GIS larsen

Laboratoire d'Analyse économique des Réseaux et des Systèmes Energétiques

Working Paper n°2

November 2006

Electricity and long term capacity adequacy. The quest of regulatory mechanism compatible with electricity market

Dominique Finon

Virginie Pignon

November 2006

Electricity and long term capacity adequacy. The quest for regulatory mechanism compatible with electricity market

Dominique Finon[1] & Virginie Pignon[2]

Summary : The paper deals with theoretical aspects of long-term electricity supply security. Market prices and contractual arrangements on the physical and financial electricity markets do not allow the creation of sufficient incentives to invest in adequate capacity for guaranteeing the appropriate level of supply in every circumstance. Long-term security of supply by capacity adequacy must be conceptualised as a public good. Alternative solutions to reach capacity adequacy having been adopted in different markets are successively considered: public procurement of strategic reserves, capacity payments, capacity obligations with tradable rights. Each presents theoretical limits and practical drawbacks when implemented in complex markets. That brings out the interest of a mechanism of centralised auctions for forward capacity contracts (or reliability options); it combines controls by quantity and by price while stabilising investment in peak power plants and is compatible with energy and reserves markets, which is not the case with the three other mechanisms.

[1] CIRED-CNRS and LARSEN
[2] EDF R&D

1. Introduction

Insufficient attention was paid to the issue of investment in generating capacity during the period of designing the competitive electricity reforms. The progressive erosion of excess generating capacity in industrialised countries, along with the ensuing crises that translated into power outages, raise questions regarding the ability of the electricity market to provide market actors with incentives to invest efficiently in terms of timeliness and equipment mix (IEA, 2002, 2003; Finon, 2006). Even though, in theory, a competitive market is assumed to send the correct price signals to encourage investment in new capacity, imperfections in the electricity market, short-term demand inelasticity, very specific price volatility, and incomplete market rules cast doubts on that presumed ability of electricity prices.

The risk of underinvestment is particularly acute in the case of peaking plants. Flexible to load following, they are also the best suited for coping with unanticipated variations in supply and demand, and are system operators' preferred means for ensuring the reliability of systems in real time. Little used and episodically called on, they present high profitability risks to investors. The development of peak units for purposes of safeguarding supply reliability in the long term raises the classical problem of a public good, the supply of which cannot be left to the interaction of private transactions and simple coordination by the market, as this leads to suboptimal investment with regard to the collective demand for reliability.

The problem of the reliability of supply encompasses two elements that can be conceptualised in terms of the provision of two public goods. On one hand, we have short-term reliability - or "the ability of the electric system to withstand sudden disturbances," in particular during peak hours - and on the other hand "capacity adequacy" - or "the ability of the electric system to supply the aggregated electrical demand and energy requirements of costumers at all times," - which conditions the supply of the former good (according to the working definitions of the NERC (2001). The first aspect is the dynamic side to reliability. It reflects system operators' technical balancing, which specifically relies on operating reserve services supplied by generators and the administrative power to ration by load shedding to prevent the collapse of the system.

The second aspect is investment in peak and base-load plants that allow growth in demand to be met and that provide reserve capacity margins adequate for dealing with extreme situations. These two public goods must be separated for three reasons. First, the supply of each good involves very different timeframes: decisions covering at least two years in the development of peak facilities on the one hand and intra-day, or even real time, decisions on the other. Second, short-term security is not only conditioned on past investment decisions, since several means (balancing, load shedding, emergency imports) are available to the system operator. Finally, the former can only be ensured in a centralised fashion by the system operator, while the latter depends on the investment decisions of decentralised agents.

Capacity adequacy - which could also be termed long-term security - is a public good since it is nonrival and nonexcludable.[1] This good is nonrival, since everybody benefits from the security provided to the system by new facilities provided by anyone: operators who invest in peak units that are never used do not earn any revenues, and yet they contribute to the security of the system by expanding the scope of potential action for ensuring balancing of supply and demand in random events, and thus provide a positive externality to operating reserves (Perez-Arriaga and Meseguer, 1997). This is a nonexcludable good since, as we shall see, it is impossible to personalise contractual arrangements governing long-term provision with respect to the preference of each consumer for supply reliability and their willing to pay.

Thus, it is necessary to define a regulatory mechanism that will ensure the necessary investment so that, in the medium term, demand can be satisfied by the supply of competitive producers in the broadest possible configurations of situations of high-load and generators availability. In the EU, To address this problem, the second European directive 2003/54 liberalising electricity markets proposes a method of direct intervention that consists of the regulator and the system operator issuing a call for tenders for the creation of reserve power facilities. Other mechanisms based on price control or on quantity control are possible, raising the issue of what criteria to prefer for implementing this type of mechanism in electricity markets, which are already complex enough.

In the following, we first examine the market imperfections that justify regulatory intervention, then we define selection and design criteria for capacity mechanisms in the complex environment of electricity markets in order to compare three classical mechanism types: public procurement of strategic reserves contracts, capacity payments, and capacity obligations with markets for exchangeable certificates, that ensure decentralised steering of capacity adequacy either by price or by quantity. The limitations of each of these lead us to analyse proposals for centralised mechanisms of forward capacity contracts or reliability options.

2. Why incentive mechanisms for investments ?

The requirement for new capacity raises the issue of how suitable regulations and market institutions are for providing the incentive for efficient investment. The risk of underinvestment is particularly acute in the case of peak power plants—which are, however, indispensable for satisfying hourly electricity needs in a maximum of random situations. These facilities have very short periods of operation, increasing the share of fixed costs in the total generation cost. Thus, their prospects for profitability depend on stochastic prices, which need to be extremely high for very few hours.[2] These prospects for profitability are reduced even more if government authorities impose a price cap out of fear of strategic pricing behaviour. Imperfections that are intrinsic to electricity markets then translate into earnings

[1] Among these two characteristics of pure public goods, non-rivalry and non-excludability, the latter is the determinant property that justifies regulatory intervention to correct sub optimal supply of this good in the private market owing to the indivisibility of benefits, motivating agents to adopt free-rider behaviour (Cornes and Todd, 1996; Lévèque, 1998).

[2] Thus, for a 250 €/kW cost of investment in a peak unit, amortized over 5 years, and a fuel cost of 10 c€/kWh, we obtain the following relationship between the mean number of hours of operation per year over the lifespan of the economic depreciation (5 years) and the price required to make the equipment profitable: 10 000 €/MWh for 5 hours per year, 1000 €/MWh for 50 hours per year, and at least 200 €/MWh if the duration is 250 hours.

shortfalls for peaking units, the "missing money" discussed by Stoft and Cramton (2006) and Joskow (2006), which ultimately exacerbate short-term relability problems. Finally, the price can also be biased by the out-of-market supply of ancillary services under the system operator' control, who seeks to ensure the flexibility required for balancing supply and demand without using efficiently existing market mechanisms (Joskow, 2006).

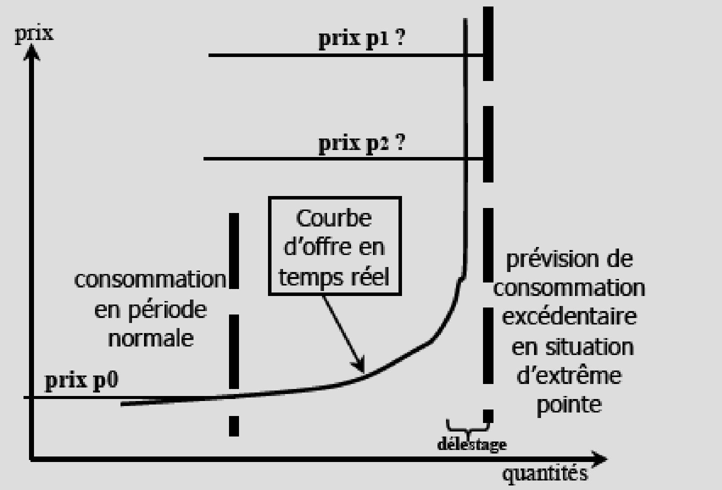
2.1. The limitations of price signals in energy markets

The market model is assumed to establish an optimal level of reliability on the basis of an instantaneous price-elastic demand function, which captures consumers' differentiated demand for supply and reliability and yields a price-quantity equilibrium. This stands in contrast to the old model of the regulated monopoly, which does not always have an equilibrium because of the inelasticity of the demand function. Reliability is obtained in extreme situation from administrated rationing, which is equivalent to considering the monopolist as confronting undifferentiated demands and assuming that all consumers suffer the same disutility from a disruption of supply, the value of lost load (VOLL). This simplification made it very easy for the utility to seek out a social optimum in terms of reliability while optimising its investment choices and managing the system in real time with regard to the danger of outage. In fact, policy makers identified and selected a probability of lost of load (LOLP) that equated the mathematical expectation of the VOLL with the long term marginal cost by an additional peak unit, after having determined the mean VOLL from polls. Subsequently, since all consumers are assumed to incur the same disutility from involuntary rationing, load shedding allowed an optimum to be achieved in the event of very unfavourable conditions in order to avoid black-out. We shall see that, in liberalised markets, the impossible build-up of a price-elastic demand function, the physical reality of electricity and market imperfections will lead us back to this approach, delegating long-term security responsibility to the transmission system operator (TSO) in surplus of its short term responsibility of the system reliability[1] .

Under "pure" competition with price elastic supply and demand, agents invest as a function of market signals, which reveal short-term generating costs and consumers' marginal propensity to pay. In the case of electricity, revenues are generated on the spot electricity market and associated markets, such as operating reserves and balancing markets on which the demand function issues from the system operator (Hunt, 2002; Chao and Huntington, 1998). However, a first market imperfection: since wholesale prices are not transmitted to the retail price in real time, owing to the unavailability of metering and pricing infrastructures to perform that on a large scale, instantaneous demand is price inelastic and this fact, along with the impossibility of storing electricity, causes prices to be very volatile. During periods of strain on the grid, physical market equilibrium may not exist (Figure 1). The dual inelasticity creates opportunities for strategic behaviour on behalf of any operator, of whatever size, when the system is constrained, thus complicating investment decisions, particularly in peaking units (Stoft, 2002; Pignon and Tarbé, 2003).

[1] We refer here to a transmission system operator instead of a simple independent system operator given the dominance of the first regime in the EU and the responsibility of the capacity adequacy that ideally it should have to share with the regulator.

Figure 1. The possible absence of market equilibrium in peak periods



Source : auteurs

Energy market proponents consider that the revenues generated during high load periods are adequate to attract investment into peak units. This reasoning postulates that prices must reach extremely high levels, ultimately set by demand and very far from the marginal cost (for example, 10 000 €/MWh, in contrast to a total cost of 100 €/MWh for a peak unit with a variable cost of 70€/MWh), over several price spikes, for investments to be profitable. This, however, raises two problems. On one hand, such high prices for a good like electricity are difficult to accept, socially and politically, since the transfer of surplus toward generators implied by them are perceived as very excessive during the periods of peak pricing. On the other hand, the timeframes required for licensing and for construction force agents to relevant anticipation of the price signal to avoid under capacity in peak and long periods with high probability of load shedding. The great volatility of prices and difficulties inherent in predicting long-term price trends make these calculations very complex. This creates a real problem of social acceptability (Besser et al., 2002; Hughes and Parece, 2002), leading regulators to opt for fixing a price cap: 1000$/MWh on North American regional markets, and 350 to 500€/MWh on certain European markets.

These price caps are subject to criticism, as they exacerbate the disincentive to investment in peak power plants by reducing the revenues generated by new base load and semi-base load equipment. In fact, the price cap placed on energy market to prevent the exercise of market power would equal the shadow price of the capacity constraint (J.P. Bouttes, 2005). Returning to our previous calculations, two observations follow. If we aim for a high level of reliability in the medium term (one or two days of power failure every ten years, or 2.5 to 5 hours per year in expectation), then the implicit value of the power failure (10 000 €/MWh) is higher than the price cap that is usually set. Conversely, if we suppose the value of disutility to a level of 1000 €/MWh, then the reliability level is one-tenth of what we just saw, (i.e. an average of some 50 hours of power outages per year).

One solution that can be envisaged involves recourse to complementary mechanisms designed to promote mainly the development of peak power plants. However, proponents of the point of view that "energy-only" markets and operating reserves markets should regulate all the capacity development in generation have several proposals for how to correct market imperfections and avoid resorting to these mechanisms.

2.2. Correcting market imperfections

To avoid resorting to complementary incentive mechanisms for investments in peak units, an accurate energy demand curve should be expressed in order to reveal the true marginal willingness to pay. However, this solution raises a number of technical and institutional difficulties.

• **Revelation of a price-elastic demand function.** Expressing such a function would necessarily require the transmission of wholesale prices changes on retail prices in real time, implying real-time metering of consumption.[1] Consumers would ration their electricity consumption as a function of prices, as they do in the case of other goods and services characterised by temporary shortages (foodstuffs, transportation, housing, etc.). The need for reserve capacities would be reduced, along with the potential for exercising market power. This price elastic demand function would allow to achieving market equilibrium during peak periods in an optimal generation mix (Cases 1 and 2 in Figure 2). In the event of an increase in variable and stochastic demand, prices would adjust to maintain this equilibrium.

In order to manage price fluctuations for risk-averse consumers, some also propose developing supply contracts that provide to producers an incentive to guarantee the supply of energy to consumers up to a price that is a function of each individual's requirement for energy reliability. There are two possible paths. The first is based on capacity subscription contracts under which the consumer is supplied with power until a remote circuit breaker cuts off the current when the market price surpasses the energy price contractually negotiated between the supplier and consumer.

The second uses purely financial contracts between suppliers and consumers which incite in turn suppliers to establish option contracts with producers (Oren, 2001, 2003). The options allow the last ones to pay the market price in real time up to a certain limit. Beyond that, they can exercise the option and are reimbursed for the difference by their counterparty. The premium for the options paid to suppliers is all the more expensive than the guarantee offered by the insurance is high, i.e. the level of the strike price chosen by the consumer is low. Finally, these financial contracts must promote investments in generating capacity: The premiums reduce the volatility of suppliers' incomes and provide them with an overview of consumers' marginal willingness to pay. If there is insufficient physical investment, market prices of energy must increase, which may in turn imperil the financial equilibrium of this type of contract for suppliers having signed them. This acts as an incentive to contract equivalent option contracts with producers which in turn proceed with investments in physical capacity to hedge their risk.

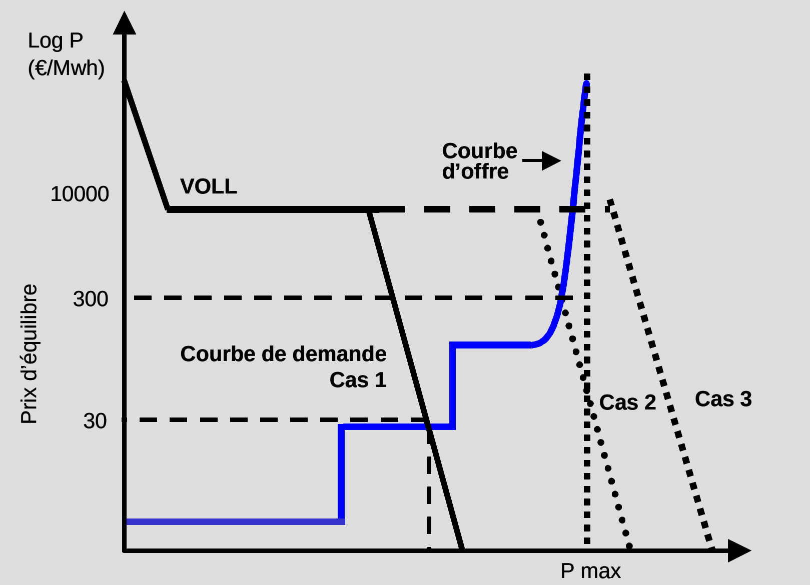
[1]We must clearly distinguish between this real-time pricing and suppliers' programs to curb demand. These pay eligible consumers to curtail their demand during peak periods at their opportunity cost, at the price it would cost to supply them with one kWh at the margin during extreme peaks and under constraint (for example 500 €/MWh, while the full cost of a kWh from a peak unit is 100€/MWh). These programs mostly involve industrial purchasers who are inclined to accept this amount to withdraw: a disposition that is not symmetric to the willingness to pay. Unlike voluntary price-based rationing, these curtailment programs impose a cost on all consumers. They are to be treated as supplementary resource suppliers for dealing with peak demands.

Currently, however, these solutions are scarcely feasible. They need: to be extended to the whole of consumers. They would necessitate establishing a very complex control infrastructure and generalising real-time metering. Assuming they were to become possible, the learning curve and social acceptance remain problematic. On the economic level, these approaches imply significant transaction costs, especially in the case of mass consumers, with the proliferation of risk-management mechanisms and the reinforcement of the role of intermediaries.[1]

Furthermore, the provision for energy subscription contracts has no physical basis. Suppliers are no longer responsible for the physical provision of electricity, the distribution network is assumed separate, and the technical capability to disconnect is in the hands of distribution and transmission system operators. Also, in the new institutional contexts of separating supply and network activities, the implementation of power subscription contracts necessitates that suppliers inform the distributor of signed contracts, further increasing transaction costs.

This does not diminish the interest in seeking a real-time price-elastic demand function for part of electricity consumers. Hobbs, Inon and Stoft (2001) reveal that, in an oligopolistic market, we can achieve a better outcome with no price cap and in which demand is rendered partially price elastic than with a price cap of 1000$/MWh in a context of inelastic demand when it is necessary to complete the market by adding a capacity mechanism.

Fig. 2. Equilibre en pointe dans trois cas de courbe de demande semi-élastique



Source : authors

Nonetheless, demand elasticity in a market segment does not eliminate the risk of outages, as is revealed by the formalisation of a two-segment market - one with real-time pricing and the other with flat prices - by Joskow & Tirole (2004).

[1] See, especially, the study by Borenstein (2004) on the impact of different scenarios of development of real-time pricing on the California market.

Most of the real-time demand function remains price inelastic. If equipment failures combine with inclement weather to boost hourly consumption during peak periods (case 3 in figure 2), prices will rise to the VOLL level and the TSO will continue to need to administer rationing.

• **Improving the demand function for reserves and balancing[1].** For the producer, these services amount to a commitment to make available a plant for operation or to curtail generation by order of the TSO. They are sold on a specific market on which the TSO is the only demander of services. Contributing facilities are, in particular, peaking units that thus significantly top up the revenues yielded by the energy market. Using only this way of revenues for peaking units poses two difficulties. First, the system operator's demand for reserves is very price inelastic, leading to high price volatility that distorts the revenue forecasts of potential investors. Also, the TSO will be inclined to resort to off-market sources of generation to ensure physical balancing, thus impeding the revelation of the true price of operating reserves (Joskow, 2006).

The stake centres on the long-term capacity of the energy-only market to organise the conditions that will guarantee supply in any situation of generation availability and high load, with a very low probably of outage. As to reliability, price levels, and price volatility, it appears difficult at present to substitute signals from energy and operating reserves markets for benevolent monopolistic programming of the former electric industries. If public authorities wish to retain these price and reliability levels, we must return to the same procedure considering an uniform demand for short- and long-term reliability by the regulated monopolies. The TSO's responsibility for balancing the system must be complemented by the definition, by the government or the regulatory bodies, of a responsibility to ensure "capacity adequacy" and the means for acting on this responsibility, in order to incite decentralised players to act in the same direction.

2.3. Design principles for a capacity mechanism

Government bodies need to establish a regulatory mechanism to provide agents with incentives to maintain capacity adequacy. Such a regulatory mechanism is then introduced into a market whose rules are already complex, with prices that are intrinsically volatile and, we may speculate, subject to strategic behaviour. The choice and design of the mechanism must account for this reality. Drawing on various comparative analyses (Perez-Arriaga, 2001; De Vries, 2004; Oren, 2005; Cramton and Stoft, 2006; Joskow, 2006), several issues arise.

- First, the classical issue of public economics in an environment of uncertainty on the supply and demand curves of a public good (Weitzman, 1974): Should preference be given to controlling the supply of this good by price (a capacity price is added to the energy price) or by quantity (the imposition of an obligation to contract

[1] We must clearly distinguish between capacity reserves and operating reserves. The first include capacities that are built to confront randomness in generation and demand in high-load periods (peak units, old power plants) and are an element of capacity adequacy. The second include system services provided by generators and managed under the technical authority of the system operator. Part of the supply of operating reserves can be ensured by a market mechanism with competitive tendering of upward and downward adjustments by market players In particular, the tertiary reserve includes a series of means that the system operator can mobilise in timeframes varying from 30 minutes to half a day to top up secondary reserves that can be mobilised on very short notice.

on a capacity level independent of any given level of generation) to minimise regret in terms of social surplus loss if an error in the marginal cost function or the marginal benefit function ?

- Since the issue is to incite investment in peaking units, how should the revenues of investors or entrants be stabilised or foreseeable with this mechanism ?

- In light of the interdependence between the energy and the capacity in the identification of the value of generating capacity, can we isolate the treatment of a capacity-good from the markets for energy and reserves ?

- How to design a mechanism so that it does not open new ways of strategic behaviours? Does it help to reduce incentives to exert market power on the energy market on the other side ?

- And finally, in an open market how robust the mechanism will be to interactions with the neighbour systems having a different approach of capacity adequacy when submitted to shortage ?

• **The efficiency of capacity adequacy targeting.** We must first assess the mechanism’s capacity to guide the electrical system toward the level of reliability desired by the public authorities. In the classical dilemma of “price controls versus quantity controls” in Public Economics, the results in Weitzman (1974) for situations of uncertainty on supply and demand curves favour the choice of the quantity-mechanism with steep demand curve or flat cost curve. Indeed, when the demand curve is steep, as it is here where marginal damages increase rapidly with the level of outages, it is preferable to control quantities, i.e., to choose a level of reserves capacity to be attained. The same result obtains when the supply function is flat, which is also the case here since the marginal cost of capacity is the capital cost for a peaking unit, which is added to the capacities already in place. A small error in fixing the capacity price could result in a huge difference in installed capacity in particular towards deficit. Conversely, too much capacity during peaks has a relatively small impact on the surplus.

Public economics also proposes hybrid mechanisms. These combine control by quantity (quotas with tradable certificates) and control by price (price cap) to control the costs of complying with quota obligations. In an environment of uncertainty in costs and benefits (avoided damages), this type of mechanism allows better control of forgone social surplus in the event of error than either of the two aforementioned pure mechanisms (Roberts and Spence, 1976).[1]

[1] We must underline that Weitzman results and Roberts & Spence ones are refereed to the supply of pure collective goods completely independent of any other goods, that is not the case of the capacity-good here. Capacity and energy are linked products supplied by any generation equipment. In concrete terms infra-marginal surplus on the energy market during price spike constitutes a part of the economic value of the equipment. Other example producers could arbitrate between bidding on operating reserve market (the main service of capacity that can offer any generation units) and energy market. So Weitzman and Spence results must only help to establish some fruitful hypothesis. But complete separation of capacity and energy are misleading, as it will see in the case of capacity obligation mechanism in which in the original design the demand of “capacity” was completely inelastic, without relation with tight supply situations.

- **Securing the profitability of new peaking units.** The purpose of a capacity mechanism is to provide an incentive to invest in peak capacity and to complete the profitability of base-load equipment in order to start their investment. The revenues jointly yielded by the energy market and the capacity mechanism must be foreseeable, at least over the capital recovery period after the equipment has been installed. This presents a two-fold problem. The first aspect involves the duration, i.e., the stability of the investment in peaking units by four-to-five year commitments that enter into effect after the equipment is in place (two years later). The second aspect involves smoothing the revenues generated by the peak unit by limiting the investor's price-risk on the energy market and ensuring complementary compensation defined *ex ante*.

- **The consistency of the mechanism with the energy market**. We are not dealing here with a market for a single commodity governed by clear property rights. Electricity markets are complex and encompass various complementary levels (energy, operating reserve service, congestion capacity). Markets for energy and reserves will also be affected by the use of a capacity mechanism, if only through the new opportunities for arbitrage it creates between goods and services. Similarly, the cost of investing in peak power plants is compensated and rendered profitable by the surpluses generated by the hourly market for energy and operating reserves,[1] and not only by compensation for the public good of capacity adequacy. Thus, there are traps to avoid in the design of the mechanism. First, the supplementary compensation that the capacity mechanism yields must not doubly pay for the cost of the equipment, which would ultimately be passed through in the price to the consumer. It should, in particular, vary with the level of surpluses on the hourly energy and reserves markets, which constitute the basis for the revenues of peak units. In other words, when prices are very high on the energy market and reflects stochastic scarcity, the compensation yielded by the mechanism must subtract increased revenues from the energy market. Second the capacity mechanism must not distort generating companies' investment decisions (with respect to the size and the type of generator) and the formation of prices on the energy market. Thus, a mechanism designed without understanding the interdependence between the capacity-good and the energy-good is liable to lead to excess capacity during peaks, and consequently depress energy prices and limit surpluses on the energy market. We must ensure that the compensation to capacities interacts with the movements in energy prices.

- **Robustness to the exercise of strategic behaviour** A capacity mechanism is a set of supplementary rules that is added to the already complex rules of electricity markets. Thus, new opportunities for strategic interactions between actors can be elicited. A mechanism must also be judged by its transparency and reduced barriers to entry.

- **Institutional feasibility**. The design of each capacity mechanism must be consistent with the principal rules of electricity markets. For example, the system of bilateral options contracts proposed by S. Oren (see above) cannot be established without the prior dissemination of real time pricing. For another example, an mechanism that directly adds compensation to all kWh generated by producers is much less easily integrated into a non-mandatory market exchange than into an mandatory market pool.

[1] In theoretical economics this is called "inframarginal rent," since it is generated by equipment that is used prior to the marginal equipment that sets the equilibrium price on the competitive market.

- **Effectiveness of the mechanism in an open market.** The collective good approach of capacity adequacy is conceived for a closed market. But each market is more or less in interaction with its neighbour markets. So the effectiveness of the mechanism in an open market must provide a means for preserving its benefits in terms of reliability and lower prices during high load and tight supply periods, in relation to possibilities of arbitrage and electricity exports towards neighbouring systems when they are short of capacity. The least to say is that each mechanism would need to have an equivalent mechanism which copes with the capacity supply in neighbouring systems, unless national or regional regulation would allow to temporarily curtail exports during a crisis. It is the case between the EU countries which have this possibility allowed by the EU legislation (Art. 24, Directive 2003/54/EC).

3. A comparison of classical capacity mechanisms

By its very existence, the capacity mechanism is premised on the public authority (government or regulator) having assumed responsibility for the capacity adequacy and reliability of long-term supply. As a preliminary condition, it organises a coordination of expectations with the TSO through a forecast of long-term peak hourly supply and demand. On the basis of identified capacity requirements, and without needing to await the spontaneous decisions of agents who are assumed well informed by the aforementioned anticipation, it has the choice between three principles of action : reserve capacity procurement directly commanded by the TSO, price action based on the mean value of power outages or on the cost of a marginal peak unit, or quantity action based on a capacity obligation with suppliers, which is completed by a capacity certificate trading procedure for capacity.[1]

3.1. Procurement for long-term strategic reserves contracting

The simple way for the supply of a collective good is for the government to provide it and for all beneficiaries to pay for it through a targeted tax. In our case, two solutions can be envisaged: long-term capacity reserve contracting by the system operator, which may be complemented or substituted by direct installation by the TSO of peaking units. However, this centralised acting is criticised because it distorts the natural functioning of the energy market and the incentives to invest in peaking units[2].

[1] S. Stoft (2002) has studies a fourth mechanism, ***the operating reserve pricing***: it consists to use the operating reserves market in a systematic way to give a stable surplus to the peaking units, and in particular, the new ones. The TSO purchases reserve capacity services on a daily basis. It purchases more than it is needed for short term operations alone. The extra reserves will help to sustain the operating reserve price and will give increased revenues to old and new peak units, that provides an incentive to producers to create more production capacities. The TSO is only willing to pay up to a certain maximum reserve price and will adjust its extra reserve demand to the situation of the load and the price on the energy market. That means that it reduces its extra reserve demand during high loads. So the extra reserve capacity which is contracted during off peak on a daily basis is available to produce during peak demand. The necessary condition this system be effective is the long term commitment of the TSO on the maintenance of the maximum reserve price.

[2] This approach is different of the short term standing reserve contracting used in a number of countries (UK, Sweden, Norway, etc). The TSO is allowed to auction short term option contracts in order to be guaranteed against the risk to have insufficient reserve margin in real time and to have sufficient available capacities. Given the difference of focus, the major

- **The design of the mechanism.** This approach was adopted by some European countries and is explicitly mentioned in the texts of the European Commission, in particular the second Directive of 2003 on electricity market liberalisation (Art. 7.1) and the Directive of 2006 on the security of the electricity supply. The principle is that the system operator contracts for the right to mobilise a capacity, which the generator commits to make available or to make run for bidding on the market when specifically asked for it, by signing a contract with this latter that covers the medium and long term. Thus, the law creates a right to dispose of "strategic reserves," which the TSO can mobilise to satisfy hourly energy demand and the requirements for operating reserves during periods of system stress. The TSO can launch this programme after the public authority (ministry, regulator) have identified the need for reserve capacity. The level of this contractual capacity reserve is fixed by it, the minister, or the regulator. Market-based selection by calls for tenders is made, as in New Zealand, France, or Sweden. In the first two cases, contracts are backed to new units to be installed as security. In the third case, candidates are not obligated to link their commitments to new capacities, this is possible because of the maturity of the system.

The long-term contract provides annual compensation for capacity based on the marginal bidding price of the selected candidates, regardless of the facility's actual output. Additional payment, in the event that the peak unit is called on and has variable costs, depends on the design of the contract. The simplest solution is to limit the TSO's claims to a right to dispose of the capacity but not of the energy. In this case, the peak facility must produce and sell energy on the exchange, where it is remunerated at the market price, which will be high in these situations of tight supplies. The other situation involves the TSO buying kWh directly, either at the marginal cost or at the hourly market price, and to dispatch them out of merit at a price of zero, the second way being more incentive to not distort the market functioning. In both cases, the costs the TSO incurs will be reimbursed by an uplift imposed on the transmission price of all kWh transmitted.

The constitution of a contractual strategic reserve can be complemented by two other, much more direct, means which are the most well-known and the most criticised, but not the most developed as long term reserves. The TSO can take over old units that the owners have decided to close. However, owing to their inflexibility in terms of responding to real time balancing needs, these facilities are limited to coping with inter-seasonal variations, especially in systems that are largely hydroelectric. The TSO can also purchase peak units and operate them itself. These facilities will then complement the means provided by strategic reserves contracts. This approach, provided for in the legislation in Sweden, Norway and Finland, is far from market principles somewhat. It can only be invoked provisionally in order to enable rapid response to threats to the security of supply.

- **Evaluation of the mechanism.** This mechanism shows a high degree of effectiveness with respect to the goal of long-term security of supply, but it contravenes market principles.

---

differences are the absence of relation to specific new equipments or old ones to be managed in seasonal de-mothballing, and the short duration of the contracts. The TSO defines each year a level of operating reserve that he wants to be sure to call up on the operating reserve market with the help of a system of reliability contracts of different terms (one month to one year). If a generator is selected, he is paid at the marginal bid price for the capacity service he will offer, plus the operating reserve price when it will bid on this market.

- *The efficiency of adequacy targeting.* This type of provision guarantees that the peak-capacity reserve ratio desired by the TSO will be met. It safeguards investments in peak capacity by ensuring that the investor's costs will be recovered over the life of the contract. In acting in a centralised and direct way on quantity, the target defined by the public authority (on the advice of the TSO) will probably tend to overestimate growth in consumption and reserve capacity requirements owing to the technical risk aversion of the TSO and the public authority, who will seek to reduce as much as possible the likelihood of power outages. However, this can be justified from an economic perspective to the extent that the supplementary cost associated with a power outage exceeds the supplementary cost of sporadically excess capacity.

- *Compatibility with the energy market.* By its very nature, this provision is not intrinsically compatible with the interplay of competitive forces on the markets for energy and for operating reserves for three reasons. First, there is a risk that the regulator (and the TSO as his adviser) behave prudently by anticipating on mid term a need of new peak units to comfort the reserve margin in a context of uncertainty of the electricity demand growth. So, a new requisite for the use of this policy is an efficient anticipation of capacity needed to complement the reserve margin (see below). Second the administered introduction of peaking units may also affect decentralised decisions to invest in base- and semi-base load equipments, since generation by these equipments may limit the prices of energy and operating reserves during peaks and reduce infra-marginal rents to them (Joskow and Tirole, 2004). Third, bringing the energy generated by these facilities to market directly distorts the functioning of the market if the provision does not include safeguards against too large discretionary intervention by the TSO.

▪ An ideal form of protection of the energy market functioning would be the definition of a energy price ceiling, above which the TSO would be authorised to require that the contracted facilities bid on the energy market (and so generates), or to supply operating reserve (or balancing) service on the reserve markets, i.e. to guarantee availability in case of shortages (e.g., it could be around 300 €/MWh up to the marginal cost of the last peak units which is around 70€/MWh). This trigger price will have a function of price cap on the energy market and will be by this way the price which will pay the kWh produced by the strategic reserves.

▪ An alternative form of controlling the TSO intervention is to restrict the call to strategic reserve to a level below which reserve margin must not decrease without dramatically increasing the probability of black-out. For that purpose the regulation should specify the exceptional physical conditions of the system under which these units can be called on for reserves service or to bid on the energy market.

Such a protection is difficult to define in an optimal way, as shown by DeVries (2004).[1] In the absence of these two types of protection, the uncertainty created by

[1]The regulator with the TSO must define the optimal volume of generating capacity in relation to extreme situation (load, availability), the optimal level of the energy price to trigger the call up to strategic reserve and then, by guessing the capacity to be developed by the market with this trigger price, to deduct the level of the strategic reserve. The regulator would have to fix the trigger price, then to let the market to determine spontaneously the amount of capacity to be present during high load, then after to decide the amount of the strategic reserve needed to respect the reserve margin ratio in relation to the optimal volume of generating capacity. The main difficulty is that these parameters are interrelated and not independent for the optimality of the system. So this lets no way to define the optimal couple of parameters,

the possibility of calling on the strategic reserve facilities may dissuade investment in peak power plants. There is a high risk that the actors will expect the TSO to behave "prudently"—i.e. to call on these available strategic reserve facilities in times of tight capacities in the system (or of congestion) in order to avoid technical imbalance. This intervention will indirectly affect the market price and the revenue of equipments during peak. But even with such a protection which curbs the intervention of the TSO, the mechanism affects the revenues of the different equipment by reducing infra-marginal rents during peak period whereas the price could climb above the price limit of recourse to the strategic reserve. It is the same critic which is addressed to the use of price cap.

- *Robustness to the exercise of strategic behaviour.* This provision may curb strategies to restrict investment in peak supply. However, since actors are able to foresee the actions of the TSO, we can show that the latter cannot completely eliminate this behaviour and bring about a market optimum (Meunier and Finon, 2006). This mechanism can also generate a windfall for risk-averse operators who would have been inclined to invest in peak equipment even without this provision. They have an incentive to await these calls for tenders to benefit from the income security they ensure. The efficiency of the system is thus closely linked to the conditions of the call for tenders. The long-term competition envisaged by the competitive reforms thus tends to mutate into competition for public contracts.

- *Institutional feasibility.* This mechanism is adaptable to all market types, non-mandatory markets as in the majority of European markets, and eventually mandatory or semi-mandatory markets as in the United States. The fact that it does not require creating any new regulatory provisions, and it is easy to implement, is doubtlessly its principal benefit. However its introduction will be more consistent with a TSO which owns transmission infrastructures and could add directly or contractually reserve units in its assets than an ISO which does not own any one.

- *Effectiveness against regional shortage in an open market.* As shown by De Vries (2004), it is not robust against regional energy shortages in an open decentralized market. Scarcity in a neighbouring system will also lead to high prices in the system and the reserve margin will be the same in the two systems. When the spot price is higher because a tighter situation in the other system, there is an additional demand addressed on the energy market in the first system and the generators will be incited to sell on the energy market rather than to stay in the operating reserve market. The TSO will be obliged to call up strategic reserve sooner than in a situation of isolated system when the energy price arrived to the trigger price.

To summarise, this mechanism allows direct steering toward a reserve capacity target and ensures stability in investment in peak units. However, it casts doubts on the neutrality of the system operator and introduces distortions into the functioning of markets for energy and reserves. If implemented efficiently, it allows close scrutiny of the TSO's decisions to issue calls for tenders (or directly construct peak units) and strictly encompasses how facilities are called up to produce. The regulator must complete the provision with safeguards to guarantee freedom of action to actors on the hourly energy market, that is generally not clearly established by the regulation.

---

the trigger price and the level of reserve capacity, which depends on the anticipation of capacity level developed by the market.

3.2. The capacity payment

In this approach, which is used by a number of countries (Spain, Argentina, Chile, Colombia, Peru, and, until 2001, the Anglo-Welsh market), explicit reference is made to a reliability optimum that the ideal market would generate and that would equate, in mathematical expectations, the shadow price of the capacity constraint in situations of tight supply and system strain with the cost of the marginal peaking unit to be installed - as the regulated monopoly would do at the optimum. Here, a capacity price is fixed centrally so that decentralised capacity decisions would lead to this optimum.

- **The design of the mechanism.** The price mechanism must be defined to yield a compensation, especially during high-load periods, in which the danger of an hourly supply-demand deficit is greatest. Depending on the specific provisions at work, the price of capacity is either fixed *ex ante* by the regulator in reference to the cost of a peaking unit and ignoring the VOLL, or fixed *ex post* in a flexible manner in reference to the average disutility of the consumers according to the probability of lost of load estimated after the realisation of the hourly market.

The peak capacity level is assumed to adjust to this long-term price by the installation of peak power plants thanks to the incentives provided by this additional compensation. It also features the benefit of providing an incentive to generators to declare themselves available during periods of strain on the grid, even though - unlike with the other mechanisms - there is no commitment or control over the availability of equipment at these times. In its optimal formulation, the equilibrium capacity price should be virtually nil during hours of excess capacity, while it should rise during periods of scarcity when reserve margins are low, providing a greater double incentive for generators to declare their capacities available and to invest in peak units. The mechanism also contributes to revenues from base-load and semi-base load equipment and thus to initiating earlier investment decisions.

- **Evaluation of the mechanism.** This mechanism presents difficulties, both theoretical and practical, making it a less than ideal tool.

\- *The efficiency of adequacy targeting.* Ideally, producers add peaking capacity to their generation equipments when the sum of expected market revenues and capacity payments is greater than the cost of installing a new peaking unit. The price is fixed as a function of the regulator's estimate of the intersection of the cost and demand curves for capacity adequacy and long-term security, the supply of which is its responsibility. However, there are two reasons why a high degree of security remains elusive: First, the choice of level to set for any price-mechanism in a situation of uncertainty on demand and supply curves for the collective good presents a risk of error (Weitzman, 1974). Nothing guarantees that installed quantities will adjust to the desired level of capacity. Furthermore, given the moral hazard vis-à-vis the regulator, decentralised agents may continue to underinvest in peak units so as to simultaneously benefit from capacity payments and extreme premiums on the energy market. Thus, in the case of Spain, there is no proof that it has resulted in new facilities having been built (Rivier, Vasquez and Perez-Arriaga, 2006). The mechanism only encourages older facilities, which would have been closed in the absence of capacity payments, to be kept online.

- *Compatibility with energy market.* This depends on how the mechanism is implemented. To approach this principle of compatibility, we must adopt a rule for computing the market clearing price, as in the mechanisms provided for on the mandatory Anglo-Welsh market until 2001. The per-facility compensation to available generating capacities must be a function of the probability of lost of load determined by the difference between real hourly consumption and available capacity notified on the mandatory market. The capacity payment, CP, is computed *ex post* hour by hour, and equated the expected social gain of avoiding lost of load minus the expected revenue on the market *SMP*:

$$CP = LOLP \times (VOLL\text{-}SMP) - (1\text{-}LOLP) \times SMP = LOLP \times VOLL - SMP$$

where *LOLP* is the probability of outage computed ex-post as a function of generators' hourly notified available capacities and the actual consumption of energy, *VOLL* is the value of outage as fixed by the regulator, and *SMP* is the equilibrium price on the energy market at J–1.

The modalities for implementing capacity payments in other countries (Spain, Argentina, Chile, Colombia, Peru) consist of defining the capacity price *ex ante* in relation to an average availability of the equipment technology without accounting for the hourly supply-demand situation. The capacity price is thus set in relation to the investment cost of a peak unit. This creates a situation in which the link with hourly market prices reflecting the system state is tenuous. However in some countries, in which the bulk of generation is hydroelectric, the value of the compensation is adjusted in light of seasonal and annual variability of hydraulic supply.[1] The only exception is the Italian system, in which compensation to capacity is partially adjusted *ex post* to market prices.[2]

- *Robustness to the risk of strategic behaviour.* This is closely linked to the conditions for implementing capacity payments. The English design was criticised because it made possible manipulation of the market rules, since the results of the computations were predictable. Experience revealed that the dominant operators effectively succeeded in driving up the capacity price by declaring themselves available, while simultaneously bidding inadequate hourly offers in order to increase the probability of lost of load, which in turn determined the hourly price of capacity (Newbery, 2006). Thus, the recommendation is to fix the price with an *ex ante* calculation, which presents the disadvantage of not accounting for the hourly supply-demand situation, but has the advantage of being predictable and not exposed to risk of gaming with rules.

- *Institutional feasibility.* This mechanism, which directly adds a compensation to all generators' kWh, is less difficult to integrate into a mandatory, or semi-mandatory, market than into a non-mandatory one. Conversely, it features more institutional difficulties than other mechanisms based on control by quantity. It is, in fact, a

[1] We should mention a practical criticism of the use of a capacity price that fluctuates in real time with the hourly supply-demand situation—its irrelevance to systems that are dominated by hydroelectricity. During wet years, the capacity term will be small and yield a very small compensation to thermal equipment developed to partially meet the needs of dry years

[2] In France, a proposal was advocated by one of the entrant in the supply (Choné, 2004) to add a capacity price dependent on the annual period, to the price per kWh. However, the difficulty with that suggestion was in implementing it in a mandatory market on which only a small proportion of transactions are on the power exchange (cf. institutional feasibility, below).

mechanism that directly generates rents for established generators, since it compensates existing equipment that is already amortised in the same way as new equipment. And, as is the case for any mechanism of public policy that yields such a rent, it is also vulnerable to rent-seeking behaviour (Stigler, 1966). This consideration could, in fact, play a role in the difficulty in reforming the mechanism and in its perpetuation under the influence of generators seeking rents.[1] The debate of the Spanish provision is of some interest from this perspective. The assessment by Batlle et al. (2006) reveals that, owing to the magnitude of the rents at stake, it is very difficult to reform or abandon. This can be explained by the design of the mechanism, which did not seek only to provide a supplementary incentive to invest in peak facilities, but also to provide compensation for "sunk costs" in existing equipment. Quantity mechanism based on a capacity obligation does not present such a risk of rent-seeking.

- *Effectiveness against regional shortage in an open market.* As for the former mechanism, during a regional shortage, this mechanism provides no means to protect the capacity-good developed in one market. Energy price increases will be the same as in the neighbour markets and the reserve margins will be adjusted in the same way.

3.3. Capacity obligation with exchangeable rights

This mechanism, used in regional electricity markets in the USA, is based on decentralisation by quantities. A capacity obligation is imposed on the main suppliers, those known as load servicing entities in the US markets. These suppliers are treated as representatives of consumers and their reliability requirements.

- **The design of the mechanism.** As with the strategic reserves mechanism, public authorities (the minister or the regulator, in collaboration with the system operator) fix a reserve margin as a function of the supposed average value of lost of load. Each supplier must demonstrate to the regulator that it has secured its capacity to deliver its high load demand plus a reserve ratio that is fixed administratively, by its own capacity and long term contracts.[2] To ensure the flexibility of these provisions, those bound by them can trade capacity credits. The price of capacity will be determined by the trading of capacity credits between suppliers in the control area. This mechanism is complemented with a penalty fixed higher than the investment cost of a peak unit, to be imposed on the suipplier if it does not meet its commitment or if the producer committed by contract with a supplier is unavailable. This penalty increases with the shortfall in generating capacity and reserves, as in the PJM market.

In practice, the level of the obligation must be fixed first. For this, system operators use peak demand forecasts from suppliers in their control area. Subsequently, obligations must be imposed on a type of agents, to wit, suppliers. It is also necessary to impose credible capacity-right commitments on generators: Producers

[1] In a principal-agent analysis, D. Perrot and V. Pignon (2005) compare possible efforts to capture rents between capacity mechanisms, on the basis of the strategic reserves mechanism and capacity payments mechanism

[2] This is physical, rather than financial, securing as in the provision proposed by Oren (2002) for option contracts between generators and suppliers (representing consumers) since, through this commitment, suppliers are not ensuring a future purchase price of energy, but rather the existence of a supply capacity that must be adequate to cover their demand in all situations, and thus in high peak.

who are bound to the obligated suppliers must be in a position to maintain their units in operating reserves or to generate electricity if the TSO requests it, otherwise they will be penalised. When they generate, the electricity is sold on the hourly market price. Finally, we must design the mechanism for trading rights (bilateral transactions, organised markets on energy and futures).

Capacity credits correspond to a generator's commitment to supply reserve services or electricity on energy markets in response to the system operator's request during situations of shortage in the system. In other words, if the energy market is in a position of shortage or near-shortage in terms of hourly supply, the system operator can impose an obligation on generators having sold capacity credits to supply the corresponding energy by bidding on the day-ahead market.

- **Evaluation of the capacity obligation.** This provision has been the subject of a vast literature in the United States, owing to the federal regulator's former plan to impose it on all regional markets, along with the Standard Market Design (cf. in particular, FERC, 2002). It presents limitations that essentially spring from how it has been initially implemented in practice, especially the limited duration of capacity certificates and the design of a capacity obligation that is totally disconnected from the energy market.

- *Efficiency of adequacy targeting.* As an mechanism that acts on quantities, this would be expected to ensure satisfactory steering toward the desired peak capacity and the adequate reserve ratio. However, in concrete applications in the three U.S. regional markets, it would be difficult to document the impact of current provisions on investments in peak units, owing to imperfections in the design of the mechanism (Oren, 2005; Cramton and Stoft, 2006; Pignon, 2006). In fact, for an investor in peaking units, this provision does not allow for clear forecasts of the capacity price, which is highly volatile on markets for capacity certificates, owing to the fact that the supply and the demand of capacity certificates are inelastic to the future price. Neither does it allow a price to be guaranteed over a time-span compatible with the amortisation of a peaking unit from its entry into service.

The supply of certificates is inelastic because of the unpredictability of capacity certificates: initial maturity too early, limited duration of fixed-price certificates (one week, one year), characters justified by the need of liquidity. Ideally, for investors in peak units and new entrants, some rights should be tradable in advance, i.e. for the market two years ahead, so as to accommodate the timeframe for building and authorising a peak facility (Joskow, 2006). Furthermore, the period covered by this type of mechanism (12 months maximum in PJM market) is criticised by agents as too short in light of the payback periods for investments in peak units (3 to 5 years).

Demand is inelastic, being totally defined by the fixed quantity suppliers are compelled to secure on the one hand, and by the fixed penalty level if it is not secured on the other. When the supply-demand equilibrium becomes taut, the value of the certificates suddenly jumps to the level of the penalty, signalling to generators the need for new peak units. In the opposite situation, when capacity constraints are relaxed, the supply of credits is greater than the demand by suppliers and the equilibrium price approaches zero. The price-signal for capacity credits is thus highly volatile, owing to the structure of the mechanism.

If we establish a more suitable right with much longer maturity, it is also important that the provision be flexible with regard to other, more short-term, tradable rights, so suppliers will be able to contend with movements in their clientele. There is a difficult arbitrage inherent in this provision.[1]

- *Compatibility of the mechanism with the energy market.* The capacity right is totally separated from the energy-good. The value that the capacity market yields is disconnected from the price of the energy markets. From another perspective, when a scarcity occurs and the price of energy rises to high levels and yields very important infra-marginal rents, which will contribute to recovering most of the cost of peak power plants, the price of energy has no impact on the value of capacity rights, leaving them overvalued.

- *Exposure to the risk of strategic behaviour and entry barriers.* On the capacity market, inelastic demand for rights creates a danger of strategic behaviour. However, when all of a generator's capacity is under contract, the level of the penalty (1.5 to 2 times the cost of the marginal facility) may provide an incentive to sell the corresponding energy on the market. Even greater limitations exist: Owing to the relatively short maturity of capacity certificates, the mechanism may provide little incentive to new entrants. Furthermore, as in the case of capacity payments, it favours incumbents by failing to differentiate between generating capacity that is already amortised and new construction.

- *Feasibility of the mechanism.* This mechanism is applicable to all market types, since it acts on suppliers in a decentralised fashion. Conversely, relative to other mechanisms, it presents greater difficulty. Implementation requires defining a trading mechanism and imposing new obligations on suppliers, including pure suppliers who may be dissuaded from entry or motivated to change their business plan.

- *Effectiveness against regional shortage in an open market.* As the two preceding instruments, this mechanism presents an explicit risk of capacity rights "leakage" into neighbouring interconnected markets with delisting of capacity on short notice if committed producers or suppliers want to export in another market with tighter supply and higher energy prices. The US North East regional markets have experimented problems of this nature (Stoft, 2002). Even if a rule in the provision is supposed to protect the system operator by giving it the right to recall capacities designated for export in exchange for compensation, it remains possible that arbitrage with a neighbouring energy markets would be more lucrative after payment penalty for non compliance by the suppliers (or producers) has been paid (Creti and Fabra, 2004). It could be the case in a system with price cap below the price caps of the neighbouring ones. An answer to this difficulty should lie in creating precise and homogeneous rules for capacity obligation in neighbouring markets and coordination between TSOs in situations of system stress, two conditions that are

[1] For completeness, we also point out that this provision is not suitable in terms of incentives to making equipment available. In the case of a capacity payment, it has a direct effect of suppressing earnings owing to non-functioning when the mechanism's design incorporates this objective. In the case of a capacity obligation, effective capacity payments are disconnected from effective performance at the point in time in which the equipment is called into service (Cramton and Stoft, 2006). Capacity calculations only account for the mean record of availability by equipment type. The generator's incentive to limit outages in real time is nothing other than foregone earnings on the market for energy and reserves. Penalty does not act directly on the availability of equipment, but merely on holding adequate capacity certificates.

hard to satisfy. To summarise this analysis, the record of these three capacity mechanisms (cf. Table 1) reveals that they all present theoretical limitations and suffer from significant difficulties in application.

Table 1. Comparison of capacity mechanisms

	Long term Strategic reserve contracting	**Capacity payment (with flexible variant)**	**Capacity Obligation with exchangeable rights**	***Auctioning Forward capacity contracts /reliability options***
Countries	Sweden, Nor., France, New Z., Portugal	Spain, Italy Argentina, Chili, Colombia, Peru	US regional markets : PJM, NY, New England	*Proposals*
Efficiency of capacity targeting	+	-	0	+
Stability of investment in peaking unit	+	0	0	+
Consistency with the energy market	-	-/(+)	-	+
Robustness to strategic behaviour		0/(-)	+	*0*
-----------	------------------	------------------	------------------	------------------
Disincentive to market power on energy market	0	0	0	++
Feasibility : compa-tibility with non-mandatory market	yes	no	yes	*Yes, with significant adaptation*
Effectiveness against regional shortages	no	no	no	*yes*

Source : authors

These flaws lead us to seek out a joint steering over quantities and prices by establishing centralised coordination directed at the capacity goal desired for high peaks by resorting to calls for tenders (a procedure that has a dual function: allowing the desired security rate to be achieved and revelation of a reference capacity price through the bids) and ensuring a stable revenue to the generator during high-peak hours, so as to initiate investment in peak power plants in reference to the said price.

4. Centralised auctioning for forward capacity contracts / reliability option contracts

The proposed centralised mechanisms for auctioning capacity contracts or for reliability options seek to avoid the four main problems raised by the choice and design of a capacity mechanism: the requirement for a predictable economic signal (by stabilising the combined revenues of peak power plants from energy markets and from the capacity mechanism, and by organising an interaction between the markets of energy and “capacity rights”), the need for visibility of the revenues from peak units in the medium term, the incentives to efficiency by market mechanisms, and the provision of incentives to new entrants. These proposals rest on two pillars: on one hand, contracts that create a capacity commitment between generators and the TSO who, in this framework, represent consumers’ interests and, on the other hand, an intermediation of the TSO with the help of a centralized auction-based attribution of these contracts for the full capacity desired for extreme load period.

Thus, this type of provision conjugates a dual steering—by quantity (the capacity obligation that TSOs impose on suppliers and manage for them through calls for tenders) and by price (yielded by the auction). It rests on a centralised coordination that, in some senses, echoes the provision of calls for tenders for strategic reserves used in certain European markets. Nonetheless, it is clearly different for two reasons. First, this approach necessitates involving all generators' capacities in a contractual framework, and not only a stock of reserve equipment with a special status that is set aside from the other generation capacities of the market players. Second, this type of mechanism is in consistency with the energy market in the way the revenue given to capacity price (or the reliability option) is defined, deducts the infra-marginal rent and thus avoids distributive effects and distortions in the long term.

These principles are captured in the two proposals: the forward contracts mechanism proposed by Cramton and Stoft (2005; 2006), to be implemented on the New England market, and the centralised markets for reliability options contracts proposed (in particular) by Vasquez, Rivier and Perez-Arriaga (2001; 2003) and more recently by Oren (2005, 2006) and studied for the Spanish, Colombian, and Dutch markets. We present the two mechanisms as two variants of the same one.

4.1. The design of the mechanism

These two mechanisms are based on an explicit delegation to the TSO of the capacity obligation placed on suppliers, unlike the preceding provision for tradable capacity obligations currently in existence. One is based on forward capacity contracts between the TSO and generators, the contractual nature of which is distorted by the administrated subtraction of inframarginal rents obtained on the energy market during peak and calculated *ex post* by the regulator. The other draws on a call option between the same two parties, the intrinsically financial risk-management function of which is complemented by a physical right assigned to the TSO. This provides a strong incentive to the units that are party to these contracts to build the facilities necessary to satisfy all their commitment and to make available their power plants in times of tight supply. We can define the principal characteristics of these two mechanisms in six points.

- TSOs assess future capacity requirements during stochastic maximum peak, including the reserve margins they deem necessary. They proceed with auctions for long-term capacity contracts - or reliability options contracts -, between them and generators, for capacities in excess of total capacities owned by suppliers already bound vertical generation-supply integration or by forward contracts.

- The TSO's cost of capacity contracts (or reliability options) thus selected is subsequently distributed among suppliers as a function of their client-portfolio participation at the system's peak (or in the second mechanism, by the rule under which all kWh transmitted are allocated an uplift charge based on the transmission price).

- Each proposal is backed with a precise physical capacity. Selected generation capacity owners are compensated at the marginal bid price over a duration of four or five years, starting after two years. The capacities committed by the generators under these capacity contracts, after selection by auction, are subject to the standard obligation to bid on the balancing market or on the hourly energy market at the TSO's behest. Importantly they are liable to a penalty that is a function of the state of the system if they cannot.

- This penalty also provides a dual incentive: not only for the construction of plants to comply with commitments, but also to make all plants available during periods of high prices. In the "reliability options" mechanism in which the TSO exercises the option when the price of energy on the market exceeds the strike price of the option contract, generators having sold these options must bid on the energy spot market for the amount of energy contracted if they are able to provide it. If they cannot immediately produce this capacity, they pay the penalty.

- The price of the forward capacity contract (or the reliability option premium) is assumed to compensate for the shortfall in expected revenues to investors in peak facilities on the energy market by subtraction of the infra-marginal rent – or the revenues above the strike price in the reliability options system. It also adds compensation to base and semi-base load equipments for these shortfalls.

- However, the original aspect that needs to be underlined is that this provision ensures stability of revenues from the combination of the capacity contract and the hourly price of electricity to investors in peak units. In the design of the forward capacity mechanism, the amount of inframarginal rents on the energy market is, in fact, subtracted from the capacity price: *Ex post*, the TSO deducts the inframarginal rents that generators earn during peak and high-peak periods from the revenues generated by forward capacity markets. These rents are calculated by deduction of a "reference" cost-price of a marginal peak plant.

Revenue stabilisation is more explicit in the mechanism of reliability options, since by the very nature of the reliability option contract a reimbursement accrues to the generator, amounting to the difference between the strike price fixed by the regulator for all the reliability option contracts and the market price when the option is exercised. This mechanism thus refers to real inframarginal revenues, while in the first provision is based on an ex-post calculation with quite discretionary rules.

These two variants thus differ in the type of contract implemented between the system operator and the generators, in how the inframarginal rents are deducted from the capacity price (or the option price), and in the means by which the cost of the provision are distributed amongst consumers. However, beyond these differences, these two mechanisms are by nature a hybrid centralised mechanism combining control by price and control by quantity. Relative to traditional hybrid mechanisms, its specificity is that it centralises quantity-based management by mimicking the distribution of the responsibility for long-term security across suppliers and then distributing the cost of this security amongst them. Also, it reveals the equilibrium price of capacity through the bidding process.

We note that these design principles can serve to reform by far the capacity payment provision if it proves ineffective in eliciting investments in peak facilities. This is the solution advanced by Batlle et al. (2006) to reform the Spanish provision for capacity payments by focussing calls for tenders exclusively on the requirement for new capacity in peak facilities estimated by the regulator on the advice of the TSO. The capacity payments guaranteed for each new plant on a number of years is aligned on the marginal bid price.

4.2. Evaluation of the mechanism

This mechanism avoids some of the pitfalls presented by the other provisions (cf. Table 1).

- *The efficiency of adequacy targeting.* This mechanism combines effectiveness of the mechanisms based on a capacity obligation put on suppliers (here compliance with the obligation is delegated to the TSO, who backs the expected peak demand with generating capacities including a reserve margin) and those that centralise the development of capacity reserves. The TSO assumes the task of guaranteeing the level of desired capacity by auctioning the forward capacity contracts or long-term reliability options.[1] Thus, this provision can ensure good steering of total capacity by a quantity control. Generators who are bound by forward capacity contracts (or by reliability options) have an incentive to invest since, if they have not hedged all their contracts with physical capacities, they risk paying the penalty.[2]

- *Stabilising investment.* Forward capacity contracts or reliability option contracts are designed to stabilise investments in new peak power plants. These contracts enter into effect after two years, and then last four to five years in the case of new equipment. This contract design makes it possible for new entrants and small-scale generators to participate in the auction. Revenues from peak units - generated by both the capacity contract to which they are linked and the sale of electricity generated on the energy market - are also stabilised by this contract.

- *Compatibility with the energy market.* These two mechanisms explicitly links earnings from capacity to energy sales. In both variants, there is a trade-off between relinquishment of infra-marginal rents and the guarantee of a capacity price (or the option premium). The reference price used for computing these rents in the forward capacity mechanism (and the strike price in the reliability option mechanism) exercises a dual function in reducing risk and stabilising revenues (Cramton and Stoft, 2006). For generators, the function of stabilising revenues suppresses incentives to strategic behaviour because additional infra-marginal rents will be deducted from revenues yielded by capacity contracts, or by reliability option contracts. A generating company's revenues are limited to the strike price, for instance with the second mechanism. So there is no incentive to withhold power. For downstream buyers and consumers, it limits the level and volatility of energy prices and so the cost of their electricity supply.

[1] In order to avoid undue complexity, we will not delve into the bilateral long-term energy contracts involving generators and suppliers. Let us give two precisions about the forward capacity mechanism. Firstly the power capacities involved by these contracts are deducted from the amounts of capacity contracts put up for auction. Secondly to harmonize these two contract types for generators, those who are bound by bilateral wholesale contracts could also be required to offer their capacity in a capacity auction and to sign a dual option contract, call and put (also called a contract for differences, or CFD), with suppliers with whom they have previously contracted in this manner. (Joskow, 2006)

[2] By the way there is also supplementary incentive to be available during high load period in the reliability options mechanism. A generator who sold a reliability option, but is unavailable during price spikes during which the options is called is required to pay the differences between the market price and the strike price. It will bear a loss thaht increases with market prices.

- *Exposure to the risk of strategic behaviour.* This provision limits the possibilities of strategic behaviour on the energy market. On the markets of forward capacity contracts (or reliability options), risk is limited by the potential arrival of new entrants during the auction, owing to the long maturities of the contracts. However on a market with high degree of concentration, it is intrinsically exposed to risk of market power of the dominants generators.

- *Institutional feasibility.* This type of provision seems best suited to a mandatory market. In a non-mandatory exchange, the existence of bilateral sales contracts for energy complicates the definition of physical responsibilities vis-à-vis the capacity or reliability commitment. This market type requires significant adaptations for applying the mechanism, as was demonstrated in a study on the adaptation of the reliability options mechanism to the Dutch market (Vasquez, Perez-Arriaga et al., 2004). [1]

- *Effectiveness against regional shortage in open markets.* These two mechanisms are designed for mandatory markets with the system operator having greater control over market transactions than in non-mandatory markets, as points out previously about capacity obligation. But they present an advantage on the previous one to limit the effects of shortage in other interconnected systems, and they are more robust to risk of capacity and energy “leakages” to a neighbouring market with tighter supply during high load period. In the mechanism of reliability options, as the options contracted by the TSO hedge consumers for full demand of power capacity, generators know that they will have to reimburse all the revenues on the energy market over the strike price and will not respond to supplementary bids by the foreign buyers. So local suppliers can bid any price, if necessary, to obtain electricity for their consumers, despite tentative of foreign purchases on the pool or to local generators, because they are implicitly hedged against price spikes by the strike price used by the TSO. “This allows them to outbid competitors from neighbouring markets, who should not be willing to pay more than their reference VOLL” (De Vries, 2004). Because of this protection the suppliers should always be able to obtain all the electricity

5. Conclusion

Changes to the regulation of the electricity industry radically altered the approach to reliability of supply, which had been a major benefit of the centralised coordination by the regulated monopoly. With the de-verticalisation of activities and the decentralisation of investment decisions of competitive agents, the market is unable to ensure long-term coordination to guarantee an adequate level of peak capacity corresponding to the reliability requirements of consumers, as the regulated monopoly was able to do. A market solution that uses the spontaneous development of purely bilateral financial options contracts between producers and every consumer to attain an optimal level of reliability appears neither doable nor viable.

[1] Proposal by Vasquez et al. is that producers who hold bilateral contracts in non-mandatory markets sell reliability options for all their productions and to let them make option payments for the same volume when the options are called, but to return these payments through parallel contracts to the extent that the producers could demonstrate that they have sold their production through bilateral contracts. So when the options are called, they have to produce as much as they have committed. Then they receive the strike price for the electricity that they sell on the energy market and receive the contract price for the electricity sold under bilateral contracts If they cannot respect part of their commitments, they pay the market price minus the strike price for the capacity that was committed but is unavailable.

Henceforth, responsibility for long-term security must be clearly assigned to public authorities. But it must not be without seeking to partially correct the inelasticity of the instantaneous demand function that makes the establishment of a physical equilibrium impossible during extreme peak load under certain conditions of supply availability. It must also include a capacity mechanism, so as to ensure the existence of appropriate incentives for developing peak-unit capacities. However, the design of an efficient mechanism is rendered complex, as much by theoretical problems as by its practical implementation, owing to the complexity of electricity markets. Its insertion into a specific environment of market rules imposes compatibility constraints.

We could, as is the case in some European countries, make do with a simple procurement approach to contractually constitute strategic reserves if the market does not ensure an adequate development of peak units, but it forces to partly backtrack on the principles. It would, in any case, be necessary to accompany the provision with rules that would place boundaries on short- and long-term actions by the public authorities and the TSO to order strategic reserve plants to produce or to open call for tenders for reserve contracting, so as to avoid distorting the competition.

Moreover, the price-instrument and quantity-instrument, both of which decentralise the pursuit of the capacity goal, present theoretical limitations and serious hurdles to implementation. Steering by price provides absolutely no assurance with respect to the physical goal. The quantity-mechanism too easily leads to adopting a design of rights that is incompatible with the investment timeframe for peak power plants, and to separate energy and capacity goods.

This has led to proposing an mechanism that combines centralised quantity-based action with a payment for capacity that is set by a market mechanism. The rule of payment ensures revenues that are stable and adequate, combining the inframarginal rent yielded by the energy market with the capacity price net of this rent. Thus, we return to a centralised procurement of the capacity-good, but with it being strongly linked to the responsibility of producers to dispose of sufficient capacity and ensure its availability in high load period and also being thoroughly integrated into the energy market. We would have squared the circle, but for how complex these provisions are to define and implement. They require the definition of new obligations, new transactional infrastructures, and the establishment of mechanisms to render these obligations credible, which may lead market actors to fundamentally review their actions. The supply of long-term security in electricity supply is a most thorny problem.

References

Abbot M., 2001, “Is the security of supply a public good?” , *The Electricity Journal*, August-September, p. 31-33.

Batlle C., Rivier C., Vasquez C. & Perez –Arriaga I., 2006, *A regulatory mechanism to enhance security of supply in the Spanish wholesale electricity market*, Proceedings IAEE Conference, Potsdam, juin

Besser J. G., Farr J. & Tierrey S., 2002. “The political economy of long term generation adequacy: why an ICAP mechanism is needed as part of standard market design ? ” *The Electricity Journal,* August/Sept., vol 15, n°17.

Boiteux, M., 1949, "La tarification des demandes en pointe: Application de la théorie de la vente au coût marginal", 1949, *Revue générale de l'électricité.*

Borenstein, S. and Holland, S.P., 2003. Investment Efficiency in Competitive Electricity Markets With and Without Time-Varying Retail Prices. Center for the Study of Energy Markets working paper 106R.

Bouttes J.P. et Trochet J.M., 2002, *Marché de gros et bourses électriques en Europe et aux Etats-Unis*, Présentation aux Conférences Jules Dupuits, Paris, Décembre 2002

Bouttes J.P., 2005, Sécurité d'approvisionnements et investissements dans l'électricité, *Revue de l'Energie*, novembre.

Castro-Rodriguez F., Marin P.L., Siotis G. , 2001, *Capacity choices in liberalized electricity markets*, Discussion Paper 2998, Center for Economics Policy Resach (CEPR), London

Chao H., 1983. Peak Load Pricing and Capacity Planning with Demand and Supply Uncertainty. *The Bell Journal of Economics* 14 (1), 179-190.

Chao H.P. et Huntington H. ed, 1998, *Designing competitive electricity markets*, Boston (Mass.): Kluwer.

Choné F., 2004, *Proposition d'évolution d'équilibre du mécanisme d'ajustement*, Paris, Direct Energie, 31p.

Cramton P. et Stoft S., 2005, *A capacity market that makes sense*, Mimeo, March 2005.

Cramton P. et Stoft S., 2006, *The Convergence of Market Designs for Adequate Generating Capacity,with Special Attention to the CAISO's Resource Adequacy Problem*, White Paper for the Electricity Oversight Board, Mars

Creti A. et Fabra N., 2003, *Capacity markets for electricity*, Communication to the Colloquium "Efficiency on Electricity Markets", Toulouse, December 2003.

Demsetz H., 1968. Why regulate Utilities?. *Journal of Law and Economics* 11, 50-65.

De Vries, L., 2004, *Securing the public interest in electricity generation markets, The myths of the invisible hand and the copper plate,* Ph D thesis, Delft University, May 2004

Dubroeucq F. et Pignon V., 2005, *Retour d'expériences sur les mécanismes incitatifs en investissements en production*, EDF RD, ….

EC 2003, *Proposition de directive du Parlement européen et du Conseil concernant des mesures visant à garantir la sécurité de l'approvisionnement en électricité et les investissements dans les infrastructures* {SEC(2003) 1368}/* COM/2003/0740 final - COD 2003/0301 */

E.C., 2005, *Directive 2005/89./EC of the European Parliament and the Council concerning measures to safeguard security of electricity supply and infrastructure investment,* Brussels, 18 January 2006.

Finon D., 2006, "Incentives to invest in liberalised electricity industries in the North and South. Differences in the need for suitable institutional arrangements", *Energy Policy,* 2006, Vol.34, n°5, p.601-618

Hobbs, B., Iñón, J. & Stoft, S.E. 2001b. Installed Capacity Requirements and Price Caps: Oil on the Water, or Fuel on the Fire?, *The Electricity Journal* 14 (6),p. 23-34

Hugues W.R. et Parece A., 2002, " The economics of price spikes in deregulated power markets", *The Electricity Journal,* July, p.31-44

Hunt S., 2002, *Making competition works in electricity*, New York, J. Wiley Publishers

IEA , 2002, Security of supply in electricity markets: evidence and policy issues, Paris, OECD/IEA.

IEA, 2003, *Power generation investment in electricity markets*, Paris, OECD/IEA.

Joskow P. et Tirole J., 2004, *Reliability and Competitive Markets*, Working paper CEEPR-MIT

Joskow P., 2006, *Competitive electricity markets and investment in new generating capacity*, CEEPR-MIT, Working Paper 06-009 WP, April

Lévêque F., 1998, *Economie de la règlementation*, Paris, La Découverte.

Meunier G. et Finon D., 2006, *Market Power in Electricity Markets, Suboptimal Investment and ISO Procurement*, Presentation à la 26° USAEE/IAEE Conference, September 2006,

Neuhoff K. et De Vries L., 2004, *Investment Incentives for investment in electricity generation,* Cambridge Working Paper in Economics, CMI WP42.

Newbery D., 2006, Electricity liberalisation in Britain, the quest for a satisfactory design, *The Energy Journal*, Special Issue “European electricity Liberalisation”, p. 43-70

Oren S., 1999, *Ensuring generation adequacy in a competitive electricity market*, Workshop “Long-term guarantee to supply”, Madrid, Comission Nacional de la Energia, 7 June 1999.

Oren S., 2003, *Ensuring generation adequacy in competitive electricity markets*, Univ. of California Energy institute, Working Paper UCEI, EPE 007

Oren S., 2005, *Generation adequacy via call options: safe passage to the promised land*, Univ. of California Energy institute, Working Paper UCEI, EPE-016

Oren, S., 2006, *Generation adequacy through call options obligations*, Power Presentation, EDF-RD, Clamart, 17 Juillet 2006

Perez-Arriaga I., 2001, *Long term reliability of generation: a criotical review of issues and alternative options*. Woking Paper IIT-00-0981T, Instituto de Investigation Technologica, Univ. Commillas Madrid

Pignon V. et Tarbé M.,2003, *Problématique économique des investissements en production*, Working Päper EDF R&D, HR-19/03/054/A

Stigler G.J. , 1971, The theory of economic regulation, *Bell Journal of Economics and Management Science*, n°2 (1), p.3-21

Stoft S., 2002, *Power System Economics, Designing Market for Electricity*. New York, Wiley IEEE

Turvey R., 2003, Ensuring adequate generation capacity*, Utilities Policy*, Vol. 11, p. 95-102

Vasquez C., Perez-Arriaga I., Rivier,C., 2001, A market approach to long-term security of supply, *IEEE transactions on Power Systems*, Vol. 25, n°2, p. 349-357

Vasquez C., Battle C., Rivier M., Perez-Arriaga, I., 2003, *Security of supply in the Dutch electricity market : the role of reliability options*. Report IIT-03-0841C, (prepared for the Office for Energy Regulation of the Netherlands -DTE), Comillas university, IIT.

Weitzman, M L, 1974, Prices vs. Quantities, *The Review of Economic Studies*, 41 (4), pp. 477-491.

Wolak F., 2004, *What’s wrong with capacity markets?* Mimeo, UCEI.