available for Elbas trading are published. Elbas is a continuous market, and trading takes place every day around the clock until one hour before delivery. Prices are set based on a first-come, first-served principle, where best prices come first—highest buy price and lowest sell price. The products traded are one hour power contracts. Nord Pool Spot AS acts as counterparty in all contracts traded on the Elbas market and all trades are physically settled with the respective transmission system operator (TSO).
- Italy: The *Mercato infragiornaliero* (MI) takes place in four sessions: MI1, MI2, MI3, and MI4.[27] The sessions of the MI1 and MI2 take place after the closing of the DAM. They open at 10:45 a.m. of the day before the day of delivery and close respectively at 12:30 p.m. and 2:40 p.m. of the same day. The sessions of the MI3 and MI4 open at 4:00 p.m. of the day before and close at 7:30 a.m. and 11:45 a.m. of the day of delivery. The results in the four cases are made known half an hour later the closure.

***Balancing markets***

In real-time power system operation, the SO or TSO must ensure that supply matches demand at all times. Consequently, competitive electricity markets

[27] http://www.mercatoelettrico.org

generally feature a balancing mechanism that enables SOs to take measures geared to maintaining the supply/demand balance for which they are responsible. Imbalance pricing arrangements can be used to encourage market players to maximize their efforts to match supply and demand. Balancing markets therefore form an integral part of overall wholesale electricity trading arrangements and timetables.[28]

Notification and bidding deadlines differ depending on the country. Gate closure is a concept common to almost all SOs, although the details may differ, primarily in two respects [12].

- Gate closure may be a rolling deadline programmed at fifteen-minute (Netherlands), half hour (England and Wales), or hourly (Sweden) intervals, or a deadline set at specific times during the day (France, Germany, Spain).
- The time lapsing between gate closure for a given period and its start time also varies from one country to the next. In Sweden it is 1 min (i.e., the gate for the 10:00–11:00 period closes at 09:59). In Netherlands, Norway and England and Wales gate closure is one hour before the period begins.

After gate closure, market players may not vary in the expected physical position of their generators. In real time, the SO may change the physical positions of generators (or demand) to balance the system or manage congestion. If an interconnector links two systems that use different gate closure times, the earlier gate closure is generally applied.

Furthermore, actors may submit bids and offers on the balancing market specifying the extent to which they are willing to be paid to deviate from these positions and what they will charge for this service. These markets just remunerate the energy delivered but do not provide for any payment for availability.

In many countries a balance responsible party (BRP) is designated for this purpose. BRPs are market players that are financially responsible for balancing injections and withdrawals (including possible purchases and sales) of electric power. For instance, if one of the generating plants for which a given BRP is responsible fails to deliver electricity as scheduled in the generation plan, the imbalance cost is charged by the SO to that BRP. In a way, each balancing responsible party is like a "virtual network" that must keep its schedules balanced.

For the SO to exercise full control over system stability, all wholesale market players connected to the transmission system are usually either compelled to be a balance responsible party or to trade through an aggregator with balance responsible party status. Market participants are then either directly or indirectly bound by balancing market rules.

Following gate closure, the SO calls for generation bids and load offers to balance the system at the lowest cost. Where intraday markets exist, SOs need to take further bid and offer restatements into consideration when extending these requests.

---

[28] Part of this discussion is taken from ERGEG [11].

The purpose of balancing markets is to provide short-term operational security of supply (security of grid operation) from a market approach and settle energy imbalances. Hence, balancing markets must be economically efficient. SOs purchase balancing power using market criteria.

Balancing markets are generally designed to encourage market participants to manage their exposure to imbalance. For example, generators who run short of their notified position are usually required to purchase the difference between energy notified and delivered at a price established by balancing mechanism rules. This price is likely to reflect the costs to the SO of acquiring the energy shortfall and may be higher than the price on the intraday or day-ahead energy markets. Such arrangements are designed to minimize the amount of balancing power needed and lower overall balancing costs. Similarly, SOs themselves may be the object of incentive mechanisms designed by the regulator to encourage them to conduct balancing in a least-cost manner.

Market power may be a real risk in balancing markets, even where the market is relatively "non concentrated". Relatively small players may be able to impact the market heavily when the supply/demand margin is small, or where they cover a specific geographical position or have unique technical characteristics, particularly where demand parties or other generators are already committed or unable to respond to price signals on very short notice. An efficient balancing market should be as resistant as possible to any such exercise of market power, given market structure, and concentration. Transparent market operations enable all players as well as regulators to identify and discourage any unfair behavior.

Imbalances may occur only during operating hours and they are balanced using the balancing power provided by balance responsible organizations such as SOs. The costs of handling imbalances may be distributed across all users, allocated to the market participants involved in the imbalance, or a combination of the two. Parties incurring imbalance are, in any event, subject to some manner of 'imbalance charge'.

Imbalance arrangements and pricing must be simple and transparent so that the underlying principles are easily understood and justified in ways that enable market participants to readily assess economic risk. Imbalance arrangements must enhance balancing market and wholesale market efficiency. Balance settlement arrangements must provide for swift and accurate settlement and invoicing.

In principle, balancing markets must:

- deliver short-term operational security
- operate to economically efficient standards
- use market methods
- further effective competition
- not contribute to market power
- be non-discriminatory
- have clearly defined roles and responsibilities
- ensure transparency.

***Imbalance pricing***

In most European power systems, imbalances are presently settled via dual imbalance pricing, in which positive and negative imbalance volumes are priced differently, depending on the hour. A fictitious but representative pricing scheme is used in the following discussion. The price of imbalances in each half-hour period is assumed in this example to be determined by two factors:

- the absolute value of the imbalance for each balance responsible agent,
- the sign on the imbalance (positive or negative) for the system as a whole (the opposite sign is often referred to as the balancing trend, which is upward when the volume of energy involving upward balancing operations is greater than the volume of energy involving downward balancing operations, and vice versa).

In a given half-hour period, the settlement price for balancing agents whose imbalance has the same sign, positive or negative, as the system as a whole is computed as the average weighted cost of the balancing actions adopted by the System Operator to correct the imbalance. This value is adjusted by a factor of $1 + k$, which raises the price for negative (and reduces it for positive) imbalances, further penalizing imbalanced parties. The day-ahead price is applied to imbalances, positive or negative, which bear the sign opposite the sign for total system deviation.

This dual imbalance pricing, reinforced by the use of the factor $k$, is designed as an incentive for market agents to try to avoid deviating from their scheduled programs. A measure intended to improve system security, it nonetheless has certain adverse effects for market operation and overall efficiency.

The ultimate purpose of this measure is to safeguard system security, on the assumption that single pricing methodology reduces the incentives for agents to avoid deviating from their day-ahead programs. In fact, imbalances that bear the sign opposite the sign of overall system imbalance reduce net system deviation and contribute to mitigating the problem. In theory, then, the agents involved should be entitled to receive the balancing price, like any of the producers presenting bids to the system operator to solve the imbalance problem. Such uninstructed deviations should not be encouraged, however. The fear is that, on the one hand, rewarding these unintended imbalances that incidentally solve the system problem may create an incentive for some agents to deviate intentionally from their schedules to capture this additional income; and on the other that the single-price mechanism may weaken the incentives to maintain a balanced schedule.

This reasoning is sound, in principle. The dual price methodology induces costs, however, which may be significant in some cases. The dual pricing approach allocates security costs asymmetrically, for instance. Since imbalances are measured for BRPs only, netting is allowed for deviations within the same balancing party, but not for deviations between balancing parties. Take, for example, one generator that produces 10 MW more and another one 7 MW less than expected. If they are owned by the same firm (and thus under the same balancing umbrella), only a 3 MW imbalance is recorded. But if the generators are owned by two small companies, two imbalances are computed, one for 10 MW and the other for 7 MW. When the overall balancing results are calculated for the two generators, the larger firm is better off by an amount of 14 MW times the difference between the two balancing prices. The larger the spread between these two prices, the larger is the penalty for the small firms.

More dramatically, assuming no changes in the unit operating efficiency in a firm owning a portfolio of generators, if, for imbalance settlements purposes, the firm is artificially divided into two, the resulting payments for imbalances would be larger. Therefore, if no technical improvement is made, the impact on the firm is greater.

This effect is highlighted by the fact that intra-firm netting might be a result not only of accidental imbalances with opposite signs attributable to two generators belonging to the same firm, but also of each firm's ability to self-adjust its schedules, i.e., to compensate for a potential problem in one of its plants with its own generators.

New entrants or, more generally, generators owning small portfolios, are far more sensitive to imbalance prices, and therefore more intensely affected by the asymmetric nature of the dual pricing methodology. In other words, this regulatory measure allocates a higher portion of the implicit costs of security to smaller agents, creating an entry barrier for potential new actors and consequently hindering power market development. In most European countries, where a small number of agents control an extremely high percentage of system generation resources, this should be viewed as a matter of major concern.

Due to the peculiar strategic characteristics of electric power supply, it would not be difficult to defend the premise that no market efficiency objective is justified if it implies a cutback in security standards. At the same time, however, the impact and effectiveness of this kind of measures depend significantly on the particular characteristics of the power system where they are implemented. In a fully developed, mature and quasi-perfect market, the asymmetric impact on agents might be less troubling, since in principle the agents would be more size-comparable and the costs of market intervention less significant. But the opposite is true when agents' market shares differ radically.

## 7.4 Generation Ancillary Services

### *7.4.1 Introduction: Ancillary Services Products*

The foregoing sections reviewed the sequence of market-like transactions that take place prior to gate closure, i.e., when the final production schedule is determined for all participants after bilateral, spot or day-ahead market trading and possible schedule adjustment on the intraday market. From this time on, responsibility for generation scheduling and dispatching is transferred to the System Operator, which makes all the pertinent decisions and defines and manages a series of procedures to ensure the delivery of electric power to suitable quality and security standards.

Generation and network ancillary services are the services associated with the production, transmission, and distribution of electric power necessary to guarantee the quality, security, and efficiency of supply. Quality of supply is understood to mean maintaining voltage and frequency within acceptable margins for the system, security of supply to mean non-interruption of supply in the short term (not to be confounded with long-term reliability of supply), and efficiency to mean supplying electric power at the lowest possible cost.

Ancillary services are very closely associated with power production: traditionally, the provision, purchase, and remuneration of such services were fully integrated into basic power generation, from which they were considered to be inseparable. In the context of liberalization, a need is gradually being identified to establish separate provision and remuneration mechanisms for these services to minimize anticipated operating costs. While it is incumbent upon the regulator to develop the regulatory models for these services, this institution often delegates the responsibility for their implementation to the System Operator, as an independent agent with an in-depth understanding of system operation.

Following [20], operating reserves are defined as the real power capability that can be given or taken in the operating timeframe to assist in generation and load balance, and frequency control. There is also need for reactive power reserve, but it will not be discussed here. The types of operating reserves can be differentiated by: (a) the type of event they respond to, such as contingencies, like the sudden loss of a generator or a line, or longer timescale events such as net load ramps and forecast errors that develop over a longer time span; (b) the timescale of the response; (c) the type of required response, such as readiness to start quickly a plant or fast response to instantaneous frequency deviations; (d) the direction (upward or downward) of the response.

In spite of their relatively small financial significance (from 1 to 10 % of total generating costs), ancillary services, listed below, are absolutely necessary for system operation. Terminology and subdivisions into different services may differ from one country to another. A European-oriented way to classify these services is[29]:

- Frequency control,[30] which consists of the following three elements:
  - Primary reserve: automatic maintenance of the balance between generation and demand, using the rapid response governors built into generators to handle frequency deviations.
  - Secondary reserve: centralized and automatic function whose objective is to regulate generation output in a control area to:

exchange energy with other control areas at the programmed levels, and
return the frequency to its set value in case of a (major) deviation, thus restoring primary control reserve.

  - Tertiary reserve[31]: automatic or manual change of the generator operating point (mainly by re-scheduling) to restore an adequate level of secondary control reserves.

- Reactive power for voltage regulation is essential to establish and sustain the electric and magnetic fields of alternating-current facilities and has a direct effect on system voltage.
- Black-start capability (power restoration) is the ability of a generating unit to start up and deliver power without external assistance from the power system.

Most of the ancillary services discussed in this chapter fall under one of these three headings. But as stated, it is also possible to come up with different classifications. A more US-oriented approach can be found in Milligan et al. [20], where all types of reserves are classified into five categories, in decreasing order of quickness of reaction:

(i) *frequency response reserve* (to provide initial frequency response to major disturbances; also called primary control or governor response, acting in seconds);
(ii) *regulating reserve* (to maintain area control error within limits in response to random movements in a timeframe faster than energy markets can clear; also termed frequency control or secondary reserve, acting in seconds);

[29] See UCTE [38] for a collection of operating principles and rules for transmission system operators in continental Europe.

[30] As explained in Chap. 1, electric power systems have regulation mechanisms to keep frequency within an acceptable range around the nominal value. The aim of such controls is to maintain the equilibrium between the mechanical power delivered to the generators and the electric power demanded by the system.

[31] Tertiary reserve is sometimes classified under a different heading, “balancing energy”. Its inclusion in the upper hierarchical level of the process that controls system frequency is believed to be preferable, for this also avoids confusion with the balancing market explained below.

(iii) *ramping reserve* (to respond to failures and events that occur over long timeframes, such as wind forecast errors or ramps; also termed deviation reserve, balancing reserve or forecast error reserve, acting in minutes to hours);
(iv) *load following reserve* (to maintain within limits area control error and frequency due to non-random movements on a slower time scale than regulating reserves; also named tertiary reserve, acting in several minutes); and
(v) *supplemental reserve* (to replace faster reserve to restore pre-event level reserve; also called tertiary reserve and replacement reserve, acting from minutes to hours).

Regulating and load following reserves are used during normal system operation. Frequency response and supplemental reserves are used during contingencies. A mix of spinning and non-spinning reserves can be used for the slower reserves (ramping, load following, and supplemental) while the faster reserves (frequency and regulating reserves) require strictly spinning reserves.

It is incumbent upon the System Operator to establish the volume of service to be provided. The operator asks generators to provide the system with a certain reserve capacity (active power and reactive power reserve capacity or power with autonomous start-up capacity). Since the proportion of this capacity that will be used cannot be known in advance, the volume of service required must be established on the grounds of probabilistic criteria such as used in centralized planning decisions to install generating capacity.

### 7.4.2 Frequency Control

Frequency control is conducted at three levels, distinguished in terms of response time, as explained below.

***Primary regulation***
This is the automatic, local regulation provided by generating unit speed regulators. These regulators control frequency in the unit terminals by adjusting the turbine mass flow control valve: when frequency drops to below its established value the regulators widen the valve opening and narrow it when frequency rises. The variable that relates frequency variations to increases in power is known as speed regulation or droop and is a characteristic parameter of generating unit regulators. This level of regulation sustains frequency levels, preventing large deviations from the scheduled value. The response time for this type of regulation is measured in seconds. This type of regulation is sometimes expanded to include the natural response of the system as a result of the inertia of the generator rotors, i.e., the kinetic energy released or absorbed when network frequency varies. It also covers the automatic load shedding control performed by minimum frequency relays installed by distribution companies on certain loads under SO supervision.

***Secondary regulation***
This is the automatic, regional regulation provided by automatic generation control (AGC), which sends signals from the control center to certain generators to re-establish the nominal frequency value and restore the primary reserve capacity (and power exchanges with adjacent control areas to their original values). The response time for this type of regulation ranges from 5 to 15 min.

***Tertiary regulation***
This manual, regional regulation, provided by generating units and controlled by the System Operator, attempts to restore the secondary reserve capacity by dispatching the units that entail the minimum incremental operating cost for the system. The response time for this type of regulation is upward of 15 min.

Ranking regulation by response time optimizes the use of the available resources. Figure 7.6 summarizes frequency and power regulation in the event of an abrupt generation outage.

***Procurement and payment of frequency control services***
Europe's Transmission System Operators (TSOs) can purchase ancillary services in several different ways, depending on the type of service [12]. Ancillary services are acquired by the SO or TSO both before and after gate closure in short-term markets. Prior to gate closure, the SO may conclude long-, medium-, and short-term agreements, whereby generators commit to availability for providing primary, secondary, or tertiary reserves. These agreements may be awarded in the context of a monthly or annual market, or negotiated bilaterally. They typically specify the technical characteristics of the service, the availability level required (a certain number of hours, or between specified time periods), and a price.

Primary control services may be obtained either commercially or imposed as an obligation. In Germany, Poland, Sweden, and Denmark, all primary frequency control services are purchased commercially and generators are under no obligation to provide the service. In other countries, all the major generators are obliged to provide primary frequency control services, although payment arrangements vary. In France and the UK mandatory frequency response is remunerated on a

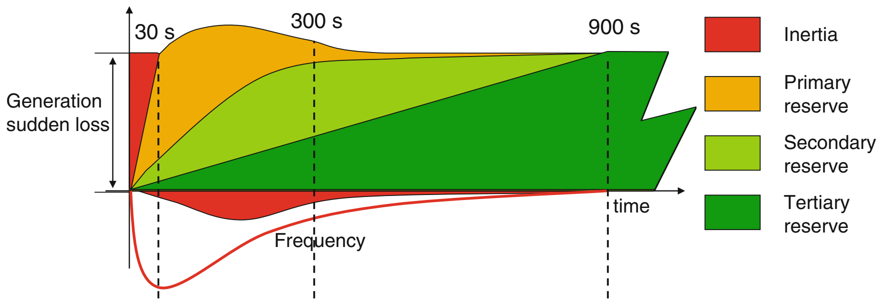


**Fig. 7.6** Frequency and power after a sudden loss in generation

cost reflective basis. In Norway, mandatory primary frequency response is partly remunerated under an agreement negotiated yearly and partly in accordance with a market-based arrangement in shortage periods. The service carries no explicit charge in Austria, Spain, Italy, Norway, Netherlands, Switzerland, or Slovenia.

Secondary control is generally a commercial service with no obligations attached. The exception is in France, where secondary control is required of generators exceeding a certain size and is paid on a cost reflective basis. In Germany, frequency control services are purchased under semi-annual tendering arrangements. Pre-qualified generators submit one price for primary control, and two prices for secondary control—one for reserve capacity, and the other for the energy delivered when the generator is called upon to provide the service.

As described in NYISO [24], the NYISO selects Regulation Service in the Day-Ahead Market from qualified Resources that bid to provide Regulation Service. Market Participants may submit bids to the NYISO for Regulation Service up to the Real-Time Market market-close time (75-minutes prior to the operation hour). Bid information includes:

- Regulation response rate, in MW/min, with the exception that Limited Energy Storage Resources (LESRs), such as flywheels and batteries, are not required to provide a regulation response rate.
- Regulation availability/price, in $/MW.
- Regulation Availability MW—regulation capacity available in one direction. For example, a bid of 5 MWs is a bid to provide 5 MWs of regulation up and 5 MWs of regulation down.

As it has been described in the previous section, some manner of ad hoc market is frequently used to purchase tertiary frequency control services: the real-time and balancing markets.

In many systems, to complement these markets, the SO purchases additional reserves, in line with the ones enumerated in the previous section.

For instance, ISO New England acquires three different types of these operating reserves:

- 10-minute non-spinning reserve (TMNSR)—A form of operating reserve provided by off-line generation that can be electrically synchronized to the system and increase output within 10 min in response to a contingency.
- 10-minute spinning reserve (TMSR)—A form of operating reserve provided by on-line generation that can increase output within 10 min in response to a contingency.
- 30-minute operating reserve (TMOR)—A form of operating reserve provided by on-line or off-line operating reserve generation that can either increase output within 30 min.

Additionally, ISO-NE runs a Locational Forward Reserve Market (FRM). FRM provides revenue to peaking resources that operate infrequently. Participants commit in the FRM through auctions held twice a year. The auctions set the prices and procure the reserve capacity to meet system-wide TMNSR and TMOR requirements and local TMOR requirements

### 7.4.3 Other Ancillary Services

***Reactive power and voltage regulation***

Another element of the electricity system, known as voltage–reactive power (V/Q) regulation, maintains the voltage in the transmission grid nodes within an acceptable range. This type of regulation is likewise phased to optimize the existing resources.

Primary voltage regulation is the automatic, local regulation conducted by an automatic voltage regulator (AVR). This regulator controls the voltage in the generating unit terminals, adjusting the excitation current to restore nominal voltage values by varying the amount of reactive power generated. This type of regulation is virtually instantaneous.

Secondary voltage regulation is the automatic, regional regulation involving the calculation of a reference voltage value which is sent to the generating units by the AVR every few seconds. The reference value is calculated on the basis of criteria that ensure system operating security. Unlike secondary frequency and power regulation, this mechanism is not installed in most electric power systems.

Tertiary regulation is centralis\zed, manual regulation, in which the System Operator sends reference voltages to generators at intervals measured in hours. These reference voltages are calculated as a compromise between technical and economic constraints.

Both of the above types of regulation, frequency/power, and voltage/reactive power, range from quick and less "intelligent" control (primary regulation) to slower control actions that require more computation time (tertiary regulation) but which yield solutions that ensure nearly optimum system operating conditions.

***Black-start capability***

The other traditional ancillary service is the restoration of the supply of electric power after outages. This is the system's capacity to return to full operation after a massive failure or blackout involving generation resources. Certain generators have independent start-up capacity, for which purpose the transmission grid is equipped with synchronizers.

## 7.5 Market Power in Power Markets

Once a comprehensive view of wholesale electricity markets has been provided, this section will extend the general introduction to market power in Sect. 2.5, by focusing on more specific features of power systems. The reader is advised to consult Sect. 2.5 before reading the present section.

### 7.5.1 Market Power and Market Structures

Market power fundamentally depends on the structure of the market rather than the rules, under sound market designs. Regulators should understand that when the market structure is not adequate for competition, the solution is not a change in the market rules, but a change in structure.[32]

As it was mentioned in Chap. 2, prior to the restructuring that led to the implementation of wholesale markets, most electric power systems were organized around a rather small number (often just one) of vertically integrated monopolies. These traditional electricity utilities were frequently fully government-owned (this was the case for instance in the majority of cases in Europe, being Spain one of the exceptions).

Turning this existing structure into another one that would be suitable for the implementation of a sufficiently competitive market happened to be tall order. When the utility was not fully controlled by the government but by private investors it was legally difficult to mandate the company to divest part of its assets to allow for the creation of a sufficiently significant number of relevant competitors. In the majority of cases utilities could not be legally required to divest, except when divestment is established as a condition for authorizing mergers and acquisitions (M&A). Therefore the most common outcome was the negotiation of complex agreements, frequently involving the acceptance of stranded generation costs by the regulators.

In those cases in which the government fully controlled the traditional utility, in principle it would have been possible to get to a sufficiently competitive market structure: in theory, it would have been enough to break the traditional company in a sufficiently large number of pieces (generation portfolios), in order to later sell them to different shareholders to guarantee an atomized and thus very competitive market structure, but experience shows that this was not always the case. In some power systems, the governmental decision was either to keep the full control of the company with no significant structural changes[33] while in others the decision was

[32] As we shall see later, interventionist changes in the market rules—i.e., new rules that limit the freedom of the market agents, like a price cap—do mitigate market power. But this happens in detriment of the freedom of the market that it was intended to promote, in the first place.

[33] Take for instance the case of France, in which EDF remains under the Government control and whose market share in the French wholesale market can be qualified as huge, above 85 % according to Eurostat [14], and the case of Vattenfall, still state-owned and whose market share is larger than 40 % in Sweden according to the same source, Needless to say that these market shares cannot be only judged in the context of the national market. The Nordpool in which Vattenfall is embedded turns to be a rather competitive and liquid regional market while the interconnection capacity of the French system with its neighboring markets is still far from being enough to deter the local market power of EDF.

to turn the utility into a publicly owned company within the new competitive market, divesting in some cases part of the generation assets (e.g., Enel in Italy) or a large portion of the shares in the market, but still maintaining a large size of the company (with the purpose of keeping a high market value of the company with a view to a possible future privatization, or to maintain the position of a "national champion" with a strong international presence).[34] One of the pioneer experiences was in the UK, where the former state-owned company's non-nuclear generation capacity was split into two companies, National Power and PowerGen, which were subsequently required to sell some of their assets (see Newbery and Pollit [22] and Thomas [37] for a description of the process and further examples in the UK).

It can be thus stated that in the majority of cases electricity markets are characterized by a concentrated structure, see for instance Eurostat [14] and US Environmental Protection Agency [39].[35] As a result, these oligopolistic schemes are prone to abuse of market power. Thus, looking for the means to identify the existence and detect the abuse of market power, and then to design prevention measures, has always been one of the main concerns of electricity regulators. Next, we introduce the main regulatory measures for the mitigation and control of the exercise of market power in electricity markets.

When markets are interconnected, it is not obvious to define the geographical scope of the market to which monitoring and mitigation measures should be applied. This leads to the concept and definition of "relevant market". According to the ruling of the European Commission, "A relevant product market comprises all those products and/or services which are regarded as interchangeable or substitutable by the consumer, by reason of the product' characteristics, their prices and their intended use. The relevant geographic market comprises the area in which the undertakings concerned are involved in the supply and demand of products and services, in which the conditions of competition are sufficiently homogeneous and which can be distinguished from neighboring areas because the conditions of competition are appreciably different in those areas".

The US Department of Justice & Federal Trade Commission in the "US Horizontal Merger Guidelines (1997) proposes the Small but Significant and Non-transitory increase in price test (SSNIP) to identify the scope of a relevant market: "A market is defined as a product or group of products and a geographic area in which it is produced or sold such that a hypothetical profit-maximizing firm, not

[34] The market share of Enel in Italy, still under the control of the Government (who owns more than 30 % of the shares) has evolved from 70 % in 1999 (the year in which it sold 15 GW) to below 30% [14], In Belgium, the market share of Electrabel, owned by GDF Suez, is above 75 %, while in Chile the market share of Enersis, bought by Endesa now part of the Enel group, as for 2010 was around 35 %.

[35] An excellent review of market power issues in the US is provided by Helman [16]. An easy to read, tutorial text is authored by Rose [30]; another useful reference is The Brattle Group [35]. FERC Order 697-A establishes the conditions to allow market-based rates, depending on market power mitigation issues. Although very technical in legal terms, this document illustrates the terminology and the issues involved when examining market power in actual systems.

subject to price regulation, that was the only present and future seller of products in that area likely would impose at least a "small but significant and non-transitory" increase in price, assuming the terms of sale of all other products are held constant. A relevant market is a group of products and a geographical area that is no bigger than necessary to satisfy this test". In the US a "small" price increase is normally defined as 5 %, while in the EU is 5–10 %.

### 7.5.2 Monitoring the Existence and Exercise of Market Power

Market power may be exercised in short-term electricity markets in two equivalent ways. In the first one, the price that is bidden for a given amount of energy is higher than what it would have been under competitive conditions. In the second one, the amount of energy offered at a price is lower than what had been competitively offered. The effect of market power in this second approach is usually the withdrawal of supply from the market.

Two long-term effects of the higher prices that result from the exercise of market power can be identified.

Generation firms have a natural incentive to install new capacity. Any additional capacity will ultimately place downward pressure on prices, therefore reducing the incentive for the exercise of market power. The easier it is for potential agents to install new capacity, i.e., the lower the entry barriers, the greater this effect is. In the absence of any entry barriers (i.e., when the cost of market entry is nil), the "virtual competition" from possible newcomers may ultimately suffice to ensure that prices are close enough to the competitive level. An oligopolist would not charge overly high prices for fear of attracting new competitors. Although incumbents on electric power markets often maintain that this is the case and that potential competition from possible new entrants guarantees competitive prices, significant entry barriers usually exist that are difficult to avoid. A long construction time for new power plants is an obvious one.

The exercise of market power eventually leads to lower electricity consumption than on a perfectly competitive market. The greater the elasticity of long-term demand, the lower the consumption.

***Market power metrics***

Certain special characteristics distinguish electric power markets from other traditional markets. First, the underlying asset is a practically irreplaceable good that cannot be economically stored. Generation side elasticity is very low and demand-side elasticity is presently nearly nil in the short term in most markets. Production capacity is exposed to high operating risks and investments are very capital-intensive. The transmission grid introduces complex market constraints. Market power is difficult to analyze because of all these reasons.

Concentration measurements in the context of power markets have been thoroughly studied. Newbery et al. [23] provides a good review. Another excellent reference is Hope [17]. Traditionally, simple general-purpose indicators have also been applied to power markets. The reader can find a comprehensive description in Sect. 2.5.4 of this book.

***Market power monitoring***

Proper market monitoring is therefore a necessity in the context of electricity markets,[36] although it faces an almost insurmountable obstacle: the calculation of the (supposedly) correct cost (i.e., competitive bid) of a generation unit is a difficult, fuzzy, subjective and thus, in most cases, pretty arbitrary task.

As stated above, the electricity supply cost/bid/price formation functions depend on such a variety of intricate drivers, that it is almost impossible for the regulator to properly supervise this process. In electricity markets the monitoring duty of the regulator turns to be a permanent pursuit of smoking guns. As stated by Barker et al. [6]: "When structure is not conducive to competition, the regulator and pool operator will find themselves unsuccessfully chasing after conduct. The solution is not a better rule, but a change in structure". If feasible, it would always be more efficient to implement an explicit *ex ante* regulatory measure to tackle the structural problem when detected. The objective of these solutions should be to mitigate the market power of dominant players while at the same time affecting the least possible the right marginal signals that lead generators behavior and system operation to the maximization of economic efficiency.

### *7.5.3 Measures to Tackle Market Concentration*

It is remarkable, the difference in the high level approaches to market power mitigation that have been adopted in the US and in the European Union. The US FERC has a statutory obligation under the Federal Power Act of 1935 to ensure that individual State Regulatory Commissions manage liberalization to ensure that wholesale prices remain "just and reasonable". Therefore, before an electric utility could be allowed to sell at wholesale market prices, any market power has to be adequately mitigated, and the authorization can be withdrawn (& regulated prices would be used instead) if "there is any change in status that would reflect a departure from the characteristics the Commission has relied upon in approving market-based pricing".

On the other hand, articles 81 and 82 of the EU Treaty examine mergers and acquisitions, as well as anticompetitive behavior, but they do not limit market power *ex ante*. The EU Electricity Directives aim to remove the barriers to create a competitive market, without (apparently) seeing the need to ensure that the

[36] See The Brattle Group [35] and Adib and Hurlbut [1] for two good descriptions of the role and function of market monitoring units, mostly from the US perspective.

resulting market structures were sufficiently competitive before introducing liberalization. The Electricity Directive 2003/54/CE establishes that “Member States will create adequate instruments to prevent abuses of dominant position”.

The regulatory measures for mitigation of market power can be categorized into two main sets: short-term (price or bid) caps and long-term structural interventions.

***Short-term caps***

A straightforward and very common way (particularly at the start of electricity markets back in the 1990s) to try to tackle market power has been the imposition of limitations in the generators’ bids (the regulator may, for instance, through the Market Operator, amend or even eliminate certain offers considered to reveal anti-competitive strategic behavior) or (more often) in the resulting market prices.

Price caps can be “hard” (a maximum price that demand can pay) or “soft” (when the price limit is defined as a function of some index, as for instance fuel prices in international markets).

Roughly speaking, these caps are far from being a good idea. The main and first reason is that, since it is not easy to differentiate between high prices due to scarcity and due to market power exercise, such a rule directly affects the key market marginal signal for generation and demand. This leads to an inefficient economic dispatch in the short term and a significant distortion of the long-term signal that should attract investment to achieve a well-adapted generation mix.

In addition to affecting market efficiency, limiting the maximum (hourly) price according to which generating units can be remunerated curbs the potential exercise of market power in an extremely partial way. Price caps are set at reasonably high levels to avoid market power abuse under a scarcity event. But experience suggests that market power in electricity markets (if present) is exercised at all price levels (not necessarily in the peak hours) in a moderate way, in order to avoid the regulator intervention.

***Long-term structural interventions***

The most efficient way to face market power consists of directly intervening on the market structure in the long term. This intervention can be irreversible (e.g., mandating the incumbents to divest) or transitory. Roughly speaking, this alternative entails forcing dominant players to sell part of their generation capacity, or part of their output for a sufficiently long period of time, under the expectation that, after that time, a sufficient number of new entrants will have reduced the concentration levels of the market.

*Divestitures*

As mentioned earlier, divestment, a traditional and one of the most drastic approaches to excessive horizontal concentration, has often been adopted. This is obviously the most traumatic procedure for companies. Although this method introduces new agents into the market, it also dramatically reduces the presence of the incumbent companies, which are traditionally more deeply committed to the system and, in principle, guarantee investment continuity.

Certain prior considerations must be addressed when implementing this measure. One of the keys to successful divestment is to ensure that timing is sufficiently flexible. A stable regulatory framework and an established market are likewise essential to attracting possible bidders.

A variation on this theme is to prohibit dominant operators from increasing their market share, a measure applied for instance in the case of Spain, in which Endesa, the former publicly-owned utility, was not allowed to grow in the market for a number of years during the early 2000s.

*Long-term contracting and market power*

The existence or the obligation to sign long-term agreements that fix the revenues earned by dominant firms for part of their capacity, also reduces the impact of excessive market concentration, see for instance Wolak [40]. Thus, in principle, long-term contracts are a helpful tool for mitigating market power. If a utility (with a large wholesale market share) concludes an agreement to sell part of its output in the long term, its incentive to raise prices by withholding part of its capacity declines. Since raising the market price would only benefit the firm for the amount not sold under the agreement (because it obviously receives the fixed price set in the agreement for the part already sold), the larger the amount sold under a long-term agreement, the lower is the utility's incentive to increase the price.

Suppose that a generator owning a large portfolio of plants and typically producing a quantity $q$ in the market, enters into a long-term contract for a portion $q_1$ of its market share. The immediate impact of this contract is the reduction of the generator's incentive to raise market prices. Indeed, if the generator attempts to manipulate the market price by withholding (or by bidding at a high price) a portion $q_2$ of its energy output, the contracted energy $q_1$ will not be affected by the price increase. The strategy of the generator should therefore consist of evaluating if the increase in income resulting from multiplying the quantity $(q - q_1)$ times the price increase outweighs the income loss due to withholding $q_2$. The long-term contract mitigates the incentive to increase the price, since it reduces the amount of energy that would benefit from the price increase resulting from withholding part of the generator's output (from $q$ to $q - q_1$).

This is only true, however, if the price laid down in the agreement does not depend on the spot market price: in other words, if when a forward contract is concluded, the spot market price is not suitable grounds for setting forward prices. But if the contract duration is not sufficiently long, the generator will clearly perceive that the short-term market price will be the reference price at the time of renewing the contract. Thus, the incentive to exercise market power for the generator will not be actually affected, since increasing short-term market prices will allow for a higher contract price

A pending issue, to be discussed later, is the way to properly price these long-term contracts that are mandated by the regulatory authorities.

***Energy release or virtual power plants auctions***
This method is based on auctioning, not generation capacity, but the right to manage production. Under this scheme a dominant company is required to auction the commercial management of part of its capacity for a limited but sufficiently long period of time, e.g., three to five years. This may materialize as a call option on a fraction of a company's (pre-defined) blocks of energy at a set price or by linking the contract to the performance of specific plants.

Energy release auctions have been used on a handful of occasions to limit the market power of dominant companies in markets such as Alberta, Canada, or The Netherlands. In other cases, such as in France, it has been used as a condition for authorizing merger and acquisition (M&A) processes to prevent significant erosion of competition levels.

This method is obviously less drastic than capacity divestment, provided that the physical facilities do not change hands and that the process is completely reversible when the contract expires. The advantage is that it attracts new agents, which can help to strengthen the retail market. It also reveals the market price of generation assets, which may be useful under certain circumstances.

As in divestment, these auctions should be spaced at sufficiently long time intervals and the maximum amount of energy auctioned should be limited to prevent prices from collapsing. A stable regulatory framework and an established market are also highly advisable prerequisites. Although to a lesser degree, this method also discourages the dominant companies from investing further in the local market.

***Administratively priced long-term contracts***
Very often, in line with the point just raised in the previous paragraph, the lack of market maturity and thus liquidity does not allow selling in the market the significant energy amount that should be released to properly mitigate the market power of the dominant player. The number of credible potential buyers in the auction could be too small, turning an oligopoly problem into an oligopsony situation, in which buyers are allowed to exert a great deal of control over the seller and can effectively drive down prices.

If this is the case, the only way to mitigate market power through forward contracting is by fixing the price externally. The regulator establishes the total volume of production to be contracted (and even the load profile), a term, and the price per MWh. This option has been implemented for instance in the Irish market, under the name of Directed Contracts (DC) and in Singapore. In the case of Ireland, according to the Single Electricity Market Committee [33] the prices of the DC are "determined by regression formulae that express the DC strike price in a given quarter and for a given product (baseload, mid-merit or peak) as a function of forward fuel and carbon prices". In the case of Singapore, the Energy Market Authority [10] states that the contract price is set based on the long-run marginal cost of the most efficient generation technology that accounts for more than 25 % of the total electricity demand and taking into consideration the key policy objective.

*Voluntary long-term contracts*
Voluntary long-term contracts would in principle have the same mitigating effect on dominant positions in power markets as virtual contracts, for under such arrangements the revenues received by the company for the volume of energy involved are independent of market price. The longer the duration of the contract, the more intense is this effect, which vanishes when the contract term is overly short (less than three years, for example). For such a contract to be accepted by the regulator as a reduction of effective capacity, it must prove to be unrelated to parallel commitments that imply uncompetitive conditions. The establishment of a set of satisfactory transparency guarantees would be no easy task, however.

***Other measures***
Another market power mitigation measure consists of facilitating the entry of new producers by removing any regulatory difficulties or uncertainties that might serve as deterrents. This measure is very important if the long-term aim of lower concentration is to be reached and is, in any event, a measure required to ensure that a market remains competitive in the long term.

Similarly, competition can be enhanced by furthering interconnections between neighboring electric power systems to heighten competition between adjacent markets.

Finally, a more elastic demand, able to react to high prices, would also reduce market power. This can be achieved by demand response programs, information, or educational activities, and it can be facilitated by the new telecommunication technologies.

These mitigation measures are unable, in any event, to fully eliminate market power. Depending on the type of instrument adopted, the ability to manipulate the market price may remain intact, although the economic incentive to do so will be smaller. More significantly, the maneuvring by the oligopolistic firm that may be required to benefit substantially from the abuse of market power may be much more readily detectable.

The stringency of market power mitigation measures should be reduced as the level of concentration declines or other regulatory instruments are introduced, such as improved market supervision mechanisms, more active demand response, elimination of barriers for new entrants, a broader margin of available generation capacity over demand, reinforcement of interconnection capacities, an adequate level of market information for all agents, and enhanced competition on operating reserve markets.

## References


1. Adib P, Hurlbut D (2008) Market power and market monitoring. In: Sioshansi FP (ed) Competitive electricity markets—design, implementation, performance. Elsevier, Amsterdam
2. AEMO (2010) An introduction to Australia's National Electricity Market
3. AIP (2007) The bidding code of practice. A response and decision paper. AIP-SEM-07-430. www.allislandproject.org. Accessed 30 July 2007
4. Baíllo A, Cerisola S, Fernández-López JM, Bellido R (2006) Strategic bidding in electricity spot markets under uncertainty: a roadmap. Power Engineering Society, IEEE general meeting, 2006
5. Batlle C, Barroso LA, Pérez-Arriaga IJ (2010) The changing role of the state in the expansion of electricity supply in Latin America. Energy Policy 38(11):7152–7160. doi: 10.1016/j.enpol.2010.07.037
6. Barker J, Tenenbaum B, Woolf F (1997) Governance & regulation of power pools & system operators. An international comparison. Papers 382, World Bank—Technical Papers
7. Boisseleau F (2004) The role of power exchanges for the creation of a single European electricity market: market design and market regulation. PhD Thesis, University of Paris IX Dauphine, Delft University Press
8. Bogas JD (1998) La convergencia de gas y electricidad en el mercado energético global (Gas and electricity convergence in the global energy market, in Spanish). Economía industrial, ISSN 0422-2784, no 321, pp 111–122
9. EFET European Federation of Energy Traders (2007) General agreement concerning the delivery and acceptance of electricity. www.efet.org. Accessed 21 Sept 2007
10. Energy Market Authority (2011) Vesting contracts. http://www.ema.gov.sg
11. ERGEG European Regulators' Group for Electricity and Gas (2006) ERGEG Guidelines of good practice for electricity balancing markets integration. Ref: E05-ESO-06-08. www.ergeg.org. Accessed 6 Dec 2006
12. ETSO European Transmission System Operators (2003) Current state of balance management in Europe. Balance management task force. Available at www.etso-net.org. Accessed Dec 2003
13. European Commission (2010) Interpretative note on Directive 2009/72/EC concerning common rules for the internal market in electricity and Directive 2009/73/EC concerning common rules for the internal market in natural gas. The unbundling regime. Commission staff working paper. Brussels, 22 Jan 2010
14. Eurostat (2012) Electricity market indicators. Available via http://epp.eurostat.ec.europa.eu
15. Green R (2010) Are the British electricity trading and transmission arrangements future-proof? Util Policy 18(4):186–194
16. Helman U (2006) Market power monitoring and mitigation in the US wholesale power markets. Energy 31:877–904
17. Hope E (2005) Market dominance and market power in electric power markets: a competition policy perspective. Norwegian School of Economics, Bergen
18. Joskow PL (2008) Lessons Learned from Electricity Market Liberalization. Energ J, Special Issue in Honor of David Newbery, pp 9–42
19. López-Velarde R, Valdez A (2008) "Mexico". In: O'Donnell EH (ed) Electricity regulation in 31 jurisdictions worldwide. Published by Getting the Deal Through. www.gettingthedealthrough.com
20. Milligan M, Donohoo P, Lew D, Ela E, Kirby B et al (2010) Operating reserves and wind power integration: an international comparison. Preprint. Conference Paper NREL/CP-5500-49019 Oct 2010
21. Newbery DM (1997) Pool reform and competition in electricity. DAE Working Paper 9734, Cambridge

22. Newbery DM, Pollitt MG (1997) The restructuring and privatisation of Britain's Cegb. Was it worth it? J Ind Econ 45(3):269–303
23. Newbery D, Green R, Neuhoff K, Twomey P (2004) A review of the monitoring of market power, for the European Transmission System Operators, ETSO, Nov 2004
24. New York Independent System Operator (2012) Ancillary Services Manual. Available at www.nyiso.com
25. O'Neill R, Helman U, Hobbs B, Baldick R (2006) Independent system operators in the United States: History, lessons learned, and prospects. In: Sioshansi FP, Pfaenberger W (eds) International Experience in Restructured Electricity Markets: What works, what does not, and why? Elsevier, 2006
26. O'Neill RP, Sotkiewicz PM, Hobbs BF, Rothkopf MH, Stewart WR Jr (2005) Efficient market-clearing prices in markets with nonconvexities. Eur J Oper Res 164(1):269–285. doi:10.1016/j.ejor.2003.12.011
27. Pérez-Arriaga IJ, Meseguer C (1997) Wholesale marginal prices in competitive generation markets. IEEE Trans Power Syst 12(2):710–717
28. PJM (2012). PJM Manual 11: energy & ancillary services market operations. Revision: 51. www.pjm.com. 8 Aug 2012
29. Rodilla P, Vázquez S, Batlle C (2012) Auction bidding, clearing and pricing schemes in day-ahead electricity markets: a comprehensive review. IIT Working Paper IIT-012-073A
30. Rose K (1999) Electricity competition: market power, mergers and PUHCA. NRRI, Ohio State University, 1999
31. Smith A (1776) An inquiry into the nature and causes of the wealth of nations. In: Cannan E (ed) 1904 Methuen and Co., Ltd., London. 5th edn. www.econlib.org/library/Smith/smWNCover.html
32. Schweppe FC, Caramanis MC, Tabors RD, Bohn RE (1988) Spot pricing of electricity. ISBN 0-89838-260-2, Kluwer Academic Publishers, Boston
33. Single Electricity Market Committee (2011) Directed contracts—2011/2012. Quantification and pricing. Decision Paper. SEM-11-045. www.allislandproject.org. Accessed 17 June 2011
34. Stoft S, Belden T, Goldman C, Pickle S (1998) Primer on electricity futures and other derivatives. Ernest Orlando Lawrence Berkeley National Laboratory, LBNL-41098, UC-1321, University of California, Jan 1998
35. The Brattle Group (2007) Review of PJM's Market Power Mitigation Practices in Comparison to Other Organized Electricity Markets. Prepared for PJM Interconnection, LLC. September 14, 2007
36. The Electricity Pool (2001) The electricity pool. www.elecpool.com
37. Thomas S (2006) The British model in Britain: failing slowly. Energy Policy 34:583–600
38. UCTE Union for the Co-ordination of Transmission of Electricity (2004) UCTE operation handbook. Version 2.5, level E. www.ucte.org. Accessed 24 June 2004
39. U.S. Environmental Protection Agency (2010) Electric generation ownership, market concentration and auction size. Technical Support Document (TSD) for the Transport Rule. Docket ID No. EPA-HQ-OAR-2009-0491. www.epa.gov. Accessed July 2010
40. Wolak FA (1999) An empirical analysis of the impact of hedge contracts on bidding behavior in a competitive electricity market. Fourth Annual POWER Research Conference, March 5, 1999