Economics of Electrical Grids

13

The electrical grid connects generators and customers. Without it, no electricity market is possible. For enabling competition among generators and retailers, third party access to the electrical grid must be assured on terms that are transparent and nondiscriminatory. From an economic point of view, electrical grids represent both a natural monopoly and an essential facility. This confers a dominant market position upon vertically integrated utilities and power grid operators that may be abused. To prevent this and the concomitant welfare losses, power grids need to be regulated.

Another issue is the network characteristic of the electrical grid. For reasons of economic efficiency, it links many countries on the European continent. The associated grid externalities require grid operators to provide system services according to common rules and standards, among others control power (also called regulation power) to keep demand and supply in continuous balance.

This chapter addresses the following questions:

– What are the economic reasons motivating grid integration?
– What are economically efficient approaches to the provision of grid services?
– What are economically efficient grid tariffs?
– What are the economic benefits and costs of unbundling?

How should interconnectors be efficiently managed?

The variables used in this chapter are:

C	Total cost
CS	Consumer surplus
c	Average cost ($= C/Q$)
c_{mc}	Marginal cost
e	Effort into cost reduction
g	Simultaneity factor
h	Full-load hour per year
L	Lagrange function

P. Zweifel et al., *Energy Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-662-53022-1_13

λ Lagrange multiplier
Π Profit
p Price for grid use
Q Amount of distributed electricity (in MWh)
RPI Retail price index
SR Sales revenue
W Macroeconomic welfare
X Efficiency factor

13.1 Grid Properties and System Services

13.1.1 Electrotechnical Aspects

The transmission and distribution of electrical energy is carried out through integrated electrical grids. With few exceptions, power lines are in alternating current (AC) operation (in contrast to direct current (DC)). AC lines deliver three-phase current with common voltage amplitude but with a phase difference of one-third of the period by combining three wires. Therefore, the sum of the three electric currents is always zero. The direction of electron flux alternates periodically with the target frequency of 50 hertz (Hz) in the European power grid with a tolerance of ± 0.15 Hz. In an interconnected AC grid, frequencies are synchronized, resulting in uniform oscillation.

Depending on the voltage level of the power line, one distinguishes between:

- Extra-high voltage grid (220,000–380,000 V) for long distance transmission;
- High-voltage grid (35,000–10,000 V) for interregional transmission;
- Mid-voltage grid (1000–30,000 V) for regional distribution;
- Low-voltage grid (220–380 V) for local distribution.

Since transmission and distribution entails losses that increase with distance, it is efficient to site generation as closely as possible to the point of use electricity. Indeed, the average transmission distance of electricity is below 100 km, with the consequence that transmission and distribution losses are below 5% of delivered electricity, the major part occurring in the low-voltage grid (see Table 13.1). However, according to Ohm's law, loss is inversely related to voltage, making long-distance transmission through extra-high voltage lines economically more viable.

Neglecting the use of electrical grids for telecommunication, power grids are factor-specific assets, meaning that they are exclusively used for delivering electric power from generators to consumers. Many types of infrastructure such as gas grids, water pipes, and railways are factor-specific assets. However, the electrical grid has a particularity which sets it apart from all other such assets: The current flow cannot be limited to one line of the grid. According to Kirchhoff's laws, the

Table 13.1 Average power transmission and distribution losses in Germany, in percent

	RWE energy	ESAG Dresden	SWM Munich
Extra-high voltage grid	1.0		
Transformation extra-high/high voltage	0.5		0.2
High-voltage grid	0.5	0.9	0.3
Transformation high/mid voltage	0.6	0.4	0.4
Mid-voltage grid	1.6	2.0	0.3
Transformation mid/low voltage	1.7	0.8	1.3
Low-voltage grid	4.5	3.7	2.3

Average power losses in percent of the delivered power in each voltage level
Source: Müller (2001)

current flow along a single line of a network depends on its electric resistance relative to all other paths that connect points of entry and exit. Thus electric currents always use all lines of an integrated network regardless of who owns them. Individual power lines can be separated from the grid for repair or to prevent damage from overload, but from an economic point of view, the integrated electrical grid is an indivisible good.

At the end of nineteenth century, electricity was initially supplied in local insulars. While the technical feasibility of long-distance power transmission was demonstrated in the 1880s, it took until in the 1920s and 1930s for nationwide electrical grids to develop in the United States and Europe. This development was fostered by supportive 'eminent domain' legislation weakening property rights of landowners affected by a power line. With the creation of the Union for the Coordination of Transmission of Electricity (UCTE) in 1951 which is now the European Network of Transmission System Operators for Electricity (ENTSO-E), grid interconnection and synchronization spread across Europe. An integrated grid offers the following advantages in terms of economic efficiency[1]:

– In case of power plant failures, customers can be supplied from distant power plants, permitting local providers to reduce their backup capacities. This is equivalent to an insurance-like pooling effect.
– The aggregation of regional load profiles results in a more uniform load, enabling power plants to operate more regularly. This is a positive externality, the so-called network externality (David 1987).
– Power plants that supply larger volumes thanks to enlarged service can be scaled up, resulting in lower unit cost of generation (economies of scale). However, recent technological change seems to have diminished scale economies in generation (Thompson and Wolf 1987), modifying the relative economic benefit of integrated power grids.

[1] The high-voltage networks in Europe are typically designed according to the $n-1$ criterion. This means that supply of all customers is still ensured, provided that a single resource (power plant, power line, transformer station) has failed.

– On the other hand, the development of offshore wind and other location-specific generation capacities has led to a renaissance of integrated power grids, with high-voltage direct current (HVDC) technology being used for reducing transmission losses in long-distance transmission.

Without access to the grid, no independent power producer can deliver electricity to customers and no retail customer can shift to a more efficient supplier. Therefore, the key condition for liberalization of electricity markets (see Sect. 12.2) is mandatory third party access to the grid on transparent and discrimination-free terms. Beginning in the 1990s, this condition was satisfied in several industrial countries.

13.1.2 Services to Be Provided by Electrical Grid Operators

When a customer purchases power from another generator rather than from the local utility, the electricity always comes from the nearest power plant connected to the grid. Currents in the integrated grid change only if generators lose (gain, respectively) customers, causing them to reduce (increase) generation. This is a consequence of non-storability (see Sect. 12.1.2). Similar requirements hold for retailers and eligible industrial customers with access to the wholesale power market who are expected to draw exactly the amount of electricity from the grid as contracted with their supplier. The electricity market is in equilibrium if all purchasing contracts are executable with the grid transmission and distribution capacities available.

In Europe, the synchronized integrated grid is divided into control areas where a single transmission system operator (TSO) has the responsibility for reliable and secure grid operation. A high quality of supply requires that all TSOs meet the technical rules and standards set up by the European Network of Transmission System Operators for Electricity (ENTSO-E).

Each TSO needs information to perform this task, which comes from retailers and eligible customers (also balancing group managers) who seek access to the grid. The data to be provided one day ahead comprise planned aggregate volumes of electricity fed into and withdrawn from all grid connecting points in their respective control areas. They typically cover for time intervals no longer than 15 min (see Sect. 3.6.3).

Based on this information, the TSO is obliged to provide the following services[2]:

– Frequency control (secured by so-called spinning reserve and control power);
– Voltage control (secured by compensating so-called reactive power);

[2]In some countries, the transmission operator is only responsible for the high and extra-high-voltage grid, whereas the mid and low-voltage power grids are controlled by distribution system operators (DSO).

– Black-start capacities for grid restoration after blackouts (secured by contracts with suitable generators);
– Compensation for transmission losses (which can be substantial in wholesale electricity purchases);
– Redispatch of generators in case of congested grid lines[3];
– Cross-border interconnection management;
– Balancing fluctuations in the supply of electricity produced from renewable sources (if required by the regulator).

While conventional power stations are mostly connected to the high-voltage grid controlled by the TSO, most of the distributed generation capacities are connected to mid- and low-voltage networks controlled by distribution system operators (DSOs). The DSO secures stable operation of the distribution grid, in particular voltage control. Grids of this type are not designed for large-scale transmission. They may even become obsolete with the implementation of smart grids.

13.1.3 Markets for Control Power

Due to the non-storability of electricity in the electrical grid, demand and supply for power must be equal within each control area. However, due to stochastic demand and supply fluctuations, permanent divergences between them occur that must be balanced by a system operator (TSO or ISO). Unexpected fluctuations arise both on the demand side (e.g. due to meteorological conditions) and supply side (e.g. power plant outages). The resulting imbalances can be recognized by deviations from the target frequency of 50 Hz (in Europe). Excess demand causes frequency to drop below 50 Hz, indicating that a positive amount of balancing power is needed. Excess supply causes it to rise above 50 Hz, calling for a negative amount of balancing power.

In Europe, the TSO has the obligation to provide balancing power to grid users which it procures on transparent and competitive markets for control power (also called regulation power); see Fig. 13.1. The TSO calls for combined tenders, specifying both volume and price, at given intervals. Control power is assured by three levels of reserve capacity:

– The primary reserve (historically also known as spinning reserve), which is automatically activated within 15 s and delivered simultaneously by committed suppliers. These suppliers are compensated for the capacity that must be available for both upward and downward regulation.
– The secondary reserve must be available within 30 s to 5 min. Auctions for it are multivariate because suppliers offer prices for both capacity (availability) and

[3]Redispatch means to change the power plant schedule by reducing generation in front of and increasing generation behind a grid congestion.

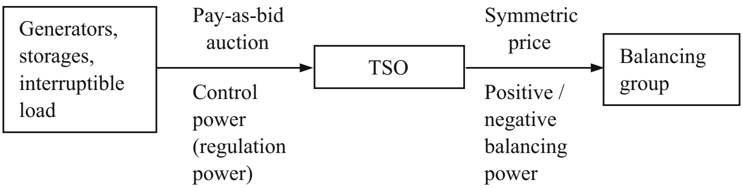

Fig. 13.1 Control power and balancing power

energy (work). Eventually, bids with the lowest prices for capacity are selected. Suppliers are activated by the TSO following the merit order based on the contracted prices for energy. A distinction (also in terms of prices) is made between positive and negative regulation power.

– The tertiary reserve (also called minute reserve) is used to substitute secondary reserves when necessary. It must be available within 15 min upon request of the TSO and is remunerated in a similar fashion as the secondary reserve. Again, a distinction between upward and downward capacities is made.

Prices for both capacity and energy are based on the pay-as-bid principle rather than according to the highest accepted bid (uniform pricing) as in the day-ahead market (see Sect. 12.2.2).

The procurement of balancing power entails additional cost of capacity and energy. The cost of capacity is charged to customers as a flat transmission grid fee. The cost of energy delivered is assigned to the parties seeking access (in the guise of so-called balancing groups) according to their individual discrepancies between registered and realized power (see Sect. 3.6.3). The energy cost depends on the total amount of balancing energy needed which corresponds to the sum of all discrepancies (with correct signs). The price of balancing energy is constant and equal to the average price of control energy supplied to the TSO with a time interval for pricing of 15 min (real-time pricing). There are no price spreads between positive and negative deviations. If this price is positive, balancing groups who exhibit net excess energy receive this price for supplying it, while parties who show net deficits have to pay this price.

13.2 Regulation of Grid Fees

In view of Kirchhoff's law, users of the grid lack control over the route electricity takes in the grid. Thus, the relevant concept is the extra cost caused by admitting an extra MW for transmission regardless of points of entry and exit. Grid fees based on this concept evidently facilitate competition among generators and have been favored by the European Union for this reason. In addition, the EC Directive 2003/54/EG on the single European electricity market establishes that access fees may be charged on the exit side, not on the entry side of the grid. Consequently, power consumers are charged for access to the grid according to the voltage level, with those connected to the low-voltage grid having to pay for all higher voltage levels. These fees are collected by the distribution grid operator, who transfers them to the respective operators managing the higher-level grids.

13.2.1 The Grid as an Essential Facility

The electrical grid is a natural monopoly, which means that the cost function that links transmission and distribution expenditure to quantity transmitted is sub-additive. Sub-additivity implies that combining grid assets of K system operators with transmission Q_k and costs $C(Q_k)$ respectively reduces the overall cost of supply, i.e.

$$C\left(\sum_{k=1}^{K} Q_i\right) < \sum_{k=1}^{K} C(Q_i). \tag{13.1}$$

Given a natural monopoly, it is therefore cost-effective to merge transmission units of a common region to become a single unit rather than having two or more separate companies compete with their grids. A sufficient (but not necessary) condition for a natural monopoly is that average cost be below marginal cost. This condition holds for power grids, at least given the current state of technology. Artificially injecting competition into the transmission and distribution industries is also near impossible because of high barriers to entry for newcomers especially in countries with already sufficient transmission and distribution capacities.

In addition, an electrical grid also constitutes an essential facility. Without it, generators cannot supply their customers unless they are located right next to the power plant. This gives vertically integrated utilities and power grid operators a dominant market position, which they may abuse by denying independent power producers access to the grid or charging them excessive fees. The result is an artificial limitation of supply that causes a welfare loss (see also Sect. 1.2.2). For this reason power grid operators must be regulated. The economic theory of regulation provides concepts and models for governments who seek to set access fees for grids in an optimal way.

13.2.2 Optimal Grid Fees

In the interest of welfare maximization, the price for access to a network should be equal to marginal cost, the so-called first-best solution. The regulator may be tempted to impose marginal cost-pricing. But the grid being a natural monopoly, its marginal cost is below average cost. Therefore, a price equal to marginal cost fails to generate enough revenue to recover total cost (see point A in Fig. 13.2). In this case, public regulation needs to find a second-best solution (Demsetz 1968).

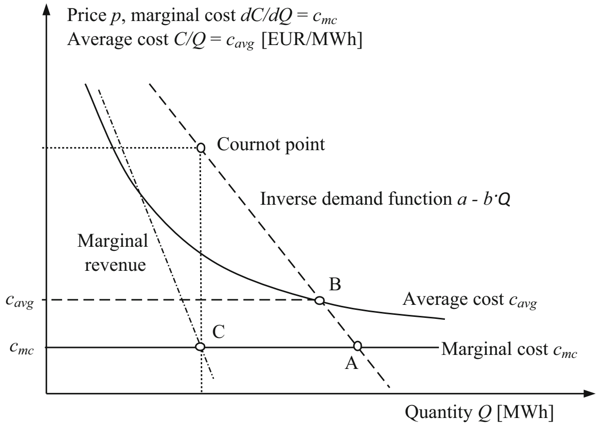

Fig. 13.2 The electrical grid as a natural monopoly

The classical proposal is to let the monopolistic network operator charge a price that covers average cost, including an appropriate return on the capital employed (so-called cost-plus regulation). However, this solution is problematic. Frist, since price depends on average cost, incentives to minimize cost are undermined. Second, there is a welfare loss because some customers who are willing to pay a price in excess of marginal cost are not served (those with a demand between points A and B of Fig. 13.2).

Both of these concerns are addressed by a split tariff, which amounts to price discrimination according to marginal willingness to pay (WTP):

– Customers with marginal WTP in excess of average cost cavg pay a price equal to cavg;
– Customers with a marginal WTP between marginal cost cmc and average cost cavg pay a price equal to cmc.

Therefore, the split tariff ensures that all customers whose willingness to pay is sufficient to cover the marginal cost of service are served. The utility breaks even since up to point B, its average cost is covered by the revenue obtained from costumers with high willingness to pay, whereas the extra quantity provided between points B and A is priced in a way to recover the additional cost.

Obviously, identifying customers with high marginal WTP is difficult. In addition, actually using this information for pricing is prohibited by the European directive on the common market for electricity (European Commission 2009a, b), which stipulates discrimination-free and transparent access to the grid. By prohibiting any kind of price discrimination in the use of grids, it in fact makes the introduction of split tariffs impossible.

Another approach is the two-part tariff:

– The first part of the tariff makes all costumers pay a price equal to the marginal cost of transmitting or distributing electricity. It preserves incentives for efficiency because by lowering marginal cost, the network operator can generate more profit.

– The second part of the tariff makes up for the shortfall in revenue. It is a separate price for access to the grid, which transfers part of the consumer surplus (given by the area below the demand function but above the marginal cost function) to the network operator. This part of the tariff is tricky because the network operator has a clear interest in appropriating as much consumer surplus as possible, beyond the amount necessary to break even. Moreover, the break-even point itself depends on the location of the average cost function, which is under the operator's control. Finally, consumers have a weakened incentive to invest in energy efficiency because the capacity fee does not depend on their consumption.

Yet several customers may use the same grid capacity if they do not call on it during the same period (see Sect. 12.3.1). To account for this, the capacity ordered is multiplied by a simultaneity factor $0 < g(h) < 1$, which is a function that increases with the so-called annual usage time h.[4] Annual usage time is an indicator of the probability that a grid customer orders grid capacity at peak load time. An example of such a function is

$$g(h) = \begin{cases} 0.1 + \dfrac{0.6 \cdot h}{2,500} & \text{for } h \leq 2,500 \\ 0.58 + \dfrac{0.42 \cdot h}{8,760} & \text{for } h > 2,500 \end{cases} \tag{13.2}$$

A distinction can be made between off-peak and peak usage of the grid. Therefore, grid companies can be regarded as monopolistic companies with $k = 2$ products or segments. In order to avoid arbitrary allocation of cost between the two segments, regulators may use the Ramsey pricing model. This model calculates a price vector (p_1, p_2) that maximizes net welfare (gross welfare W minus cost C),

$$\max_{Q_1,Q_2}\left(W(Q_1,Q_2) - C\left(Q_1,Q_2\right)\right) \tag{13.3}$$

subject to the constraint that total revenue must cover cost,

$$\sum_{k=1}^{2} p_k \cdot Q_k - C(Q_1,Q_2) \geq 0. \tag{13.4}$$

A crucial assumption is that the production cost as well as demand for one two products is independent of production cost and demand for the other (see Laffont

[4]The annual usage time h [h/a] is calculated by dividing annual amount of energy transmitted [MWh/a] by maximum capacity demanded [MW] during the pertinent period. Maximum capacity demand is measured over a fixed time unit (usually 15 min).

and Tirole 1993: p. 250 for the case of nonzero cross price elasticities for demand, involving so-called super-elasticities).

If the regulator does not want to grant the network operator any excess revenue the constraint (13.4) becomes an equality. Thus, the optimization problem can be solved using the Lagrangian approach,

$$L = \big(W(Q_1,Q_2) - C\big(Q_1,Q_2\big)\big) + \lambda\left(\sum_{k=1}^{2} p_k \cdot Q_k - C(Q_1,Q_2)\right) \rightarrow \max! \quad (13.5)$$

Here, $\lambda > 0$ denotes the Lagrangian multiplier which indicates how strongly goal attainment would suffer if cost where to exceed revenue. If gross welfare is equated to consumer surplus CS (see Eq. (12.1)), one has

$$\frac{\partial W}{\partial Q_i} = p_i, \quad i = 1, 2. \quad (13.6)$$

With this in hand, the first-order conditions of the Lagrangian function read

$$\frac{\partial L}{\partial Q_i} = (p_i - c_{mc,i}) + \lambda\left(\sum_{k=1}^{2} \frac{\partial (p_k \cdot Q_k)}{\partial Q_i} - c_{mc,i}\right) = 0. \quad (13.7)$$

The sum in Eq. (13.7) contains terms which pertain to cross-price elasticities, which however are neglected in keeping with the assumption of independent demands. Therefore, this sum can be written as

$$\sum_{k=1}^{2} \frac{\partial (p_k \cdot Q_k)}{\partial Q_i} = \frac{\partial p_i}{\partial Q_i} \cdot Q_i + p_i. \quad (13.8)$$

The following optimality condition for the price p_i results,

$$\frac{p_i - c_{mc,i}}{p_i} = -\frac{\lambda}{1+\lambda} \cdot \frac{1}{\eta_i} \quad \text{with} \quad \frac{1}{\eta_i} = \frac{\partial p_i}{\partial Q_i} \cdot \frac{Q_i}{p_i} < 0 \quad (13.9)$$

Here, η_i is the own price elasticity of demand. The price p_i resulting from Eq. (13.9) is called the Ramsey price.

According to Eq. (13.9), the Ramsey price contains a surcharge over marginal cost. It increases with λ, the so-called shadow price of a constraint, in the present case constraint (13.4) that ensures the recovery of total cost. Furthermore, the surcharge decreases with the absolute value of the own price elasticity of demand η_i. If customers in a particular market segment respond strongly to an increase in the grid fee (as is typical for off-peak customers) a surcharge over marginal cost entails a large welfare loss. On the other hand, there are market segments where the price elasticity is low (typically peak customers). Following the optimality condition, these customers should bear a larger proportion of the grid cost than off-peak customers.

For the practical application of Ramsey pricing, regulators need to have detailed knowledge of the current cost of the grid (backward-looking) as well as the future cost associated with efficient grid services (forward-looking). A large number of operational and economic parameters must be assessed as well:

– How much physical (and hence financial) capital is required for efficient operation of the grid?
– What standards of quality (e.g. interruption duration) must the grid operator guarantee?
– Is the network operator to be allowed to include the cost of future expansions of the grid in the fee?
– What method of depreciation is to be applied to grid assets (e.g. procurement cost or replacement cost)?
– What is the so-called rate base, i.e. the allowable share of equity?
– What is the allowable rate of return on equity?

There is the definite possibility that the regulator answers these questions in a way that conflicts with the assessment of the grid operator. In the case of private grid ownership, the specifications of the regulator in fact determine the company's decision to invest, blurring the division of responsibilities between the two. Eventually, the result may be a nationalization of electrical grids because it vests managerial responsibility unambiguously with the government. However, such a decision will always entail long-term consequences for economic efficiency.

13.2.3 Incentive Regulation

Another critical issue of all public regulation is the asymmetry of information between regulator and the regulated firm (the grid operator in the present context). Grid operators have detailed knowledge concerning potential for efficiency improvement and tendencies in demand that is unavailable to regulators. This is an instance of the principal-agent problem, where a principal lacks the information for controlling the agent's effort, who can therefore pursue its own interests. All the principal can do in this situation is to structure the contract in a way as to provide the best possible incentives to the agent, at least in expected value. In the present context, grid operators may use their informational advantage to obtain grid fees in excess of the level justified by minimum cost regardless of the regulation method chosen:

– Under rate-of-return regulation (also known as cost-plus or markup regulation), grid operators are granted a fixed markup on proven cost. This type of regulation creates an incentive to increase cost, notably by employing capital in excess of the economically efficient amount. This is the so-called Averch-Johnson effect (see Averch and Johnson 1962).

– Under price-cap regulation, the regulator sets a maximum grid fee. In this case, the incentive is to increase profit by reducing investment. Therefore, price-cap regulation results in underinvestment, thus hurting grid reliability in the long term.
– Under revenue-cap regulation, the regulator sets the maximum revenue. Revenue-cap regulation gives rise to an incentive to increase profit by minimizing costly grid services.

Whatever the approach of the regulator, its objectives may fail to be achieved, notably economically efficient and reliable grid operation. The popular response to this failure is tighter control and increased sanctions. However, such a response often is not helpful because supervision and compliance are not without cost themselves, resulting in an increase in the macroeconomic cost of electrical grids.

This dilemma has spawned the concept of incentive regulation, which was developed by Stephen Littlechild, who later became the first regulator of the electric industry in the United Kingdom (see Beesley and Littlechild 1989; Laffont and Tirole 1993, Chap. 4). According to this concept, regulation should be compatible with the incentives of the regulated firm. Applied to grid operators, it calls for letting them earn higher profits for a few years if they increase efficiency more than required by the regulator. After this grace period, however, they must pass the benefits from efficiency gains to their customers in the form of lower fees.

Incentive regulation determines the time path of a selected indicator, e.g. maximum allowable sales revenues SR, in the following way during a specified period,

$$SR_t \leq SR_{t-1} \cdot \left(1 + RPI_{t-1} - X_{general} + X_{\mathit{individual}}\right). \tag{13.10}$$

In this formula, $RPI_{t\text{-}1}$ denotes the percentage change in the index of retail prices over the previous period, $X_{general}$, a required rate of productivity increase, calculated over all grid operators, and $X_{individual}$, a required rate of productivity increase, applied to an individual grid operator.

According to Eq. (13.10), an operator's revenue may increase with the general rate of inflation. There are two extensions, however. The first is a deduction reflecting the rate of productivity increase in the industry. The second is designed to raise the bar for grid operators who have been lagging behind, forcing them to catch up with the rest. Conversely, grid operators who improve productivity $X_{individual}$ by more than $X_{general}$ can benefit from an increase of their allowable revenue, permitting them to earn higher profits. In this way, incentive regulation seeks to conserve incentives for dynamic productivity improvement. Grid operators can retain excessive profits, but only temporarily because the regulator adjusts the formula (13.10) at the end of a specified period. At that point, costs and profits are examined, which are (close to) their true values, providing information that would usually not be accessible to regulators.

In practice, this approach suffers from its exclusive focus on cost-efficiency. Reliability and other quality dimensions of supply aspects are not considered. Security of supply is defined here as the capability of the power transmission and distribution system to continuously maintain the flow of electricity in case of unforeseen disruptions. To account for this aspect, the incentive regulation formula (13.10) can be extended to include a bonus for high-quality grid operation which is usually based on the value of lost load (see Praktiknjo 2013). An indicator of quality is the predicted number of grid customers that can still be supplied if one element of the grid (e.g. power line, transformer, control room) fails (this constitutes the so-called $n-1$ criterion). Rather than this ex-ante indicator, most regulators use ex-post indicators. These include

– SAIDI: System Average Interruption Duration Index;
– SAIFI: System Average Interruption Frequency Index;
– CAIDI: Customer Average Interruption Duration Index.

Usually, these indicators reflect quality deficits only with a time lag. While insufficient maintenance reduces cost immediately, the quality of grid services deteriorates only in the medium term. Conversely, expenditure on investment and maintenance increases grid cost instantly but has a positive effect on quality with a lag.

13.2.4 Unbundling

The term ‘unbundling’ means undoing the vertical integration that has been characterizing the electric power industry for the past century. Its objective is to open up the market to competition between generators and to traders who are independent of both generators and distributors. However, pursuing this objective through unbundling is not without opportunity cost because the efficiency advantages of vertical integration mentioned in Sect. 12.1.3 are lost. Nevertheless, the EU Directive 2009/72/EC (European Commission 2009a) stipulates that large utilities must be at least legally unbundled, resulting in independent business units for generation, transmission, and distribution (see Table 13.2). For the time being, unbundling in terms of ownership is not required. Alternatively, grid ownership can remain within the integrated company, in return, operation of the transmission network is to be transferred to an independent system operator (ISO).

An example of the unbundling of the grid is the PJM (Pennsylvania—New Jersey—Maryland Interconnection) market in the northeastern United States, which serves an area of 13 states with 51 mn grid customers. In addition to providing the usual grid services, an independent system operator (ISO) determines transmission prices at each node where power can be fed in and taken out (so-called nodal pricing). Every 5 min and at every node (approaching a real-time market), the locational price is determined by the marginal cost of the last power plant which has to be connected to the grid in order to cover the load forecasted by the ISO without

Table 13.2 Unbundling concepts

Accounting	Informational	Management	Legal	Ownership
Separate accounts for different lines of business	Confidential treatment of sensitive data within the line of business	Division of business units into separate departments	Legal separation of business units	Spin-off and sale of grid
Regulatory requirements concerning financial statements	Separate use of information by lines of business	Functional separation of staff	Regulatory requirements concerning (in-) admissible relationships between business units	No grid ownership permitted for power plant operators
		Financial auto-nomy of departments		Possibly state ownership

violating any grid restrictions. Furthermore, the ISO performs the economically efficient dispatching of power plants using data such as maximum power gradient (i.e. the speed with which the plant can be brought up to required output), minimum uptime and downtime, and start-up and shut-down cost. Power plant operators act according to the price signaled by the ISO, which reflects the shadow price (i.e. the value of the Lagrangian multiplier) pertaining to the constraint,

$$Generation = Load. \tag{13.11}$$

This shadow price is part of the solution of an optimization problem. Power plant operators are free to not respond to this price signal, speculating to be able to extract higher capacity prices in a later period. The price signaling activities of the ISO are financed in analogy to the market for balancing power in Europe (see discussion in Sect. 13.1.3).

13.3 Economic Approach to Transmission Bottlenecks

According to Kirchhoff's laws, the transmission of electricity between a generator and a so-called load sink uses all available routes. This can lead to loop flows across linked control areas of a grid, giving rise to congestion. As a result, intended trades cannot be executed simultaneously, forcing the (independent or transmission) system operator (ISO or TSO, respectively) to modify individual delivery schedules.

The left-hand side of Fig. 13.3 illustrates such a situation. A generator (indicated at the top left) seeks to transmit 8 MW to a customer (indicated at the bottom left). The direct connection (dashed) has a capacity of 4 MW only. However, the

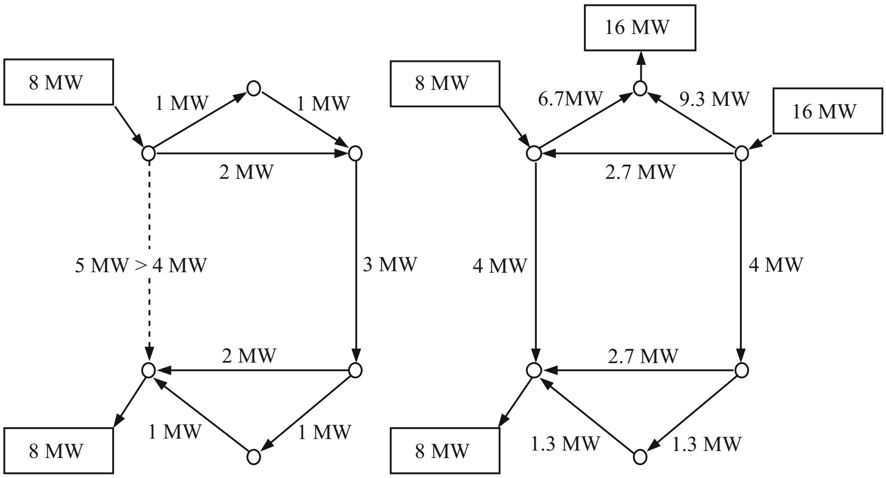

Fig. 13.3 Reverse flow and the elimination of a grid bottleneck

intended transmission would trigger a power flow of 5 MW in the dashed line, resulting in system failure.

The system operator (ISO or TSO) can avoid congestion in this example, by ordering an additional delivery between two indirectly affected grid nodes (see right-hand side of Fig. 13.3). The additional delivery of 16 MW creates an indirect counterflow of 1 MW on the congested line. As a consequence, the net demand placed on this link is reduced to 4 MW, equal to its capacity.

This is but one of several options for dealing with grid bottlenecks. Other options are the following.

– Rationing: This amounts to capping the amount of power that can be transmitted during a given period. If rationing is imposed frequently, grid customers begin to weigh the value of lost load caused by it against the value of purchasing and operating emergency backup units. They cannot be expected to undertake the investment for the elimination of a notorious network bottleneck themselves. Such an investment would benefit all other grid customers, creating a positive external effect. Therefore, this is up to the grid operator, who can be induced by the regulator to initiate the necessary investment e.g. by granting increased grid fees.[5]
– Explicit auctioning of temporal capacity rights on critical segments of the grid (see Hogan 1993): A company who has acquired capacity rights is allowed but not required to use these rights at its discretion. This gives it potential for abuse by not exercising them, thus blocking transmission by competitors. In this way,

[5]For the elimination of transborder grid bottlenecks, the European Commission envisages subsidizing investments as part of its Trans-European Networks program.

regional market areas can be insulated from international competition. A solution to this problem is for the grid operator to be able to withdraw capacity rights from non-users, applying the principle "use it or lose it". The elimination of grid bottlenecks could in principle be financed using the proceeds of these auctions.

– Implicit auctioning of capacity rights: In the absence of a grid bottleneck between two market areas, price differences between them can be removed by merging the two (so-called market coupling). If the local power exchanges cooperate, demand in the more expensive area can in part be met by supply from the low-cost market area until the price difference disappears. However, grid capacity between the two market areas may not be sufficient for price equalization. In this case, the participating power exchanges may aim at maximum possible price equalization by ensuring that power flows from the low-price area to the high-price one.
– Market splitting (nodal pricing): Grid bottlenecks may also occur within a single control area. They can be overcome by temporarily dividing the control area into separate market areas and ensuring that each of them has market prices that balance regional demand and supply. In the area with a high market price, customers pay a surcharge on the price that would prevail if the control area were integrated. This constitutes extra revenue for the generators. Conversely, customers in the area with a low market price benefit from a low price, while generators achieve less revenue. Eventually, the price differences incentivize investment in generation capacity in the high-cost area and investment in grid capacity between the low-cost and high-cost region, both alleviating future congestions. This model has been implemented in Scandinavia for years, ensuring that bottlenecks are managed efficiently by Nord Pool, the Scandinavian power exchange.

Implicit auctioning and market splitting make efficient handling of grid bottlenecks possible, suggesting that they are likely to become more common in future. However, they too fail to provide an answer to the question of how to create economic incentives for completely eliminating grid bottlenecks. In principle, a grid bottleneck hurts economic efficiency if investment in its removal is less costly than the present value of the price differences caused by it. As a result, grid operators have usually no reason to make such an investment (eliminating price differences and thus potential for arbitrage activities) unless the regulator provides them with appropriate incentives (e.g. granting a higher return on equity or exemptions from regulation imposing nondiscriminatory access to the grid).

References

Averch, H., & Johnson, L. (1962). Behavior of the firm under regulatory constraint. *American Economic Review, 52*, 1052–1069.

Beesley, M., & Littlechild, S. (1989). The regulation of privatized monopolies in the United Kingdom. *Rand Journal of Economics, 20*(3), 454–472.

David, P. A. (1987). Some new standards for the economics of standardization in the information age. In P. Dasgupta & P. Stoneman (Eds.), *Economic policy and technological performance* (pp. 206–239). Cambridge: Cambridge University Press.

Demsetz, H. (1968). Why regulate utilities? *Journal of Law and Economics, 11*(1), 55–65.

EU Commission. (2009a). Directive 2009a/72/EC of 13 July 2009a Concerning Common Rules for the Internal Market in Electricity and Repealing Directive 2003/54/EC. Brussels. *Official Journal of the European Union, L 211*, 55–93.

EU Commission. (2009b). Directive 2009b/73/EC of 13 July 2009b Concerning Common Rules for the Internal Market in Natural Gas and Repealing Directive 2003/55/EC. Brussels. *Official Journal of the European Union, L 211*, 94–136.

Hogan, W. W. (1993). Markets in real electric networks require reactive prices. *Energy Journal, 14*(3), 171–200.

Laffont, J. J., & Tirole, J. (1993). *A theory of incentives in procurement and regulation*. Cambridge, MA: MIT Press.

Müller, L. (2001). *Handbuch der Elektrizitätswirtschaft (Handbook of the electricity industry)*. Berlin: Springer.

Praktiknjo, A. (2013). *Sicherheit der Elektrizitätsversorgung. Das Spannungsfeld von Wirtschaftlichkeit und Umweltverträglichkeit (Security of electricity supply. The tension between economic efficiency and environmental compatibility)*. Wiesbaden: Springer.