CERRE

CENTRE ON REGULATION IN EUROPE

Generation capacity adequacy in Europe: What economic rationale for Capacity Remuneration Mechanisms?

A CERRE study

Guido Cervigni

13 November 2013

131113_CERRE_CapRemMech_final

Centre on Regulation in Europe (CERRE) asbl
Rue de l'Industrie 42 Box 16 - B-1040 Brussels - Belgium
Ph: +32 (0)2 230 83 60 - Fax: +32 (0)2 23 0 83 70 – info@cerre.eu – www.cerre.eu

CERRE CENTRE ON REGULATION IN EUROPE

Table of contents

About CERRE ... 3

About the author ... 4

Key messages of the report ... 5

Introduction and executive summary ... 7

1. The rationale for capacity remuneration mechanisms in electricity generation ... 14

1.1. The market mechanism driving investment decisions in power generation capacity ... 14

1.2. What makes power generation special? ... 14

1.2.1 Price patterns in wholesale electricity market ... 14

1.2.2 Vulnerability to market power ... 18

1.2.3 Interactions between generation investment and network development planning ... 19

1.3. A rationale for capacity remuneration mechanisms ... 20

1.3.1 The missing money problem ... 20

1.3.2 Coordination ... 28

1.4. Discussion ... 30

2. Capacity remuneration mechanisms ... 34

2. 1. How capacity remuneration mechanisms work ... 34

2.1.1 A stylized capacity payment scheme ... 34

2.1.2 A stylized single-buyer scheme ... 35

2.1.3 Conditions for the effectiveness of capacity remuneration mechanisms ... 36

2. 2. Capacity payments ... 37

2. 3. Capacity requirements ... 41

2.3.1 The standard model ... 41

2.3.2 Reliability options ... 45

2. 4. Strategic reserve ... 47

3. Additional issues ... 52

3. 1. Coordinating capacity remuneration mechanisms across multiple countries ... 52

3.1.1 Enforcement of capacity rights in an interconnected system ... 53

3.1.2 Capacity remuneration mechanisms and cross-border competition ... 60

3.1.3 Discussion ... 62

3. 2. Capacity remuneration mechanisms and flexibility of the generation fleet ... 62

3. 3. Capacity remuneration mechanisms and demand side response ... 63

Appendix: The EC Consultation paper on generation adequacy, capacity mechanisms and the internal market in electricity, 15 November 2012 ... 67

CERRE CENTRE ON REGULATION IN EUROPE

About CERRE

Providing top quality studies, training and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network industries. CERRE's members are regulatory authorities and operators in those industries as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross sector approach;
- the widely acknowledged academic credentials and policy experience of its team and associated staff members;
- its scientific independence and impartiality.
- the relevance of its contributions to the policy and regulatory development process applicable to network industries and the markets for their services.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological and social environment. CERRE's work also aims at clarifying the respective roles of market operators, governments and regulatory authorities, as well as at strengthening the expertise of the latter, since in many member states, regulators are part of a relatively recent profession.

This study has received the financial support of a number of CERRE members. It has been completed under the supervision of Professor Catherine Waddams, Joint Academic Director of CERRE and Professor at the Centre for Competition Policy – University of East Anglia. As provided for in the association's by-laws, it has, however, been prepared in complete academic independence. The contents and opinions expressed reflect only the author's views and in no way bind the sponsors or any other members of CERRE (www.cerre.eu).

CERRE CENTRE ON REGULATION IN EUROPE

About the author

Dr Guido Cervigni is a CERRE Research Fellow and a Research Director at IEFE-Bocconi.

He is an economist and former regulator who provides qualitative and quantitative analyses on regulatory and competition-related issues, with a particular focus on energy markets. As a member of the System Operation Expert Group advising the European Regulators' Group for Electricity and Gas (ERGEG), he has been involved in the development of the EU framework guidelines on System Operations.

Guido is a former Director at LECG Consulting, Head of Business Development in an energy trading company and Head of Market Development at the Italian Power Exchange. Guido started his career at the Italian Energy Regulatory Authority where he was Head of the Competition and Markets Division.

Guido Cervigni holds a PhD in economics from Bocconi University, Milan.

Key messages of the report

- Capacity remuneration mechanisms are policy measures aiming to ensure that adequate level of capacity be available at all time, by providing reward to electricity generators in addition to the income obtained by selling electricity and ancillary services in the market.
- Some technical and economic features, specific to electricity markets, may provide a rationale for capacity remuneration mechanisms. These include, first, non-storability of electricity and price inflexibility of a large part of electricity demand, which cause rare but extreme prices spikes. Second, some features of the market design and regulatory system may prevent energy and operating reserve prices from rising to levels that correctly reflect conditions of scarcity. Third, capacity adequacy concerns are sometimes motivated by the specific risk structure of the generation business, such that small changes in demand or supply conditions can have a dramatic impact on generators' profitability at times of scarcity.
- Any empirical assessment of the nature of the market failures justifying implementation of capacity remuneration schemes, particularly in Europe, is likely going to be very difficult and highly contentious.
- Some administrative intervention impacting on power generators' income, and therefore on the level of installed capacity such as setting the price for electricity in case of physical rationing (based on the value of lost load) and the operating reserve requirements, is unavoidable. However, capacity remuneration schemes take public intervention in the industry a step further, and discussion on such measures ultimately boils down to the (largely political) question of whether governments or private investors should decide on the level, type and location of generation capacity.
- Three broad types of capacity remuneration measures can be identified: capacity payments, capacity requirements and strategic reserve.
- Irrespective of the mechanism selected, credibility of the public authorities' commitment is crucial to the effectiveness of capacity remuneration mechanisms.
- Capacity remuneration mechanisms should not be regarded as substitutes for a sound market design and a predictable regulatory environment. Eliminating any market design flaws or unnecessary regulatory uncertainty remains crucial to achieving efficient

electricity markets and attracting capital in the industry, independently of capacity remuneration measures.

- Capacity remuneration measures may impact on energy and ancillary service prices, and therefore possibly on cross-border flows. However, some factors may mitigate the distortion in the generators' siting decisions caused by unilateral or asymmetric capacity remuneration mechanisms. In particular, limited import capacity into a country reduces the scope for profitable exports of energy and ancillary service by generators sited in neighbouring countries. All other things equal, then, limited interconnection reduces incentives to divert investment from countries where capacity remuneration schemes are implemented to countries where such schemes are not implemented. Further, it is reasonable to assume that a remuneration scheme independently introduced by a country pursues a capacity target based only on the country's energy and reserve requirement, i.e. it does not reflect the objective of subsidizing capacity that would mainly export production. Under that assumption, if the mechanism turned out to divert investment from neighbouring countries, the capacity target of the country implementing the remuneration scheme would be exceeded. In that case, capacity remuneration will reduce, either as an automatic consequence of the scheme's design or as a result of policy decision. This perspective should deter (excess) investment in the country running the capacity remuneration scheme.

- Flexible loads, capable and willing to give up consumption in case of high prices, may reduce the system capacity requirement. Intraday and balancing markets are meant to reward demand side, as well as supply side flexibility and therefore to provide efficient incentives to invest in flexible resources. Regulatory or market imperfections may nevertheless justify compensating consumers for making their load flexible. However, it is hard to find obvious reasons why the market for energy services could not be relied upon to achieve the efficient outcome.

Introduction and executive summary

In several European countries, governments and regulators appear concerned that in the future the level of electricity generation capacity delivered by the market may not ensure that there is sufficient supply to meet demand at all times.

In some member states, these concerns have been spurred by the reduction of generators' profitability due to the demand slow-down caused by the global economic crisis. Further, the ambitious sustainability targets of the European Union are expected to dramatically increase the stock of subsidized renewable generation capacity. This will affect the profitability of non-subsidized generators, both by creating additional capacity and by increasing the volatility of demand, net of renewable production. Finally, capacity adequacy concerns in Germany are associated with the accelerated phase-out of the nuclear installations, following the Fukushima accident.

Fears that the lights might go out, which raise strong political sensitivities, have generated a lively debate on the appropriateness of policy measures aiming to reward generators in addition to the income obtained from the sale of electricity and ancillary services in the market, or "capacity remuneration mechanisms", and on the design of such measures. Capacity remuneration schemes currently operate in several US markets, in some Latin American countries and - in Europe - in Portugal, Spain, Italy, Greece, Ireland, Sweden, Finland and Poland. France is planning to introduce a capacity remuneration system by 2016. Measures to ensure capacity adequacy are currently being discussed in Great Britain and Germany.

In its recent Guidance for state intervention in electricity[1], the European Commission acknowledges complexity of capacity adequacy assessments and recognises the possibility that public intervention might be necessary to ensure security of electricity supply. However, the Commission stresses that market forces should be primarily relied upon to ensure appropriate investment.

In this report, which was prepared before the release of the above guidance document by the Commission, we investigate the theoretical foundation for generation capacity remuneration schemes and assess alternative policy options. The main results of our analysis are summarised in the rest of this section.

[1] http://ec.europa.eu/energy/gas_electricity/doc/com_2013_public_intervention_en.pdf

The rationale for capacity remuneration measures

Some features specific to electricity markets are often used to provide a rationale for the introduction of capacity remuneration measures. First, a large part of electricity demand is currently price-inflexible in the short run. When the price-insensitive portion of demand exceeds available generation capacity, i.e. during "scarcity" events, involuntary load reduction via disconnections, or load shedding, may become necessary. In this case the price of electricity must be administratively set.

Scarcity hours are particularly important in the electricity industry because a potentially large portion of some generators' fixed costs must be recovered during these hours. Even moderate distortions of the electricity prices prevailing during scarcity hours, or in the number of scarcity hours, could have a major impact on generators' profitability.

The large scale deployment of smart-metering systems, expected to take place in Europe by 2020[2], will enable most consumers to plan and adjust power consumption in response to short-term prices. It remains to be seen whether the net benefits of such behaviour is positive for a large enough share of the consumers that a significant portion of the market's demand will become flexible, removing one of the main sources of capacity adequacy concerns.

Second, some features of the market design and regulatory system may prevent energy and operating reserve prices from rising to levels that correctly reflect conditions of scarcity.

Third, capacity adequacy concerns are sometimes motivated by the specific risk structure of the generation business, such that small changes in demand or supply conditions can have a dramatic impact on generators' profitability at times of scarcity. That could justify policy measures improving coordination of investment decisions on generation capacity, a function performed by some capacity remuneration schemes.

Empirical assessment of the nature of the failures, if any, featuring in a wholesale power market is key to identifying any appropriate policy intervention. While missing money issues would call for mechanisms that integrate generators' income, in order to attract an efficient level of investment, coordination issues can be handled by mechanisms governing the timing and possibly location of investments in generation capacity, in order to reduce the risk for investors.

[2] According to Directive 2009/72/EC: "*where roll-out of smart meters is assessed positively, at least 80 % of consumers shall be equipped with intelligent metering systems by 2020*"

The opportunity to implement capacity remuneration measures is easily assessed only in extreme situations, like for example in Latin American countries where very severe droughts take place infrequently and some thermal generation capacity is expected to be unnecessary, sometimes for several consecutive years.

In less extreme situations, and in particular in the current European context, assessing whether and to what extent the special features of electricity justify capacity remuneration is hard. It is not surprising that no empirical analysis explicitly aimed at assessing the impact of those features has been carried out so far. Providing solid evidence in favour or against generation capacity support is difficult for several reasons. First, evidence of, respectively, scarce or excess capacity at a certain point in time does not by itself prove that capacity support is or is not necessary. Second, assessing the opportunity for a capacity remuneration scheme requires figuring out the supply conditions that would be produced by the measure, a highly uncertain guess. Finally, in most European countries central elements of the electricity wholesale and ancillary service market design are still being developed. As a consequence the need for capacity support must be assessed with reference to a somewhat untested market design.

Some administrative intervention impacting on the power generator's income, and therefore on the level of installed capacity provided by the market such as setting the price for electricity in case of physical rationing (based on the value of lost load) and the operating reserve requirements, is unavoidable.

However, capacity remuneration schemes take public intervention in the industry a step further. Capacity remuneration mechanisms involve the public sector in deciding the size and possibly location and composition of the power generation fleet. To the extent that the public authority's choices do not coincide with those that market investors would autonomously make, the measure's scope extends beyond coordination to some form of planning. At that point, some of the investment risk becomes shared with consumers and/or taxpayers, through the capacity remuneration scheme. This share increases with the gap between the government's and the market's view on the desirable generation fleet. From this perspective the discussion on capacity remuneration measures boils down to the (largely political) question of whether governments or private investors should decide on the level, type and location of generation capacity.

When assessed in a broader policy perspective, insights into the opportunity for generation capacity support may be gained by evaluating the relative likelihood and cost of two errors:

- implementing a capacity remuneration mechanism when that is unnecessary. i.e. when market dynamics would result in the efficient level of capacity being built;
- not implementing capacity support when needed.

As to the cost of unnecessary capacity support, well-designed mechanisms - with capacity targets set well in advance of delivery and competitive selection of the suppliers - may result in very little additional cost for consumers in the case when the target level of capacity would have been reached without support.

However, less than perfect designs have historically been implemented in some markets. In Europe, in particular, some capacity payment schemes are more likely to implement a surplus transfer between the consumers and the existing generators and a risk transfer from generators to consumers than to impact on the investment decisions in additional capacity. So the cost of an unnecessary intervention through capacity mechanisms may be through market distortions, which are likely to impose penalties on consumers.

As for the cost of not implementing capacity support when needed, on the one hand the (social) cost in case demand cannot be met is very large, given the high value of electricity for consumers. Further, since selective disconnection of small consumers in case of scarcity is currently unfeasible, the VoLL is likely to provide a very imperfect representation of the value of electricity for many consumers.

On the other hand, large excess capacity conditions currently hold and are expected to persist in the future in most European countries[3], because of the demand reduction caused by the economic crisis and the growth of renewable capacity. This might prevent the cost of a missing capacity remuneration scheme from emerging in the foreseeable future.

In any event, capacity remuneration mechanisms should not be regarded as a substitute for a sound market design and a predictable regulatory environment. Eliminating any market design flaws or unnecessary regulatory uncertainty remains crucial to achieving efficient electricity markets and attracting capital in the industry, independently of capacity remuneration measures.

---

[3] ENTSO-E Scenario Outlook & Adequacy Forecast 2013-2030.

Alternative design for capacity remuneration mechanisms

Alternative approaches to the design of capacity remuneration schemes can be identified. The first approach, "capacity payments", sets a price that a central entity, on behalf of the consumers, commits to pay for available capacity. Capacity payments add to the revenues obtained by generators from selling electricity and ancillary services; the higher expected income is meant to attract additional investment in generation capacity. Capacity payments can solve missing money issues, but they do not address coordination issues. The second approach, "capacity requirements", sets the volume of available capacity that the central entity commits to paying for, either directly or by placing an obligation on the load-serving entity proportional to their client's (inflexible) contribution to the system's peak. This creates the demand for a product, available capacity, which generators supply. The interaction between the regulatory-driven demand and the supply of available capacity determines the market clearing price for available capacity. With capacity requirements, the public authority, through the system operator, takes direct responsibility for setting the level and possibly location and mix of generation capacity. It addresses both missing money and coordination issues.

The last approach, "strategic reserve", consists of reserving a certain generation capacity or flexible load for use only in scarcity situations, as a substitute for inflexible load curtailment. This approach, if properly implemented, has the least impact on the functioning of the wholesale electricity markets. However, it is especially suitable as a transitory measure, preventing closure of inefficient plants and using their capacity as a substitute to black-outs.

Remuneration measures that actors in the market do not believe will be maintained over the relevant time horizon will not modify investment decisions and will result in mere wealth transfers from consumers to generators. Therefore credibility of the public authorities'' commitment is crucial to the effectiveness of capacity remuneration mechanisms.

Capacity remuneration measures and market integration

In the context of liberalized and integrated wholesale power markets, different countries may pursue different reliability (or capacity) targets.

In an interconnected system the allocation among the national system operators of the rights over generation capacity may be modified by cross-border deals. The resulting allocation determines power flows during scarcity events, or, in other words, the allocation of disconnections among interconnected countries. For example, if country A purchased 1000 MW capacity from country B, in the event of scarcity in A a power flow of at least 1000 MW from B to A must be implemented, even if that results in scarcity in country B.

In order to enforce cross-border rights over generation capacity, therefore coordination among system operators is necessary.

Capacity remuneration schemes can be designed in a way such that they do not distort the generators' profit-maximizing competitive offer strategy in the energy and ancillary service markets and therefore do not cause inefficient utilization of the existing generation fleet. However, capacity remuneration measures may impact on energy and ancillary service prices, and therefore possibly on cross-border flows, by modifying the level and composition of installed capacity in the country where support is implemented. This results from the income-increasing and risk-reducing features of the capacity remuneration scheme making generation investment in the country where it is implemented more attractive.

Capacity remuneration mechanisms and composition of the generation fleet

The increasing share of renewable generation capacity, in particular solar and wind, is creating concerns that the market may not provide enough flexible generation resources to offset intermittent renewable sources. In a frictionless setting, the usual market mechanism addresses not only the level of installed capacity but also its composition. In particular, the increasing need for flexibility results in greater trading near real-time and larger demand for balancing services by the system operator, to the benefit of more flexible generators.

Issues that may justify special support for flexible capacity are, *mutatis mutandis*, the same as those identified with respect to the overall level of capacity. In the event that support to flexible generation capacity is deemed necessary, capacity payments and capacity requirements can be adapted to impact only on the more flexible subset of the generation fleet.

Integration of flexible demand in capacity remuneration measures

Flexible loads, capable and willing to give up consumption in case of high prices, may reduce the system capacity requirement.

A straightforward way to allow consumers to appropriate the value of their capability to control load is by exempting flexible load from any capacity procurement obligation or, where a centralized capacity remuneration system is implemented, from paying capacity-related costs. In exchange, flexible consumers would be required to make their consumption decision price-dependent, by bidding in the energy and/or ancillary service markets in a way such that load will (voluntarily) not be dispatched for some energy price.

Regulatory or market imperfections, such as those previously discussed may justify compensating consumers for making their load flexible. In particular remunerating demand-response may be justified by the inability of consumers to assess the future value of capacity. It remains to be seen whether (and why) the market could not be relied upon to achieve the same outcome. For example energy service companies or even the consumer’s suppliers might find it profitable to finance the consumer’s investment in flexibility in exchange for a share of its value.

This report is organized in three chapters. In Chapter 1, we investigate the rationale for generation capacity remuneration schemes. In Chapter 2, we analyse alternative capacity remuneration mechanisms. In Chapter 3, we address additional issues including cross-border effects of capacity remuneration mechanisms, the use of remuneration measures to govern the development of flexible generation capacity and the integration of demand response in capacity remuneration mechanisms. In the Appendix, we present some of our results in terms of answers or comments to the European Commission consultation paper of 15 November 2012 on generation adequacy, capacity mechanisms and the internal market in electricity.

1. The rationale for capacity remuneration mechanisms in electricity generation

1.1. The market mechanism driving investment decisions in power generation capacity

The economic mechanism driving investment in production capacity in a liberalized electricity market is conceptually the same as in most other industries. Generators obtain revenues from selling electricity and ancillary services. Investment in new generation capacity is driven by the investors' expectation of the future revenues from selling electricity and ancillary services in the market. In this context, persistently high electricity and ancillary service prices – which obtain when the existing capacity is frequently fully utilized – attract investment in generation capacity. Conversely, low electricity and ancillary service prices when part of the installed capacity is little utilized discourage capital accumulation.

Note that the market mechanism differs substantially from the traditional planning approach, in which utilities would meet reliability and resource adequacy requirements, exogenously set based on engineering standards regarding the acceptable hours of load shedding. In a market approach demand and supply rather than engineering standards determine the efficient level of installed capacity, and ultimately the level of reliability.

1.2. What makes power generation special?

In this section we discuss how electricity's technical features may affect investment in production capacity. First, we illustrate the (efficient) price pattern in wholesale power prices (Par. 1.2.1). Second, we discuss vulnerability of wholesale power markets to market power (Par. 1.2.2). Third, we investigate the potential interactions between market-based investment in generation capacity and planned transmission network development (Par. 1.2.3).

1.2.1 Price patterns in wholesale electricity market

Electricity cannot be economically stored on a large scale. It must constantly be produced in the same quantity as it is consumed. In addition a large portion of the demand for electricity is currently price-inflexible, at least in short timeframes.

As long as demand is lower than available generation capacity, the competitive clearing price of the wholesale electricity market equals the variable cost of the most expensive generator that needs to be activated to meet load, or the *system marginal cost*. On the contrary, in case of scarcity – when the demand for electricity and operating reserve is greater than available generation capacity – the wholesale market clearing price should be set at the level that rations demand. As this is higher than marginal cost, there is a contribution to service the cost of capital provision.

Since most of the electricity meters currently in place do not record hourly consumption, but only total withdrawal over a longer time period, charging a different price for consumption in different hours is impossible. Therefore in case of scarcity the price for electricity must be administratively set. The theoretical reference for the price-level in scarcity situation is value of lost load, or VoLL, the price that makes consumers indifferent between consuming electricity at that price, and not consuming. VoLL is typically estimated at several orders of magnitude greater than average electricity prices.

Administratively setting the price in case of scarcity is necessary to address what (Crampton and Stoft 2006)[4] refer to as the "reliability problem". Since the current technology makes it impossible to disconnect consumers individually, on the basis of each one's willingness to pay, capacity becomes a public good. The market would under-invest in generation capacity if consumers were left free to decide how much "reliability" to purchase, because each consumer, knowing that he cannot fully benefit from his own purchase of reliability, would have an incentive to buy less than efficient reliability or the equivalent capacity availability to provide it.

Voll-setting is the simplest, and currently unavoidable, form of capacity remuneration system, as the level of the VoLL determines the rent obtained by generators during scarcity events. Greater VoLL incentivises, ceteris paribus, larger investment in generation capacity. We will discuss issues related to VoLL assessment later in the report.

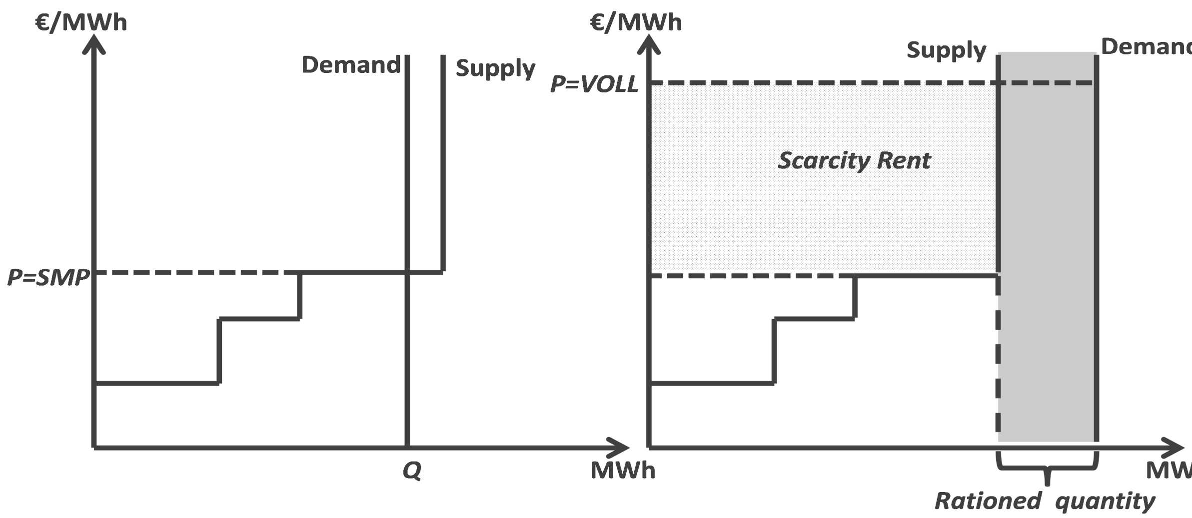
VoLL is typically estimated several orders of magnitude greater than average electricity prices, with VoLL in the range of €5,000-€10,000 per MWh commonly regarded as plausible. As noted, since the value of electricity to consumers is generally much higher than the typical production cost the efficient market clearing price in scarcity conditions is order of magnitudes greater than

[4] P. Cramton and S. Stoft, 2006. *The convergence of market designs for adequate generating capacity with special attention to the CAISO's resource adequacy problem* MIT Center for Energy and Environmental Policy Research

in normal conditions. The market outcome under normal and scarcity conditions is illustrated in Figure 1.

We depict a market with fully inflexible demand; in this case a price increase does not reduce consumption and quantity rationing must be implemented in the event of scarcity[5], and the system operator plans curtailment of service on different portions of the grid. Since selective disconnection is technically unfeasible, all the consumers connected to the same network branch will disconnected at the same time.

Figure 1 - Wholesale electricity market clearing price in regular and scarcity conditions



Scarcity conditions occur when generation capacity available for dispatch is not enough to meet the demand for energy and for operating reserve[6]. For simplicity we do not address operating reserve consistently in the paper. All results presented with reference to energy carry over to operating reserve.

We discuss in Box 1 on the next page how price spikes in case of scarcity obtain under alternative wholesale market designs.

Because of these features of electricity supply occasional capacity shortages and price spikes are normal in well-functioning wholesale power markets. As a consequence a large portion of some generators' fixed costs is covered by the margins obtained during rare conditions of scarcity (the

[5] Scarcity situations may not result in load curtailment. A scarcity situation might just make it impossible to meet one or more security constraints, which would lead to service disruptions only in case certain contingencies occur. For simplicity of exposition here we discuss scarcity situations in terms of load curtailment.

[6] The objective of operating reserve procurement is to ensure that sufficient capacity will be available in real time in order to perform balancing and to prevent major network disruptions.

"scarcity rent" in the Figure)[7]. This makes the investment, especially in peak capacity, risky, since even small changes in the number of scarcity events can have a dramatic impact on the producers' revenues. Further, those issues, combined with the long lead times of construction of generation capacity, may result in boom-bust investment pattern, as investors might wait before investing until the frequency of scarcity events provides unambiguous evidence that the additional capacity will be profitable. In the low phases of those cycles installed generation capacity may not be enough to match load, under security conditions, at all times, while in high phases of the cycles generation capacity will be greater than load all the time and profitability for generators will be abnormally low.

Box 1: Alternative market designs for scarcity pricing

Alternative spot market designs differ in the way the clearing price is set when the system is tight. We illustrate the logic underlying the two main mechanisms in the context of a simplified version of the design implemented in most day-ahead electricity markets in Europe. We assume that one spot energy market session is run for each delivery hour and that a non-discriminatory auction clears the market. In non-discriminatory auctions, also known as single-price auctions, every accepted bid and offer respectively pays and receives the market clearing price - the price of the last accepted bid or the first rejected offer.

The first methodology to achieve high prices in scarcity hours relies on market participants offering prices higher than their variable costs when conditions of scarcity are expected. With a non-discriminatory auction each generator's competitive bidding strategy entails offering to sell at any price higher than its variable cost, since the clearing price set by the market operator is independent of all the generators' offers but the highest-priced accepted one.

However, when the system is known to be stretched, each generator calculates a certain probability that demand will be greater than the total available capacity. The generators' expected profit maximizing strategy in that situation entails offering part of its capacity at prices greater than the variable cost. By doing so, the generator takes into account the possibility that its offer will set the market clearing price. In this situation the generator bears the risk that conditions of scarcity will actually not arise, and that its offer will be displaced by the competitors' offers. Productive inefficiencies may also arise if some offers above cost turn out to

[7] The peaking generating units, with the highest variable costs, rely on prices prevailing during the very few scarcity hours to cover their entire fixed cost.

be displaced by cheaper offers submitted by less efficient competitors. Finally, in this model the market clearing price may turn out to be greater than the system variable cost, even if conditions of scarcity do not actually come about.

In the alternative approach a scarcity pricing trigger is included in the market clearing algorithm: the market clearing algorithm automatically sets the market clearing price at VoLL when conditions of scarcity are detected based on the offers and bids submitted by market participants[8].

In this situation, the generators' competitive bidding strategy is to offer its variable cost, irrespective of the expected demand and supply conditions, because in the event of scarcity the clearing algorithm itself will set the price at the VoLL[9]. This approach is consistent with the broader objective of reducing the scope for inefficiencies caused by prediction errors. In general this requires that a market participant's profit-maximizing bidding strategy does not depend on its expectation of the clearing price.

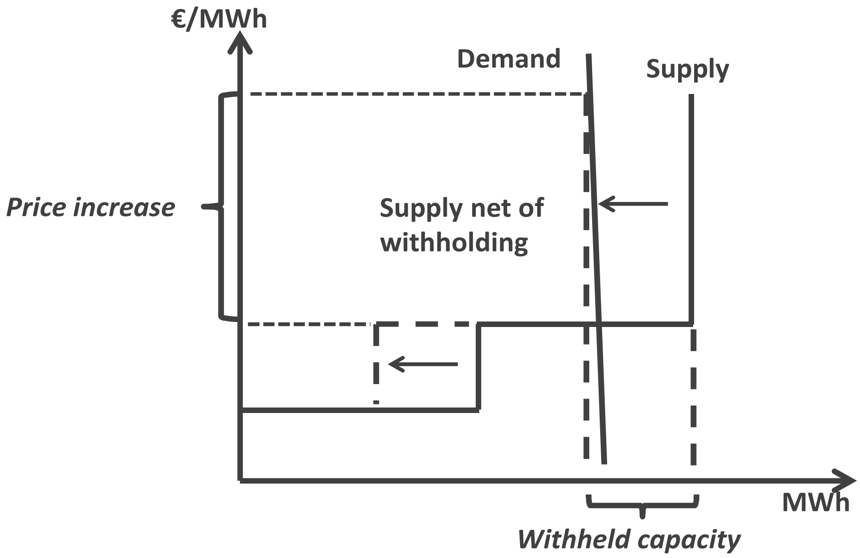
1.2.2 Vulnerability to market power

In wholesale electricity markets market power is a primary concern for two broad reasons. The first is that electricity is a primary commodity purchased by every household and business, and its price is extremely important for the economy. The second is that the unique technical and economic characteristics of electricity make wholesale electricity markets particularly vulnerable to the exercise of market power. These characteristics are little or no demand responses to price changes, the fact that electricity is not storable and tight transmission capacity constraints that reduce the scope for competition among generators connected in different locations.

As a result, even small generators may have the interest and ability to induce dramatic price increases when existing generation capacity comes close to full utilization. In that event the withholding of even a small quantity of supply from the market may cause a sharp increase in price. Figure 2 below illustrates a situation in which the withholding of generation capacity leads to scarcity, i.e. increases the market clearing price to the level necessary to ration demand.

[8] This approach is less effective if spot markets are not centrally cleared. In that case VoLL pricing can be administratively enforced only in the balancing market, where the system operator is counterparty to all transactions.
[9] In German-Austrian and Italian markets VoLL pricing is implemented by setting the bid-price by inflexible consumers equal to VoLL.

Figure 2 - Generation capacity withholding when the system is tight



The analysis of the opportunity to introduce market power mitigation measures and the merits of alternative mechanisms go beyond the purpose of this paper. In section 2 we will analyse the relationship between market power mitigation and capacity remuneration schemes.

1.2.3 Interactions between generation investment and network development planning

Public decision-making on transmission network development interacts with and creates risk to market-driven investments in power generation capacity.

Transmission and generation capacity are linked by complex relations of complementarity and substitutability. A network upgrade may be necessary in order to transfer a new generator's production from the injection node to the load centres. Alternatively, a transmission upgrade might make it possible to meet a demand increase at a certain location with production from existing generators at other locations, thus avoiding the construction of additional production capacity. Thus, total supply cost minimization requires transmission and generation investment decisions to be coordinated.

In most electricity markets, the system operator is the monopoly supplier of transmission services. The system operator is responsible for planning and often delivering network's upgrades. Therefore, the system operator's network development decisions may have a major impact on the generators' profitability.

1.3. A rationale for capacity remuneration mechanisms

In this section we assess how the specific features of electricity affect the economic mechanism driving investment in generation capacity, and investigate why this may justify implementation of capacity remuneration schemes.

Motivations for capacity support are found in two areas. Firstly, some features of the market design and regulatory system may prevent energy and operating reserve prices from rising to levels that correctly reflect conditions of scarcity. In this case the generation capacity is under-remunerated in scarcity situations, which results in under-investment.

Secondly, capacity adequacy concerns are sometimes motivated by the specific risk structure of the generation business, such that small changes in demand or supply conditions can have a dramatic impact on generators' profitability at times of scarcity.

While the first issue calls for mechanisms that integrate the generators' income in order to attract an efficient level of investment, the second issue can be handled by measures that coordinate the timing of investments in generation capacity in order to reduce the risk for investors.

We discuss the 'missing money' problem in Par. 1.3.1 and the coordination role of capacity remuneration schemes in Par. 1.3.2.

1.3.1 The missing money problem

There is a missing money problem in case the efficient level of generation capacity cannot remunerate the invested capital at the rates required by the market to invest, by selling energy and ancillary services. In other words, the missing money problem occurs when some elements of the market design, industry regulation or industry practices cause generators' revenues to be systematically insufficient to attract the efficient level of investment.

When the missing money problem becomes a structural feature of the market, the result is a drop in installed capacity. Capacity support mechanisms are therefore intended to integrate generators' income in order to counter the effects of the imperfection causing the missing money problem. Below we discuss the potential causes of missing money.

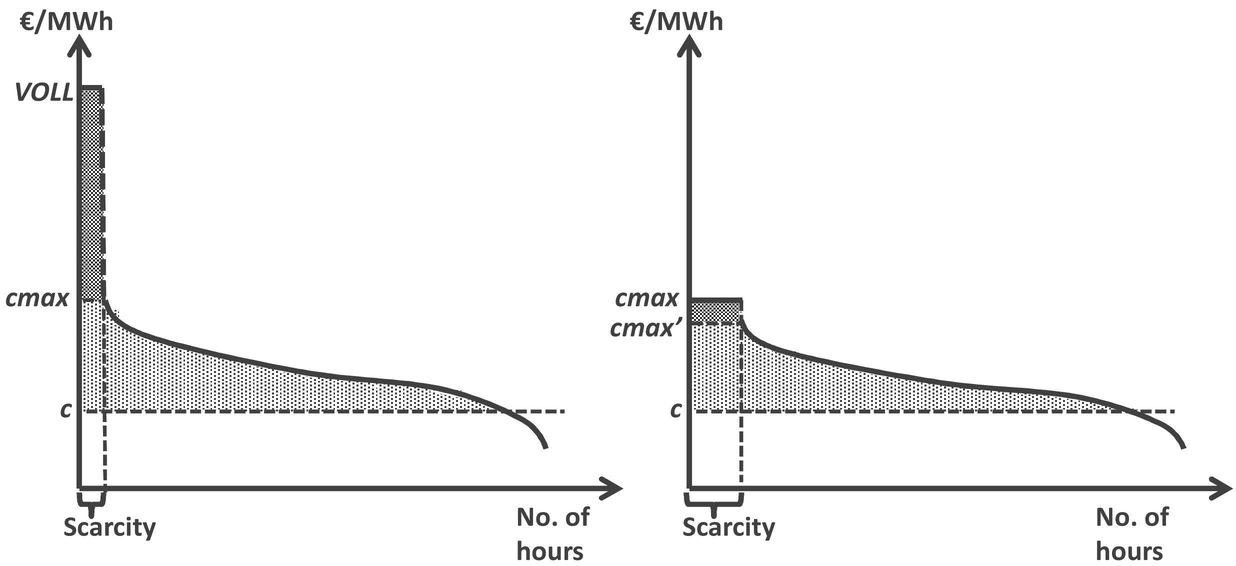
Market power mitigation measures

As we discuss in paragraph 1.2.2, wholesale electricity markets are particularly vulnerable to the exercise of market power when existing capacity is close to full utilization. When demand approaches the level of available capacity, even relatively small generators enjoy market power. Since both electricity supply and demand are to a large extent price-inflexible, withdrawing even a small amount of capacity from the market when the system is tight can be very profitable for a generator, as it may result in a dramatic increase of the market clearing price. This happens especially if capacity withdrawal results in a scarcity situation, i.e. if the market clearing price jumps from the marginal cost of the most expensive generating unit to the much higher VoLL.

Market power mitigation mechanisms may cause under-remuneration of capital invested in the efficient level of generation capacity. Consider for example an overall price cap set equal to the marginal cost of the most expensive existing generating unit.

Such measure would reduce the incentives for generators to withdraw capacity when the system is tight, to the extent that the withdrawal would not result in an extremely high VoLL price. However, because of the cap, the expected revenue for the generators falls short of what is necessary to attract an efficient level of investment. This situation is illustrated in the Figure 3, where we show a price duration curve. The price duration curve shows, for each price level, the number of hours the market clearing price is above that level.

Figure 3 - Price duration curves



The area with the grey background represents the profit that 1 MW of generation capacity with variable cost *c* obtains on the wholesale electricity market[10]. The profit is equivalent to the sum of the differences between the market price and the generator's variable cost in all hours. Over the lifetime of the generator total capacity settles to the efficient level and composition and electricity and ancillary services' prices are such that the efficient capacity level obtains the standard return on investment.

If, during conditions of scarcity, the price is set to be equivalent to the most expensive generator's cost (*cmax* in the figure) instead of the VoLL, the generators' profits are reduced by the darker area[11]. Notice that, because of the cap, the most expensive unit does not receive any contribution to fixed costs from selling energy. In the long run standard profitability conditions are re-established via the entry/exit process. Capacity settles at a lower level as the units with variable cost *cmax* are not replaced and the new system marginal cost becomes *cmax'*. This means that the number of scarcity hours increases until the new dark area is large enough to cover the generators' fixed costs[12], as shown in the right panel of Figure 3.

When the scarcity price is below the VoLL and the wholesale market is competitive, a mechanism integrating the generators' revenues is necessary to ensure that the efficient capacity level is available in the system at all times.

In some markets price caps below common estimates of VoLL are imposed. This happens for example in Nordpool (the Nordic power market), in the Australian day-ahead market and in the US in ERCOT (the power market operating in Texas). The Australian market and ERCOT also have additional price mitigation measures which limit the duration of elevated scarcity prices. If prices remain above a pre-defined threshold for a certain period of time a price cap is enforced in Australia, and the normal price cap is lowered in ERCOT.

An alternative approach to market power mitigation, involves capping the generators' offers at times when they are considered to enjoy significant market power. However, when the system is tight it is very hard to distinguish between high prices that reflect a genuine situation of scarcity and high prices that are the result of exercise of market power. This was particularly evident in

---

[10] For the sake of simplicity we omit reference to the revenues from selling ancillary services.

[11] In the example we ignore the revenues that generators obtain by providing ancillary services, and we refer to electricity spot market sessions only. These simplifications are irrelevant provided the same price cap is consistently enforced on all services and market sessions.

[12] We have assumed that the price cap has not been adjusted to the new (and lower) system marginal cost. If this happened the installed capacity would continue to shrink.

the aftermath of the 2000-01 California power crises where, according to some observers, high loads and low water availability for electricity production, combined with market manipulation, resulted in extremely high prices. In other terms, market power mitigation mechanisms based on selective capping of the generators' offers may also be activated in situations of genuine scarcity and create revenue deficiencies for the generators. For this reason, such measures are typically associated with capacity remuneration mechanisms providing an additional source of revenues to the generators.

Too small VoLL

As we illustrated in paragraph 1.2.1, when the price-insensitive portion of demand exceeds available generation capacity, involuntary load reduction via disconnections, or load shedding, may become necessary. When scarcity occurs and demand is totally price-inflexible, depending on the scarcity pricing rule implemented (see Box 1), either the price for electricity is set to an administratively defined value, the VoLL, or the VoLL operates as a ceiling to a clearing price based on the highest offer submitted in the market. The same holds, mutatis mutandis, for the price for operating reserves.

Implementing VoLL pricing and load curtailment is not without problems. Firstly, load curtailment is perceived by end consumers as being unfair. Curtailed consumers typically do not receive payments equivalent to the VoLL from their suppliers, while non-curtailed consumers are not charged for the VoLL. Secondly, although each consumer might give a different value to electricity, current technology makes it impossible to selectively disconnect consumers based on their individual valuation of electricity, for small consumers. It is then impossible to provide incentives to those consumers to reveal their individual valuation for reliability. Finally, load shedding and price spikes rapidly become a matter for political concern.

Given that the generators rely on the extremely high prices prevailing during very few hours of scarcity to cover fixed cost, if the process that sets the price for electricity in the event of scarcity results in too small a VoLL, then incentives to invest may be curbed.

Out-of-market procurement of reserve services

Some power markets have developed out-of-market backstop mechanisms for ensuring reliability and sufficient capacity. In most cases, market operators simply procure reserve capacity outside the market framework if they expect peak capacity to be short of their targeted reliability standard. In Nordpool, for example, when capacity is forecast to be insufficient on a 3-year forward basis to meet the reliability target, the transmission system operator is authorised to procure peaking resources under long-term contracts with the costs of the procurement paid for by the state[13]. In the UK operating reserve is procured by the system operator under long-term contracts of up to two years in order to provide sufficient investment signals to providers and allow enough time for the repayment of a provider's investment[14]. In some US markets out-of-market capacity purchases have been made in the form of reliability-must-run (RMR) contracts, which are intended to retain in the system capacity resources that might otherwise be retired or mothballed.

Such backstop measures may displace market-driven investment in generation capacity and inhibit the development of demand-response measures in energy-only markets if they prevent market prices rising to VoLL during scarcity events.

Lack of transparency

Some markets do not implement scarcity pricing rules that administratively set the electricity price to VoLL when scarcity conditions are detected by the system operator. These markets rely on generators to increase their bid prices above marginal costs in order to set scarcity prices. For this mechanism to be effective, it is crucial that generators are in a position to correctly anticipate scarcity situations. Lack of information about the demand and supply may cause some scarcity situations to go undetected by the market participants, exacerbating the missing money problem.

Sequential markets

When multiple related markets are cleared independently, the profit-maximizing bid for a generator in one market depends on the expected equilibrium price in the other markets where the same or a related product is traded. For example if the wholesale energy market and the operating reserve market for products delivered at the same hour t are cleared independently, arbitrage between the two markets should result in consistent clearing prices, both reflecting the

---

[13] See Nordel, 2007. *Guidelines for Implementation of Transitional Peak Load Arrangements: Proposal of Nordel*, http://www.svk.se/Global/01_Om_oss/Pdf/Elmarknadsradet/071115NordelGuidelines.pdf.

[14] See National Grid Electricity Transmission, 2013. *Short Term Operating Reserve (STOR) Frequently Asked Questions*, p.3,http://www.nationalgrid.com/NR/rdonlyres/F393E0E3-26B2-49D5-80BD-409CE2093B4E/59391/ShortTermOperatingReserveSTORFAQ.pdf.

overall supply and demand conditions for electricity and operating reserve. Efficient arbitrage would result in energy and reserve clearing prices differing, ceteris paribus, by the variable cost of the marginal unit.

However, some elements of the market designs may make arbitrage between the various markets difficult. First, energy and reserve markets are generally cleared independently in Europe.

Second, the design of the day-ahead, intraday, balancing and reserve capacity markets is not always homogeneous. In particular products traded in the different market may have different designs. This can make it difficult for market participants to assess the relationship among the prices for the products traded in the different venues.

Third, speculative trading against real-time prices is inhibited in some markets by punitive imbalance charging systems.

Imperfect arbitrage across market venues may result in scarcity conditions not being reflected in the same way in different markets. In some European countries ancillary services, in particular operating reserve, are procured by the system operator close to real time, after energy markets clear. In this situation, if information on (expected) demand and supply of ancillary service markets is not available. Generators may fail to anticipate the scarcity situation when formulating their bids in the energy market. In that event the electricity market clearing price will not signal scarcity because available generation capacity is greater than the demand in the energy market. Scarcity conditions will however emerge – and result in high prices – in the operating reserve market. The generation capacity committed on the electricity market will then receive an inefficiently low price, which does not correctly reflect its value.

Opaque system operator behaviour in scarcity hours

The prices of electricity and ancillary services could be biased by some system operator's actions. In particular, system operators may - and occasionally do - implement out-of-market measures that could depress wholesale prices during scarcity conditions.

Using an example from the US, Pfeifenberger et al 2009[15] mention curtailment actions called for by the system operator, which do not trigger VoLL pricing of energy and ancillary services and

[15] J. Pfeifenberger K. Spees, and A. Schumacher, 2009. *A Comparison of PJM's RPM with Alternative Energy and Capacity Markewt Designs*, The Brattle Group.

the practice of dispatching power from generators in violation of environmental constraints in emergency situations. Joskow 2006[16] points at voltage reductions implemented in order to delay or avoid blackouts, which again do not trigger VoLL prices.

Finally, the system operator exercises his discretion in the assessment of interconnection capacity in a zonal- (or interface-) based congestion management system. [17] Assessing interconnection capacity across interfaces entails the exercise of the system operator's judgment as to the pattern of injections and withdrawals which will obtain at equilibrium. By slightly relaxing some of the constraints involved in the assessment of interconnection capacity the system operator may prevent scarcity conditions – which would trigger cumbersome emergency procedures – from emerging at the market clearing stage, which would lead to inefficiently low prices.

Expansion of the renewable generation fleet

Over the last years a rapid change in the structure of power generation has taken place as a result of the political decisions to dramatically increase the share of renewable generation. The low marginal cost of renewable generators changes the shape of the supply function, by adding large volumes of capacity with close-to-zero variable cost, often with guaranteed remuneration for supply.

The politically driven development of renewable generation capacity increases risk for the investors in non-renewable generators in two dimensions. First, the increase in total capacity resulting from the political decision on renewable production targets can result in excess capacity, reducing utilization factor and margins of conventional generators. In the policy discussion the notion of missing money is occasionally stretched to include major revenue shortfalls for power generators caused by political intervention on renewable production. According to this approach, traditional generation capacity displaced by renewable production would be entitled to protection because the development of renewable generation capacity: a) was not foreseeable when the displaced capacity was built and b) has not been market driven but the result of a political decision. Discussing this approach goes beyond the purpose of this paper as it appears to reflect more a stranded-cost line of reasoning than genuine capacity

[16] P.L. Joskow, 2006. *Competitive Electricity Markets and Investment in New Generating Capacity*, MIT

[17] D. Perekhodtsev and G. Cervigni, 2011. *Are Flow-Based capacity allocation systems well suited for the integration of the European power market?* IEEE.

adequacy issues.

Second, renewables induce an increase in spot price volatility, since very low prices obtain in the hours when renewable generators are marginal. In that context, conventional generators will face a riskier environment as very-high-price hours become (even more) crucial to gain returns on investment. That exacerbates the issues related to pricing in scarcity conditions discussed above.

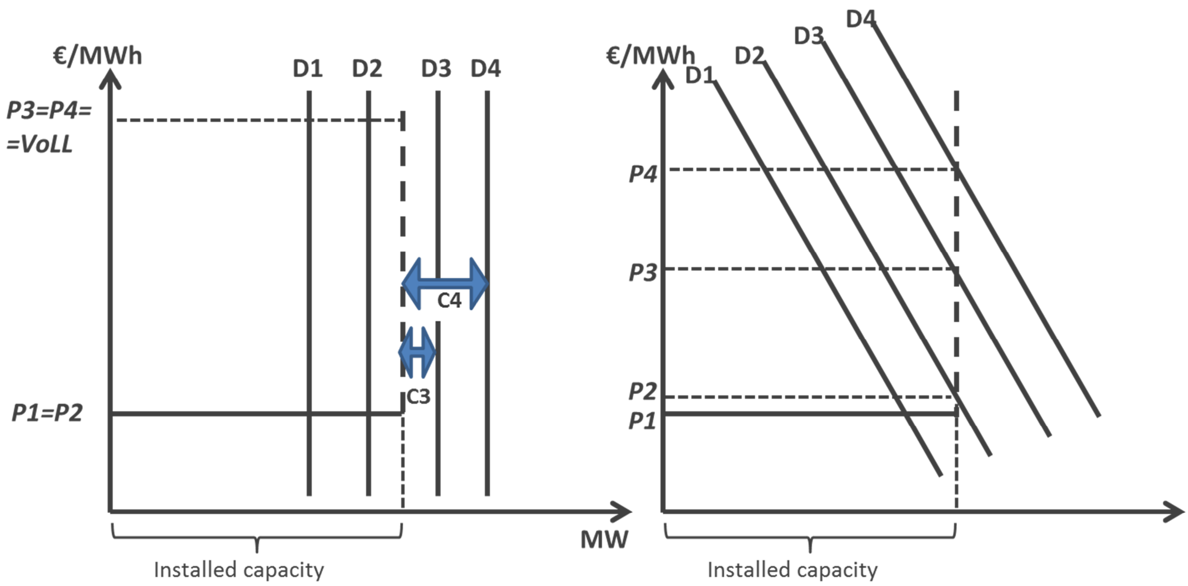
Demand response

Generation capacity adequacy concerns mostly relate to distortions in the market outcome in the event of scarcity. The insensitivity of a large share of demand to prices of in the short-term causes:

- the need for quantity rationing in case of scarcity and for a regulated price (or a price-cap) in scarcity hours;
- the extreme gap between (relatively low) prices in normal conditions and the VoLL in scarcity conditions.

Greater price-responsiveness of demand would mitigate those features of wholesale power markets, possibly reducing the need for capacity support. This is shown in Figure 4.

Figure 4 - Prices in case of scarcity with inflexible and price-responsive demand



As shown in the Figure, in case of scarcity with inelastic demand the price jumps from the system marginal cost to VoLL and involuntary service interruption occur (arrows C3 and C4). In contrast, with price responsive demand, the market clearing price when the system is tight can take any values between the system marginal cost and the VoLL; in addition those consumers that give up consumption – because of the high price – do that voluntarily.

Currently only large consumers can be exposed to wholesale spot prices and therefore be provided with the correct economic incentives to shape planned consumption consistently with the prices of electricity at the different times and to adjust those plans in response to price changes occurring as the time of delivery approaches. An even smaller subset of consumers is allowed to participate in real-time markets, i.e. to change withdrawals at in response to the real time prices. Further, typically the service provided by these consumers (interruptibility) entails giving up consumption at the system operator’s notice when critical technical conditions occur, to a large extent irrespective of prices.

The large scale deployment of smart-metering systems, expected to take place in Europe by 2020[18], will enable all consumers to plan and adjust power consumption in response to short-term prices. Whether the net benefits of such behaviour are positive for a material share of the final consumers, so that a significant portion of the market’s demand will become flexible, is still to be tested.

1.3.2 Coordination

The high level of risk in generation investments is sometimes mentioned as a reason for the introduction of capacity remuneration mechanisms (see for example Crampton and Ockenfels 2011)[19]. In this respect, the relevant feature of some capacity remuneration schemes is that they coordinate market participants’ investment decisions in generation capacity. To the extent that such coordination reduces the uncertainty faced by generators, it also reduces the required rate of return on the investment in generation capacity. As a consequence, all other things being equal, a higher level of capacity will be installed.

Coordination failures can arise in all markets and do not typically trigger the introduction of capacity remuneration schemes. The argument in favour of such schemes in power generation

[18] According to Directive 2009/72/EC: “*where roll-out of smart meters is assessed positively, at least 80 % of consumers shall be equipped with intelligent metering systems by 2020*”

[19] P. Cramton and A. Ockenfels, 2011. *Economics and design of capacity markets for the power sector*. Zeitschrift für Energiewirtschaft, 36:113-134, 2012

would be that in power markets the cost of lack of coordination can be extremely large. For investors, because small prediction mistakes about the future demand and supply condition may have a dramatic impact on their profitability. For consumers because small hiccups in the capital accumulation process may lead to large (and costly) load losses.

In case coordination is the main concern addressed through a capacity remuneration scheme, the main purpose of the measure is governing the development of the generation fleet.

In order to illustrate the effect of coordination failure on the level of risk in generation investment, consider an investor assessing the opportunity to invest in a 1,000 MW plant to be in service in year *t*. The profitability of this investment crucially depends on the decisions of other potential investors to enter the market. If just two new plants were brought into service in year *t* instead of only one, electricity prices might turn out much lower than if only one were built. The impact of the second project on market prices could be considerable for several years, potentially undermining the profitability of both projects.

Although an efficient investment pattern would include 1,000 MW additional capacity in service from year *t*, each investor will want to reduce its investment's vulnerability to other investors' decisions. Such a strategy can be expected to lead to investments being delayed when compared with the efficient path[20]. If this happened, an inefficiently low level of available capacity would be available in year *t*.

In this context, a mechanism that coordinates investment decisions could be beneficial.

The interaction between generation and transmission capacity adds a further dimension to the coordination issues among investors in the electricity industry. In the traditional vertically integrated monopoly or quasi-monopoly organization, investment in generation and transmission capacity was typically coordinated by a central entity. As a consequence of transmission and generation unbundling the decisions to build new generation and transmission capacity are made by different entities.

The locational differentiation of energy and ancillary service prices drive the siting decisions of generation capacity and the development of the transmission system. However, coordination

[20] Coordination failures might also result in overinvestment, if investors are risk neutral. We cast the problem in terms of underinvestment because a waiting strategy appears more likely to prevail in practice. In any event, the risks related to possible coordination mistakes will be reflected on the rate of return required by the investors in generation capacity. Higher expected rate of returns will in turn lead to a lower equilibrium level of generation capacity.

issues between investors in generating capacity and the institution responsible for developing the transmission system, typically the system operator, may occur. As a result too much generation capacity might turn out to be developed in areas where it is not needed and not enough in the areas where it is or, in an alternative interpretation, the transmission system might not be updated in time to allow electricity to flow from net-surplus areas to load centres[21].

1.4. Discussion

Missing money and coordination issues may provide a rationale for the introduction of a capacity remuneration scheme. However, public measures to govern investment in power generation capacity will generally come with their own imperfections. In this section we discuss a possible approach to assessing the need for or opportunity of generation capacity remuneration measures.

An empirical assessment of the opportunity to introduce generation capacity remuneration measures in Europe is beyond the scope of this paper. Further, we conjecture that such an exercise would hardly provide solid evidence in favour or against generation capacity support for several reasons. First, evidence of, respectively, scarce or excess capacity at a certain point in time does not by itself prove that capacity support is or is not necessary. Building generation capacity takes a long time and major demand as well as supply shocks can and do happen. Therefore prolonged situations of excess or scarce supply are not exceptional in the power industry. Second, assessing the opportunity for a capacity remuneration scheme requires figuring out the supply conditions that would be produced by the measure. Such counterfactual scenario is likely to be hard to identify and highly uncertain. Finally, in most European countries central elements of the electricity wholesale and ancillary service market design are still being developed, within the context of the Target Model implementation[22]. As a consequence the need for capacity remuneration schemes must be assessed with reference to a somewhat untested market design.

[21] Measures to integrate network planning across European countries are being developed in Europe. For example a ten-year pan-European network development plan is produced by ENTSO-E, the European network of transmission system operators.

[22] See, for example, the Framework Guidelines on Capacity Allocation and Congestion Management (CACM) for Electricity and the respective network codes, http://www.acer.europa.eu/Electricity/FG_and_network_codes/Pages/Capacity-allocation-and-congestion-management.aspx

For those reasons assessing whether market dynamics are such that adequate investment in generation capacity is attracted and therefore whether some form of capacity remuneration is necessary is, in all but very extreme situations of persistent scarcity, very hard.

However, insights into the assessment of the need for generation capacity remuneration may be gained by evaluating the relative likelihood and cost of two errors:

- implementing a capacity remuneration mechanism when that is unnecessary. i.e. when market dynamics would result in the efficient level of capacity being built;
- not implementing a capacity remuneration scheme when needed.

As to the cost of unnecessary capacity remuneration, well-designed mechanisms - with capacity targets set well in advance of delivery and competitive selection of the suppliers - may generate very little cost on the consumers in case the target level of capacity would have been reached even without support.

However, less than perfect designs have historically been implemented in several markets. For example, the first generation of ICAP mechanisms implemented in the United States attracted much criticism of their inability to attract investment and vulnerability to the exercise of market power by the existing generators (Battle and Rodilla 2011)[23]. Some capacity payment schemes implemented in Europe, because of their short term nature, are more likely to implement a surplus transfer between the consumers and the existing generators then to impact on the investment decisions in additional capacity. The first version of the capacity payment scheme implemented in the UK until 2000 allowed manipulation by the generators. Distortive effects on cross-border trade have been identified in the Irish and Russian capacity remuneration schemes.

Finally, planning approaches to govern the evolution of the generation fleet have historically shown their own flaws, to the point that dissatisfaction on the regulated utilities investment strategies have been pointed at, in some countries, as one of the main reasons for liberalization. We will discuss the features of an effective capacity remuneration scheme in the next chapter.

More generally, a capacity remuneration scheme places the public sector (the regulator and/or the system operator) in charge of deciding the size, location and possibly composition of the power generation fleet. To the extent that the public authority's choices do not coincide with

---

[23] C. Battle and P. Rodilla, 2010. "*A critical assessment of the different approaches aimed to secure electricity generation supply*", Energy Policy 38 (2010) 7169–7179

those that market investors would autonomously make, the measure's scope extends beyond mere coordination to a planning system. At that point a potentially large share of the investment risk is transferred to consumers, through the capacity remuneration scheme. In this perspective the discussion on capacity support measures boils down to the (largely political) question of whether governments or private investors should decide on the level and possibly type and location of investments in electricity generation capacity[24].

As to the cost of not implementing capacity a remuneration scheme when needed, on the one hand the (social) cost in case demand cannot be met is very high, given the high value of electricity for consumers. Further, since selective disconnection of small consumers in case of scarcity is currently unfeasible, the VoLL is likely to provide, at least for those consumers, a highly imperfect representation of the value of electricity for many consumers.

On the other hand, in some European countries large excess capacity conditions currently hold and are expected to persist in the future, because of the demand reduction caused by the economic crisis and the growth of renewable capacity. This might prevent the cost of a missing capacity remuneration scheme from emerging in the foreseeable future.

In any event, capacity remuneration mechanisms should not be regarded as a substitute for a sound market design and a predictable regulatory environment. Policy attention should therefore focus on eliminating any market design flaws or unnecessary regulatory uncertainty, independently of capacity remuneration measures.

Public authorities may deem that some regulatory measures potentially resulting in missing money produce benefits outweighing their cost. This could be the case, in particular, for market power mitigation measures such as an overall price cap below VoLL, whose drawback is preventing the market clearing price from rising to the efficient level in scarcity situations. Bid caps might have similar effects if not combined with a flawless triggering system and an effective scarcity pricing rule. If those market power mitigation measures are implemented, the case for integrating the generators' income to offset the compression of scarcity rents caused by market power mitigation is stronger.

---

[24] For example, since the public decision makers and the system operator don't bear the cost of capacity support – but would pay the political cost of power shortages – they might have an incentive to set an unnecessary large capacity target. Such argument, which we will not investigate further, echoes the allegation of infrastructure gold-plating recurring in the debate on liberalization and privatization of traditional electricity monopolies.

However, missing money might result from unnecessary flaws in the design of wholesale power markets. These include, for example, ineffective scarcity pricing systems, too low VoLL, poor information available to market participants on demand and supply conditions, system operator practices preventing scarcity situations from showing, heterogeneous designs of market venues operating at different times ahead of delivery. Such design flows may cause inefficiencies besides and beyond capacity adequacy issues. Therefore they are hard to see as valid justifications for a capacity remuneration scheme. Rather, policy efforts should focus on removing them, or alleviating their impact.

The same holds for some sources of risk and coordination failure in generation capacity investments, such as renewable generation support policies, which appear to some extent under the public authorities' control. Risk and coordination failures may be mitigated by ensuring that market participants have access to all available information relevant to assess future demand and supply conditions. In addition, provided renewable targets and network development decisions are announced to the market well ahead of implementation, their impact on the existing generators' profitability may be mitigated and wrong investment decisions may be avoided. Whether (the remaining) investment risk still justifies implementing a capacity remuneration mechanism is ultimately a matter for empirical investigation.

2. Capacity remuneration mechanisms

In this Chapter we illustrate how capacity remuneration mechanisms address missing money and coordination issues and we discuss the main features of alternative schemes.

2. 1. How capacity remuneration mechanisms work

In this section we discuss how capacity remuneration mechanisms work, by illustrating two extreme schemes: a simple capacity payment scheme and a "single-buyer" system, in the meaning of the first European Commission electricity Directive (96/92/EC).[25] We refer here to stylized capacity remuneration mechanism, rather than to real world examples. In the following sections we discuss in greater detail the features of alternative capacity remuneration schemes and some implementations.

2.1.1 A stylized capacity payment scheme

In this section we outline a capacity remuneration mechanism focused on addressing missing money issues. We assume a market in which an overall price-cap set equal to the system marginal cost – with market power mitigation purposes – causes a missing money problem[26]. As shown in section 1.1.3, such a price cap causes, in the long-run equilibrium, missing revenues corresponding to the fixed cost of the marginal unit, in most systems presently an open cycle gas turbine, the variable cost of which sets the price cap.

Assume that the regulator sets a capacity payment paid to all available capacity, in the form of €/MW per each hour of availability. The capacity payment level is such that a regularly available generator obtains revenues equal to the fixed cost of the marginal generator[27].

The generators' profit maximizing offer strategy in our exemplary energy and ancillary service markets, and the corresponding revenues, are not modified by the capacity payment scheme. However, each generator's total revenues are increased by the fixed cost of the marginal unit. This would neutralize the impact of the price-cap on the generators' income in the long run

---

[25] See: Directive 96/92/EC of the European Parliament and of the Council of 19 December 1996 concerning common rules for the internal market in electricity, article 2: "'*single buyer` shall mean any legal person who, within the system where he is established, is responsible for the unified management of the transmission system and/or for centralized electricity purchasing and selling*"

[26] In this section we ignore, for simplicity, the revenues collected by the generators by supplying ancillary services. This is without loss of generality, under our assumption on the cause of missing money.

[27] We ignore at this stage issues related to outages, as well as the incentives provided by a uniform payment in all hours. We will address these issues in the next section.

equilibrium, and therefore re-establish the correct investment incentives.

In addition, the mechanism provides (a little) risk hedging to generators, since what would be uncertain scarcity revenues in a market without price-cap becomes a certain capacity payment[28].

However, the mechanism does not contribute to improving coordination among would-be investors in generation capacity since it does not convey to them any information on the current and future demand and supply conditions.

2.1.2 A stylized single-buyer scheme

In this section we outline a model in which the involvement of public authority[29] in shaping the evolution of the generation fleet is greater. In this approach the public authority sets the capacity requirement, possibly by type of technology and location. Then, an auction is run in which the public authority, on behalf of electricity consumers, purchases the target capacity. The generators contracted in the auction sell electricity and ancillary services in the market, but, in exchange for the capacity remuneration, commit to return to the public authority any infra-marginal rent collected from selling electricity in the energy and ancillary service markets.[30]

This scheme addresses both missing money and coordination issues. The generators' entire fixed cost is covered by capacity revenues[31]. In addition the generators are fully hedged, as – via the public authority – consumers pay the entire fixed cost of the generation capacity irrespective of its actual use. No or very little risk is taken by market investors, as generators are left at most with some operation risks, in order to incentivise the efficient use of capacity.

Full coordination of generation, and possibly transmission, investment is obtained, as capacity development decisions are taken by the public authority. In this model the regulator decides on almost all facets of generation capacity investment, including level of installed capacity; fleet composition and location. Competition is relied upon just to procure the target capacity at minimum cost.

[28] As long as the regulator does not modify the level of the capacity payment

[29] We refer here to the "regulator" in general terms, abstracting from the allocation of responsibilities among the government, the regulator and the system operator.

[30] We ignore all issues related to the incentives to operate efficiently the generating units.

[31] The competitive offer price in the capacity auction is, for each technology, its fixed cost.

2.1.3 Conditions for the effectiveness of capacity remuneration mechanisms

The stylized mechanisms presented in the previous sections point at crucial features of capacity remuneration scheme. With respect to the missing money problem, for a remuneration scheme to induce additional investment, investors must believe that the remuneration measure will be maintained over the entire economic life of the investment. In case of capacity payments, public authorities have to credibly commit to granting (the expected level of) capacity payments indefinitely. Any risk that the scheme be cancelled during the generator's economic life will increase the rate-of-return required by the market to sink capital in the industry and ultimately result in lower equilibrium level of capacity and lower reliability.

In our stylized single buyer model, the public authority commits consumers to paying the generators' entire fixed cost, irrespective of their actual use. Therefore, to the extent that this contractual commitment is firm, all risk is moved from investors to consumers and the target level of capacity is procured at the relatively low cost of capital corresponding to such risk allocation.

With respect to coordination issues, for a remuneration scheme to reduce the likelihood of coordination failures it must convey to would-be investors information on the future demand and supply conditions. The stylized capacity payment scheme we have just described does not contribute to coordinating investment decisions. Consider the example of coordination failure described in section 1.3.2 above. A capacity payment scheme would increase each investor's expected income but it would not contribute to selecting or to informing self-selection by one of the two investors.

In contrast, the single buyer system selects the party that will make the target capacity available. In our example, the target level at time t would be such that only the additional 1,000 MW would be selected in the auction for capacity run by the public authority. Even if development of generation capacity outside the public procurement scheme was allowed, investors would be unlikely to sink money into making additional capacity available at time t, as they know that such an investment would lead to excess capacity overall, and would therefore be unprofitable.

Here the capacity remuneration scheme coordinates the timing of the investors' decisions. The capacity remuneration scheme acts mainly as a coordination device. In our example very little or no compensation might be required by the winner of the auction, if the central entity auctions off only 1,000 MW of incremental capacity. In this case, each of the potential investors would find it profitable to make 1,000 MW capacity at time *t* with no further compensation, provided no more than 1,000 MW new capacity is built in total. What the auction process delivers is only the certainty as to who will make the investment. If instead, as typically happens, the capacity target pursued by the central entity is greater than the level that would attract the investment, then the auction will clear at a higher price.

Note also that the re-allocation of risk from the generators to the central entity acting on behalf of the consumers is a by-product of this measure, not the source of its expected welfare-improvement.

In this section we have shown that stability over time and ability to coordinate investment decisions are crucial to the effectiveness of any capacity remuneration mechanisms. Three broad approaches to the design of capacity remuneration schemes can be identified. The first approach sets the price that a central entity, on behalf of the consumers, commits to pay for all available capacity. The second approach sets the volume of available capacity that the central entity commits to paying for, either directly or by placing an obligation on the load-serving entity. The third approach consists of reserving a certain generation capacity for use only in scarcity situations, as a substitute for load curtailment. In the rest of this chapter we discuss each approach in turn.

2. 2. Capacity payments

Capacity payments are administratively set payments per MW available capacity, paid to all generators regardless of whether they are dispatched to run. Capacity payments are intended to provide generators with additional revenues equivalent to the missing money. The composition and location of the generation fleet are left to market forces.

Different capacity schemes grant payments for capacity availability at different times. For illustration purposes we will consider two extreme methods. The first methodology would pay 1/8760 of the annual fixed cost of the marginal unit for all capacity that turns out to be available during each hour of the year[32]. The drawback of this approach is that it does not provide incentives to make capacity available when the system needs it most. In case of scarcity

[32] We ignore maintenance stops for reasons of simplicity. The correct assessment would allow each generator to obtain the annual fixed cost of the marginal unit in a number of hours equal to the difference between 8760 - the number of hours in a year - and the duration of a standard maintenance period.

generators would receive only the system marginal cost and a (relatively) small capacity payment. Total revenues in case of scarcity would then be well below VOLL, the level providing correct incentives to be available.

The second extreme methodology would pay 1/N of the annual fixed cost of the marginal unit for all the capacity that turns out to be available in each of the N hours of the year when the system operator expects scarcity. The advantage of this approach is that it provides stronger incentives for generators to make capacity available when the system is expected to need it most. If the N hours selected by the system operator, and only those N hours, turn out to be scarcity-hours, this methodology provides exactly the same incentives that would be provided by an unbiased energy-only market. Further, the scheme has an additional advantage over VoLL scarcity pricing: it removes from generators the burden of predicting when scarcity will occur, for example in order to schedule maintenance outages, as the capacity payment will be paid irrespective of actual load.

A drawback of the scheme is that it does not provide incentives to make capacity available in case of scarcity situations occurring in hours not included in the critical set identified in advance. Therefore effectiveness of the scheme depends on the system operator's ability to predict when scarcity conditions will occur. Typically critical system conditions could manifest 5-20 hours per year, and predicting, for example a year in advance, when those hours will be is a difficult exercise. The trade-off between the power of the incentives to make capacity available and the availability of capacity at the right times is addressed in practice by granting capacity payments in exchange for availability over a relatively large subset of a year's hours, for example one or two thousand, when demand is expected to be high. In Chile, for example, capacity payments are granted for availability in the months May-September; in Colombia, in the dry December-April season when hydropower production is limited[33]. In Italy the set of around ninety critical days when capacity availability will be remunerated is set yearly by the system operator.

Capacity payment schemes have been introduced in several countries after the electricity sector liberalization. In the boxes below we provide an overview of the main features of capacity payment schemes that are still in place in Spain and Italy and that were implemented and subsequently abandoned in UK and Argentina. Rather than providing an exhaustive description

[33] In Chile a penalty for failure to deliver capacity based on the VoLL re-establishes the correct incentives for the generator to make capacity available in the scarcity hours.

of the schemes, we focus hereunder on the features of the mechanisms which result in departures from the theoretical models.

Spain

The capacity payment scheme currently implemented in Spain has two components. The first component (or investment incentive component) provides capacity payments to new capacity or to significant upgrades to existing capacity. For capacity installed between 1998 and 2007 the investment incentive component is fixed and expires after 10 years. For capacity installed after 2007 the capacity payment depends on a measure of the reserve margin, assessed on a yearly basis. The higher the reserve margin, the lower the capacity payment provided.[34]

Granting different payments to existing and new capacity is sometimes supported in policy discussions because it reduces the initial amount of capacity payments compared with capacity payments being granted for the entire capacity. However, it is distortive, since the new capacity attracted by the selective capacity payment will exacerbate the missing money problem for existing generators that do not receive the capacity payment. This will accelerate substitution of the generating fleet faster than is efficient. Moreover, capacity payment based on reserve margin place more risk on generators, as the actual path of capacity payments may deviate from the one expected when the investment is sunk.

The second component (or availability service component) is granted to all units available for dispatch. A distinctive feature of the Spanish mechanism is that the availability service component is differentiated according to the generator's technology.

Differentiation of payments according to the technology could be interpreted as a way for the public authority to influence the composition of the generation fleet. We will discuss this feature of capacity remuneration mechanisms in section 3.2, as a way to address the larger need for flexible generation resources caused by the expanding fleet of renewable intermittent generators.

[34] This scheme still needs to be fully implemented because the mechanism for the calculation of the reserve index (or Indice de Cobertura) has not yet been approved by the regulator. See: CNE, 2012. "*Pagos por capacidad en España: situación actual y perspectivas*", p. 7, http://www.ariae.org/download/reuniones/XVI_Reunion_ARIAE_2012/Pablo%20Villaplana%20%20Pagos%20por%20capacidad.pdf

Italy

In Italy, an administratively set capacity payment is granted to conventional generators available for dispatch during a set of “critical” days, identified in advance by the system operator.

The payment is made of two components. The first component depends on the capacity expected to be available during a set of “critical” days. The higher the estimated capacity available during the “critical” days the lower is the capacity payment. The second component is granted only if the total revenue paid to a plant during all the “critical” days falls below an administratively set threshold.

The Italian mechanism has been criticized for two reasons: first, payments are smaller than the annual fixed cost of the marginal unit; second, the limited ability of the system operator to identify correctly one year in advance the days when scarcity is likely to occur.

British pool (1990-2001)

In the British pool system, between 1990 and 2001, a capacity payment was paid to all generators available for dispatch, irrespective of their activation. The capacity payment was computed as the expected value, over the probability distribution function of the demand realisations, of the scarcity rent on the day of delivery. More precisely, the value of the payment for each half-hour dispatch period was computed for each generator the day before delivery, equal to the loss of load probability, multiplied by the difference between VoLL and the bid price (if the generator is not dispatched) or the system marginal price (if the generator is dispatched). As a result capacity payments would be low when available capacity was high compared with load, and payments increased as the reserve margins shrank.

Design issues made the first implementation of the mechanism vulnerable to manipulations. Generators were able to inflate the capacity payment by withdrawing capacity in the day ahead and then collect the (inflated) capacity payment also on the with-drawn capacity, by re-declaring it available during the day.

Argentina

In Argentina a capacity payment introduced in 1995 was differentiated into two components (see Battle and Rodilla 2010)[35]. The first component was computed in a way similar to the one implemented in the British pool discussed in the previous box, with the important difference that only dispatched generation capacity would receive the payment.

The second component was provided to capacity that was shut down during most of the year, but that was kept in service as back up capacity to be used during dry years. This second component is closer to the concept of "strategic reserve" discussed in section 2.4.

Since the revenues from capacity payment depended on actual production, generators had an incentive to bid below variable costs. The structure of the scheme was later modified in order to address this issue.

2. 3. Capacity requirements

A capacity remuneration system based on capacity payments leaves installed capacity levels uncertain, since they depend on the market response to the administratively set capacity payments. An alternative approach is imposing a reserve margin requirement on all electricity retailers. In the following sections we first discuss the standard model in which the suppliers of capacity assume the obligation to make production capacity available; we then present a mechanism in which the obligation on capacity suppliers is financial.

2.3.1 The standard model

In this approach the system operator sets the capacity requirement, based on engineering security standards. The required level of total installed capacity is generally set around 115%-118% of the peak load[36].

The reserve requirement is then divided between retail suppliers in proportion to the expected contribution of their clients to peak load. Each retail supplier is responsible for procuring capacity entitlements that exceed its predicted peak load by the required reserve margin, either

[35] C. Battle and P. Rodilla, 2010. "*A critical assessment of the different approaches aimed to secure electricity generation supply*", Energy Policy 38 (2010) 7169–7179

[36] The trade-off between the cost of achieving the reliability target and the value provided by that reliability is typically not explicitly addressed.

through self-supply or by contracting available capacity from generators. The obligation placed on load serving entities creates a demand for capacity.

The supply of capacity entitlements (or capacity rights or certificates) comes from generators. Each generator is allowed to sell a maximum volume of entitlements set on the basis of their historic availability record. By selling entitlements a generator commits to make the corresponding volume of capacity available.

The obligation is fulfilled independently of the actual use of capacity. The generator can use capacity to deliver electricity sold bilaterally, on the spot market or on the real time market. Even if the capacity turns out not to be used, the obligation has been fulfilled so long as it has been offered. For that reason energy and ancillary service offers are not distorted by the capacity remuneration scheme. Under-delivery – assessed with reference to the requirements for certified capacity – leads to the application of penalties and, via the unit availability's track record, to a reduction of capacity entitlements assigned in the future.

Placing a capacity requirement on retailers creates the demand for capacity that meets the generators' supply. A market for capacity is then established, where centralized and bilateral trading of capacity entitlements takes place. The price in that market settles at the level that attracts investment in generation capacity up to the system operator's requirement. Under the usual perfect competition assumptions, generators' revenues from selling capacity availability are equivalent to the missing money.

Mechanisms based on capacity requirements have been implemented in several US markets. Earlier capacity requirement systems were implemented in the context of traditionally regulated markets, where integrated utilities carried a regulatory obligation to procure the generating capacity needed to meet the resource requirements in their (exclusive) service areas. The absence of retail competition allowed the utilities to recover the costs associated with that obligation through regulated retail rates. That meant that the need for adjusting the utilities' capacity portfolios via trading was limited to transitory imbalances, since the resource planning requirements were overseen or enforced by the state regulators.

Some features of older mechanisms have caused concern. In particular, enforcing the capacity requirement just days or months before the relevant delivery period may lead to extreme price volatility, with prices jumping from the cap (when there is insufficient capacity) to zero (when there is excess capacity). This happens because in such a short time-horizon both the demand for and the supply of capacity are highly price inelastic. The supply of capacity to deliver within a period of months does not include units not yet built. As a consequence, the entrant's cost does not act as a ceiling to the prices of capacity. In addition, suppliers and possibly buyers of capacity contracts may enjoy significant market power when the system is close to the target resource requirement: suppliers may be able to move prices from close to zero to the cap even by withholding relatively small amounts of capacity. Buyers may similarly be able to move capacity prices levels close to zero by slightly reducing their demand for capacity, for example by declaring that they will meet part of their obligation via self-supply. Finally, if capacity deficiencies are detected only slightly in advance, it may be impossible or extremely costly to the system operator to make up the missing resources.

In order to address those concerns, forward reserve requirements have been introduced in several US power markets. For example in PJM (the power market in Pennsylvania, New Jersey and Maryland) and in ISO-NE (the power market in New England) capacity is procured up to three years ahead of the delivery year. In addition, both PJM and ISO-NE allow suppliers of new capacity to lock-in capacity prices for three to five years.

Retail liberalization may make it difficult to enforce long term commitments on power retailers, given the potential instability of their customer base. For that reason PJM and ISO-NE purchase long-term capacity entitlements from generators. The corresponding cost is then passed on to load serving entities based on the customers supplied. Load serving entities and generators may also enter bilateral transactions on capacity entitlements.

In the PJM and NY-ISO (the power market in the state of New York) markets a certain degree of price-elasticity is also introduced on the demand side by implementing a downward sloping capacity demand curve, which varies the resource adequacy requirements as a function of capacity prices. A flatter capacity demand curve reduces price volatility, as a shift in the demand or supply curve leads to a smaller change in the market clearing price. However, the capacity demand curve implemented in PJM and NY-ISO is not intended to represent the consumers' (estimated) willingness to pay for capacity. Instead, consumers are assumed to be available to pay a price equivalent to the estimated building cost of new peaking resources for an administratively set target level of capacity. Then the slope of the curve near the target level of capacity is based on an administrative judgment. As a result, rather than reflecting consumer preferences, the capacity demand curve basically implements a cost-based cap on the price of capacity.

In order to address transmission congestion issues the capacity requirements are imposed on a zonal or locational basis in some markets, including PJM and NY-ISO.

Requiring resource commitments sufficiently in advance of delivery leaves enough time for either market participants or the system operator to procure additional resources if deficiencies are detected. A long time horizon also gives capacity suppliers enough time to modify their resource development plans, for example by bringing mothballed plants back online, making the capital investments necessary to defer the retirement of other plants, speeding up the development of a new power plant, or developing additional demand response capabilities. As a result the price-elasticity of the capacity supply curve rises, price volatility is reduced and competition in the supply of capacity increases.

Contrary to the stylized single-buyer model discussed in the previous section, in a market with capacity requirements investors have the right to build new capacity in excess of the system operator's requirement. However, the capacity procurement process still may coordinate generation capacity investment decisions if capacity contracts are awarded well in advance of the time of delivery[37]. In this case a would-be investor may make the decision to build new capacity conditional on being awarded a capacity contract.

The capacity procurement process acts de-facto as a selection device, to the extent that it conveys important information to the investors. Broadly speaking, would-be investors failing to sell capacity learn that:

- they are probably less efficient than those who managed to sell capacity entitlements;
- a total level of capacity equal to the system operator's requirement will be built by others;
- in the event that they decided to go ahead with the investment, they would have no hedge at all against excess-capacity situations, leading to too few scarcity hours.

The information conveyed by the capacity remuneration scheme coordinates the investors' decisions. The decision to invest or not to invest is made easier. The investors that sold capacity contracts face strong incentives to invest, since they can be confident that those who did not will not invest and bring about excess capacity; those who failed to sell capacity contracts have

[37] In Colombia, for example, the system operator sets the capacity requirement and procures the corresponding reliability options three years ahead of the start of the commitment period. The commitment period ranges from one to twenty years.

strong incentives not to invest, since they can be confident that they will not miss a profit opportunity because it is likely that their investment will result in excess capacity.

Compared with the stylized single buyer model, a capacity requirement system provides a smaller hedge. The (competitive) equilibrium price for capacity availability is not the entire investment cost but only the missing money component.

2.3.2 Reliability options

Administratively set capacity targets may also be achieved through financial contracts with the suppliers of capacity. Capacity remuneration schemes based on financial obligations are implemented in Colombia under the name of Firm Energy Obligations (OEF). In Europe a mechanism along the same lines is expected to operate in Italy starting 2017.

In this approach the capacity obligation takes the form of a call option on the generators' capacity. In exchange for a fixed fee, the supplier of generation capacity commits to pay the counterparty, in each hour *t* of the contract period:

$$Max(0, p_t^{Spot} - p^{Strike})$$

Where:

- p_t^{Spot} is the spot price of electricity in the hour, and
- p^{Strike} is the option's strike price, set as equal to the variable cost of the marginal generation unit in the system.

In other terms, the option is such that during scarcity hours, when the price rises above the variable cost of the marginal unit, the contracted generator disburses the scarcity rent for each MW of hedged capacity[38].

This creates the correct incentive for generators to make the contracted level of capacity available at times when capacity is most valuable to the system. That happens because the generator can hedge its financial position by offering its production capacity in the spot market.

---

[38] Typically for each delivery time a sequence of market venues for electricity and ancillary services takes place from day-ahead to real time. In the European markets, for a variety of reasons arbitrage across those markets is generally less than perfect. This makes selecting the reference price for the reliability options among the day-ahead market clearing price, price indexes for continuous trading intra-day markets and the real-time prices non trivial.

Consider for example the situation in an hour when the electricity price equals VoLL of a generator that has sold a 1 MW reliability option. In case the generator has not offered 1 MW capacity in the spot market, the generator suffers a loss on the option contract equal to $p_t^{Spot} - p^{Strike}$, which amounts to the scarcity rent.

However, the generator can offset this loss by offering 1 MW capacity in that hour; by doing so the generator sells 1 MWh and obtains profit $p_t^{Spot} - VC$, where *VC* is the generator's variable cost. The net profit to the generator becomes:

$$-(p_t^{Spot} - p^{Strike}) + (p_t^{Spot} - VC) = (p^{Strike} - VC)$$

Like installed capacity requirements, energy options backed by generation capacity may be procured directly by the system operator, or by placing an obligation on the load-serving entities. The options may have different duration, and may be procured more or less in advance of the commitment period. Finally, the option' strike prices may be fixed or indexed.

Reliability options perform multiple functions. First, they ensure that the target level of capacity is reached, as reliability options must be backed by physical generation capacity. In the Colombian implementation additional provisions ensure that the financial obligation placed on the supplier of capacity is indeed backed by physical generation capacity, and that the capacity is operated in such a way to ensure its availability at scarcity times. For example, contracted thermal generators must provide proof of fuel availability during the commitment period.

Second, they address any missing money issues, since under competition the option premium required by a generator reflects the difference between the total cost of supplying capacity and the expected infra-marginal rent from selling energy and ancillary services.

Third, they contribute to market power mitigation in energy and ancillary service markets by capping at the system marginal cost the net-revenues for the capacity hedged by the options. However, while it caps the hedged capacity's revenues, the mechanism does not cap the market clearing price, which in case of scarcity will rise to the level necessary to ration demand and to compensate efficiently any un-hedged producers. As a result, the incentives to develop demand response resources are not distorted as they would be for example if an overall price-cap was implemented.

Reliability options might interfere with similar commitments that generators and traders voluntarily exchange in the market. Since a generator's physical capacity hedges its position on reliability options, any further short position, such as a sale of electricity future or of a call-option, would place the generator in a speculative position. It may be argued, however, that future and option products to be exchanged in the market could be designed in such a way as to carve out the positions resulting from reliability options. Consider for example a generating unit with variable cost 60 €/MWh that entered a reliability options with strike price 300 €/MWh. Absent the reliability option, that generator's income would be fully hedged by selling a call option with strike price 60€/MWh. With the reliability option full hedge can still be achieved by selling a call option with strike price 60€/MWh and reference price capped at 300 €/MWh[39].

2. 4. Strategic reserve

This approach is based on reserving part of the installed generation capacity for use only in scarcity situations, i.e. as the reserve of last resort.

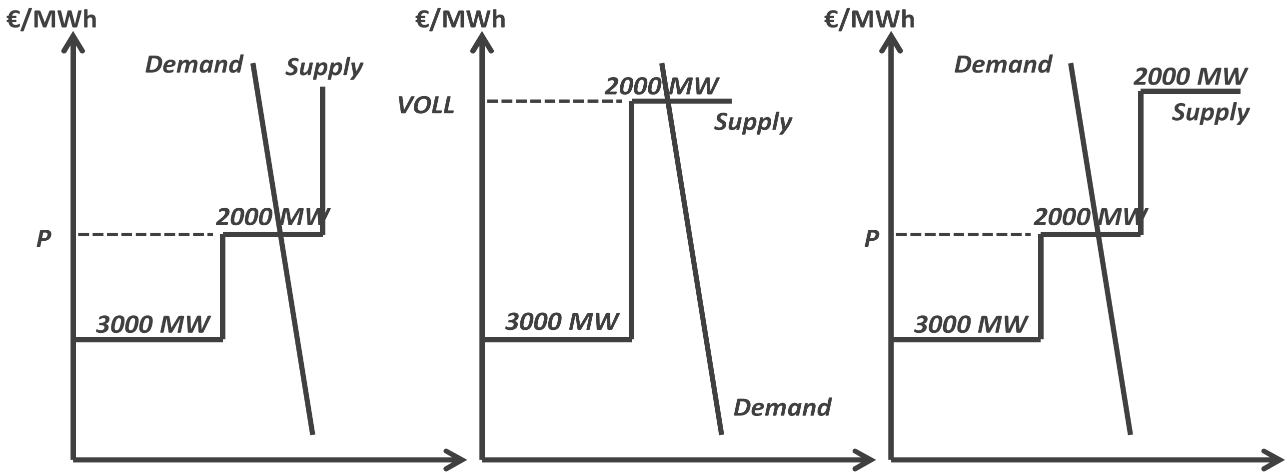
In order for the measure to bring about a permanent increase of installed generation capacity, the last resort reserve has to be effectively removed from the market. This requires that the each time the last resort reserve is activated, the market price for electricity (and operating reserve) rise to the VoLL, as in the event of scarcity. Otherwise the reserve of last resort will displace new capacity and the total installed capacity will not increase (the so-called slippery slope).

Consider, for example, the market shown in Figure 5 on the next page, where we assume that demand and installed capacity are steady. The regulator is not satisfied that the current level of installed capacity is adequate and believes that an additional capacity of 2,000 MW is necessary.

In order to induce investment in an additional capacity of 2,000 MW, the regulator therefore contracts 2,000 MW of the existing capacity as last resort reserve. The contracted capacity is then offered on the energy and ancillary service markets at a price equal to the VoLL. The result of this measure is the market supply function represented at the centre of Figure 5. Consequently, in the event that the last resort capacity is scheduled for production or to provide operating reserve the market clearing price is the same as in the event of scarcity.

[39] The latter would commit the generator to paying the counterparty, every hour, the difference $Max\left(0, Min\left(p_t^{Spot}, 300\right) - 60\right)$

Figure 5 - Reserve of last resort scheme



The new aggregate supply function is such that the market clearing prices reflect scarcity conditions more often than they would without the intervention. The profitability of the existing generation capacity then increases. This attracts investment until the total installed capacity reaches the pre-intervention level, i.e. until the capacity shifted to the last-resort reserve has been replaced. The new equilibrium is shown in the right panel of the Figure.

A specific feature of strategic reserve is neutrality on spot energy and ancillary service prices. As a consequence those markets remain the main drivers for investment in generation capacity. Other support mechanisms affect the competitive price pattern on the energy and ancillary service markets by increasing the level of generation capacity offered in the market, so that the number of very-high price scarcity hours reduces.

From that perspective, strategic reserve appears suited to situations in which the market is believed to deliver the efficient level of generation capacity. In such a context, production from out-of-the-market capacity is just used as a substitute for load curtailment, recognizing the limited and imperfect role of price signals in rationing demand. In contrast, capacity payments and capacity requirements appear to be grounded on the assessment that the level of capacity delivered by the market is fundamentally inefficient. Under that assumption, the fact that a portion of the generators' fixed cost is recovered via the capacity mechanisms – rather than through high energy prices in case of scarcity – becomes an unavoidable by-product of a measure addressing a market failure. In addition, since the market outcome is believed to be inefficient, there is no reason to enforce the same prices for energy and ancillary services that would prevail in case no capacity remuneration scheme had been implemented.

The scheme based on the reserve of last resort may cause inefficiency if the units providing reserve of last resort turn out to be more efficient than some other units. In this case, cheaper units are withdrawn from the market while more expensive generators are activated to meet load. For this reason, the reserve of last resort appears particularly attractive when the regulator has the opportunity for preventing old and inefficient units from being scrapped. The cost of keeping alive units that would otherwise be dismantled could be relatively low, and there would be little risk of technical inefficiency.

The strategic reserve approach is implemented in some Nordic countries, New Zealand[40] and Italy. The following boxes provide an overview of the mechanisms in place.

Nordic countries[41]

In Sweden and Finland the system operator procures peak-load resources in order to reduce the probability of shortage situations during the winter season. These reserves are to be used as an alternative to load shedding.

Strategic reserves can be activated either on the spot market or in the regulation power market. However given that most of the strategic reserve contracted by system operators are old and less flexible plants, they are often activated at day-ahead stage.

When strategic reserves are activated, both energy and balancing market prices are set slightly above the highest accepted offer. In this respect the mechanism departs from the benchmark discussed in the report, in which activation of strategic reserve triggers VoLL pricing.

Other Nordic countries such as Norway and Denmark also use out-of-market solutions in order to ensure security of supply. According to (Nordel 2009) part of the Danish generation is operated under special conditions and financed by a capacity payment. Moreover, the Norwegian system operator, at least in the past few years, has subscribed to bilateral agreements with production and demand resources.

[40] The Reliability-Must-Run contracts implemented in some US markets are based on the same logic.
[41] For a more detailed overview of this mechanism see: Nordel, 2009. *Peak load arrangements – Assessment of Nordel Guidelines.* Report 2/2009

New Zealand

In New Zealand strategic reserves are contracted in order to avoid costly load disconnection during dry year when hydropower production is limited. The strategic reserves are contracted through centralized public auctions and can include either new or old power plants (Battle and Rodilla 2010).

Strategic reserves are activated only if scarcity conditions arise and when this happens the price is set administratively at levels far above system marginal cost.

Italy

A capacity support scheme, in some respects similar to strategic reserve, is implemented in Italy in order to keep in service fuel oil generation capacity, to avoid scarcity events triggered by lack of natural gas – the main source for electricity production – during cold winters.

Each year the system operator selects a number of plants that during periods of gas scarcity will burn fuels other than natural gas, mostly fuel oil. These plants commit to being available for dispatch between January and March.

Since these units are activated in case of scarcity of gas supplies to Italy, the mechanism implements strategic reserve in the gas, rather than in the electricity industry, by substituting oil for gas. Consistently with this logic, when activated oil fired units are offered in the market at variable cost. A capacity payment covers the selected units' fixed cost.

Great Britain

The British electricity system is expected to face capacity adequacy issues in the near future.[42] In order to address that problem, public authorities are considering a mix of capacity remuneration

[42] Ofgem "*Electricity Capacity Assessment Report 2013*", June 2013. Available at: http://www.ofgem.gov.uk/Markets/WhlMkts/monitoring-energy-security/elec-capacity-assessment/Documents1/Electricity%20Capacity%20Assessment%20Report%202013.pdf

schemes and market design changes. Those include in particular the following measures:

- the introduction of a capacity market to ensure security of supply in the medium term (2018/19);[43]
- the procurement of additional reserve to ensure security of supply in the short term (2014/15);[44]
- a review of the balancing system rules, to provide better signals for new investments.[45]

Those measures are briefly outlined hereunder.

Capacity market

For each delivery year a capacity demand curve will be administratively set, based on security of supply analyses carried out by the system operator[46]. Then, the target capacity will be contracted through a central auction, taking place four years in advance of each delivery year. Capacity providers that are successful in the auction will commit to offer capacity in the market in exchange for a capacity payment equal to the auction's clearing price. Capacity providers failing to meet their obligations will be subject to penalties based on VoLL.

The implementation of the new capacity market is underway. According to a government's recent announcement, the first capacity auction will be held in 2014, for delivery in 2018/19.

Procurement of additional reserve

Given that the proposed capacity market will only produce its effect within four to five years, the British authorities are exploring further measures to support security of supply in the short term.

Discussion on measures to address capacity issues in the short run (2014/15) focuses on two

---

[43] DECC *"Electricity Market Reform: Capacity Market – Detailed Design Proposals"*, June 2013. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/209280/15398_TSO_Cm_8637_DECC_Electricity_Market_Reform_web_optimised.pdf

[44] Ofgem "*Consultation on the potential requirement for new balancing services by National Grid Electricity Transmission plc (NGET) to support an uncertain mid-decade electricity security of supply outlook*", June 2013, http://www.ofgem.gov.uk/Markets/WhlMkts/EffSystemOps/Documents1/Consultation%20on%20the%20potential%20requirement%20for%20new%20balancing%20services%20to%20support%20an%20uncertain%20mid.pdf

[45] See: Ofgem "*Update on the Electricity Balancing Significant Code Review (EBSCR) and request for comments on proposed new process to review future trading arrangements*", February 2013. Available at: http://www.ofgem.gov.uk/Markets/WhlMkts/CompandEff/electricity-balancing-scr/Documents1/Update%20on%20EBSCR%20and%20new%20process%20to%20review%20Future%20Trading%20Arrangements.pdf

[46] Development of this proposal is at a very initial stage.

"new balancing services" procured by the system operator. One, i.e. Demand Side Balancing Reserve, entails wider participation of demand side resources in the provision of flexibility. The other, i.e. Supplemental Balancing Reserve, seems to be a kind of *strategic reserve*, as discussed in section 2.4. of this report.

Balancing system review

In parallel with the development of capacity remuneration measures, the British authorities are also reviewing balancing rules in order to improve short term price signals. The proposed changes mainly focus on the imbalance prices (or cash out prices). Proposals

under discussion include:

- the calculation of the imbalance price based on a smaller number of the most expensive offers accepted by the system operator;
- the move from a dual to a single imbalance price mechanism;
- a different allocation of the costs sustained by the system operator in order to procure operating reserve and other services.

3. Additional issues

In section 3.1 we deal with issues related to coordination of capacity remuneration schemes across neighbouring countries. In section 3.2 we discuss how capacity remuneration schemes can contribute to ensuring that enough flexible capacity is available in the system.

3. 1. Coordinating capacity remuneration mechanisms across multiple countries

In this section we address cross-border issues related to capacity remuneration measures. In section 3.1.1 we show how the benefits of capacity supplied by generators participating in a country's remuneration scheme are appropriated by that country's consumers in an interconnected system and illustrate how the existence of capacity remuneration schemes impact on the power flows and surplus distribution between neighbouring countries in case of scarcity. In section 3.1.2 we discuss the broader impact of the introduction of a capacity remuneration scheme in one country on a neighbouring country's wholesale power market. Section 3.1.3 wraps up our results.

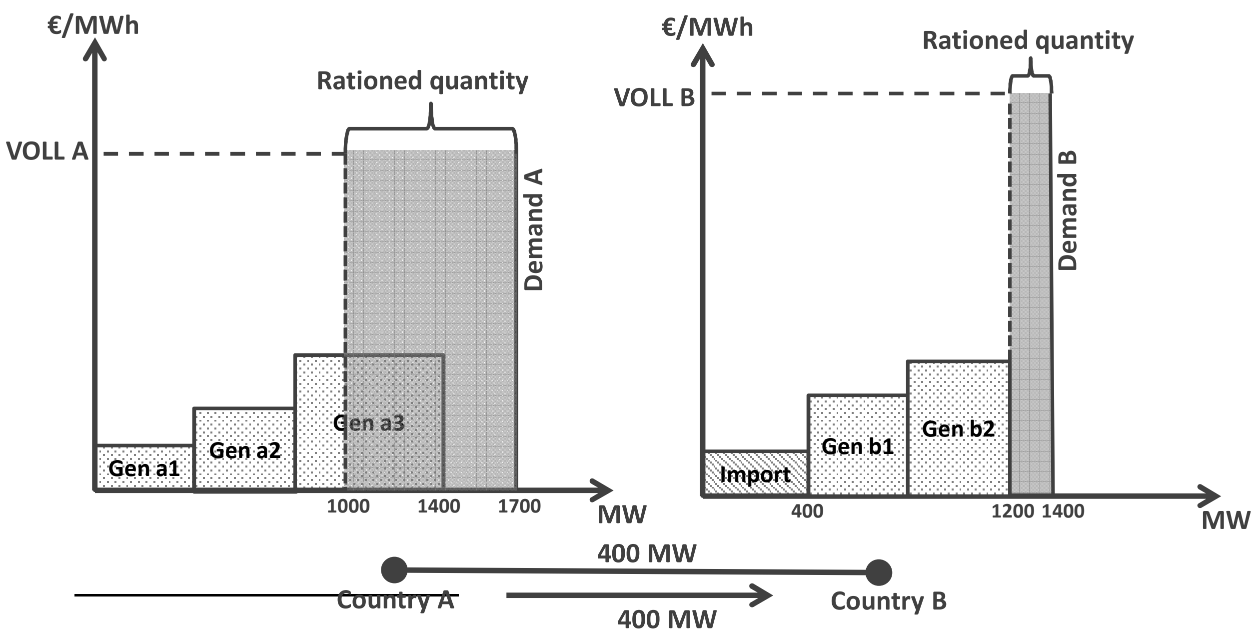
3.1.1 Enforcement of capacity rights in an interconnected system

As shown in Chapter 1, the value of generation capacity, i.e. the benefit brought to the system by an additional unit of capacity, is zero if the available generation capacity is not fully utilized and equal to the VoLL[47] in the event of scarcity. Therefore, assessing how the benefits of capacity are reaped by the consumers who pay for it in an interconnected system requires characterizing the market outcome in case of scarcity.

Consider a benchmark setting with two interconnected energy only markets, country A and country B[48], each with perfectly price-inelastic demand in the relevant range. The energy flow between country A and country B, and therefore the allocation of brown-outs between the two countries in case of scarcity, depends on the VoLLs enforced in the two markets.

In this example, the installed generation capacity in country A and B are respectively 1,400 MW and 800 MW. Demand in country A is 1,700 MW and in country B is 1,400 MWh, transmission capacity equals 400 MW in both directions. Figure 6 illustrates the market outcome in case VOLL-A<VOLL-B[49]. In case of scarcity the interconnection capacity between the countries is fully used to export from A to B. As a result, most of the disconnections (700 MW) take place in country A; only 200 MW are disconnected in country B.

Figure 6 - Power flows between interconnected countries with scarcity and different VoLLs.



[47] Net of the variable cost of the most expensive unit.
[48] For simplicity and without loss of generality we ignore ancillary services.
[49] If VoLL-A=VoLL-B

Coordination among the system operators is necessary to implement this market outcome. Each system operators schedules, and if necessary implement, load curtailment in his control area in a way such that the cross-border power flows stay at the planned level.

We consider next the effect of alternative capacity remuneration schemes implemented (only) in country B. We start with country B introducing an overall price-cap equal the system marginal cost together with a capacity payment making up for the missing money, the fixed cost of the marginal unit because of the price-cap.

In this setting, in case of scarcity the clearing price in A is unconstrained and can rise to VoLL-A, while the clearing price in B is capped at system marginal cost. That means that the market outcome is such that interconnection capacity is fully used for exports from B to A; in other terms country B exports its generation capacity to country A and suffers most of the brown-outs. This is clearly not the intended outcome of the introduction of a capacity remuneration scheme in country B, where consumers value electricity more than consumers in country A, as reflected in the VoLLs. One could interpret this situation as one in which consumers in country B pay for capacity that ends up benefiting consumers in country A. One could interpret this situation as instance of *free-riding* by country A's consumers.

This example shows that with an overall price-cap, in order for the consumers in country B to reap the benefits of capacity they pay for, system operators must coordinate, in case of scarcity, to enforce cross border power flows inconsistent with the relative price for electricity in the two countries. In particular, in case of scarcity they must prevent exports from B to A. By controlling cross-border flows in case of scarcity the system operators enforce country B consumers' rights on the generation capacity that benefits from the remuneration scheme they pay for. No *free-riding* by country A consumers takes place.

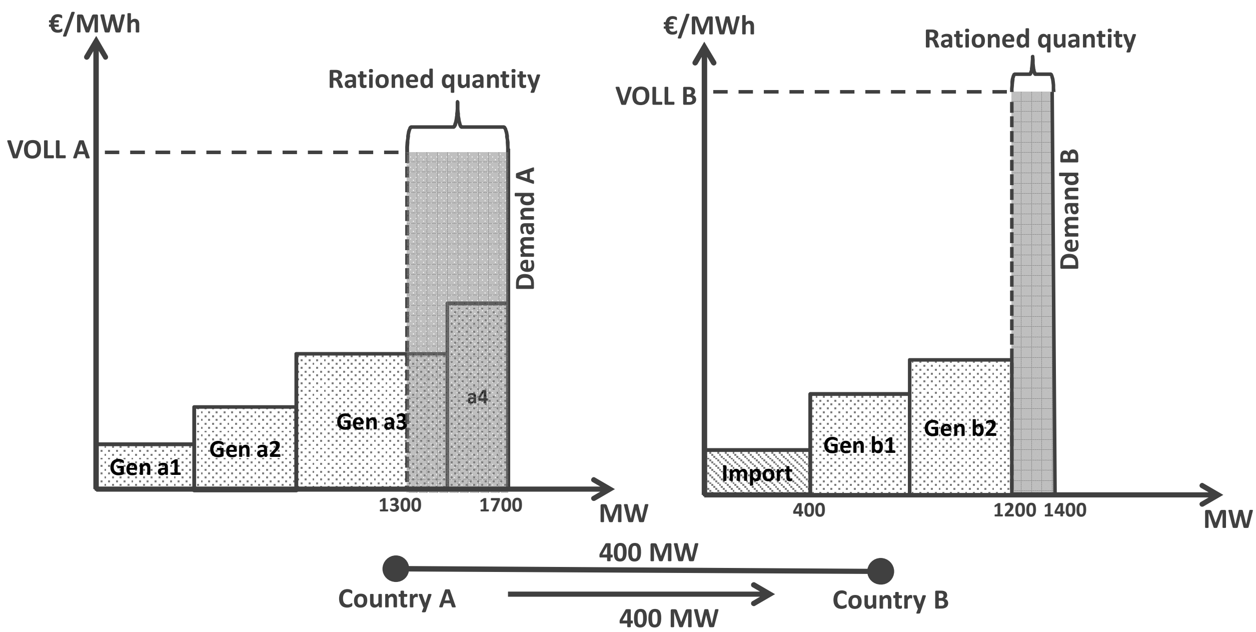
It should be noted, however, that the issue we have outlined is caused by a set of market power mitigation and capacity remuneration measures that, in case of scarcity, distort the energy market price in country B. In fact, this feature is not common to all capacity remuneration schemes. Assume for example that country B implements a capacity remuneration scheme based on financial obligations, like the reliability option mechanism presented in section 2.3.3. With this scheme, in case of scarcity power prices rise to VOLL in both countries. This capacity remuneration scheme, therefore, does not impact on the distribution of curtailment between the two countries, which is driven by the relative VoLLs.

Finally, consider a setting in which country B does not implement price-cap and capacity remuneration measures, while public authorities in country A assess that additional generation capacity is desirable. Assume that the capacity remuneration mechanism in country A causes a 300 MW increase of generation capacity in the country. Since VoLL-A < VoLL-B, in case of scarcity country A exports to country B up to full utilization of the interconnection capacity. This outcome is illustrated in Figure 7.

Interpreting this situation is not straightforward. On the one hand, consumers in country A show lower valuation of generation capacity than consumers in country B, since VoLL-A is smaller than VoLL-B. Provided the VoLLs accurately reflect the value of electricity in the two countries, it is efficient that load curtailment take place as much as possible in country A. On the other hand, consumers in country A, through the capacity remuneration system, paid for a level of capacity availability that would make load curtailment unnecessary if no exports occurred; those consumers are likely to expect to reap the benefits of capacity remuneration in the form of greater service reliability.

One may argue that a system of side-payments – from consumers not curtailed in country B to consumers curtailed in country A might make consumers in country A better off by selling capacity to consumers in country B rather than using it to prevent load shedding in their own country. In fact, such payment system might be difficult to set-up in practice, if only because of the political impact of curtailment events.

Figure 7 - Power flows between interconnected countries with scarcity and different VoLLs



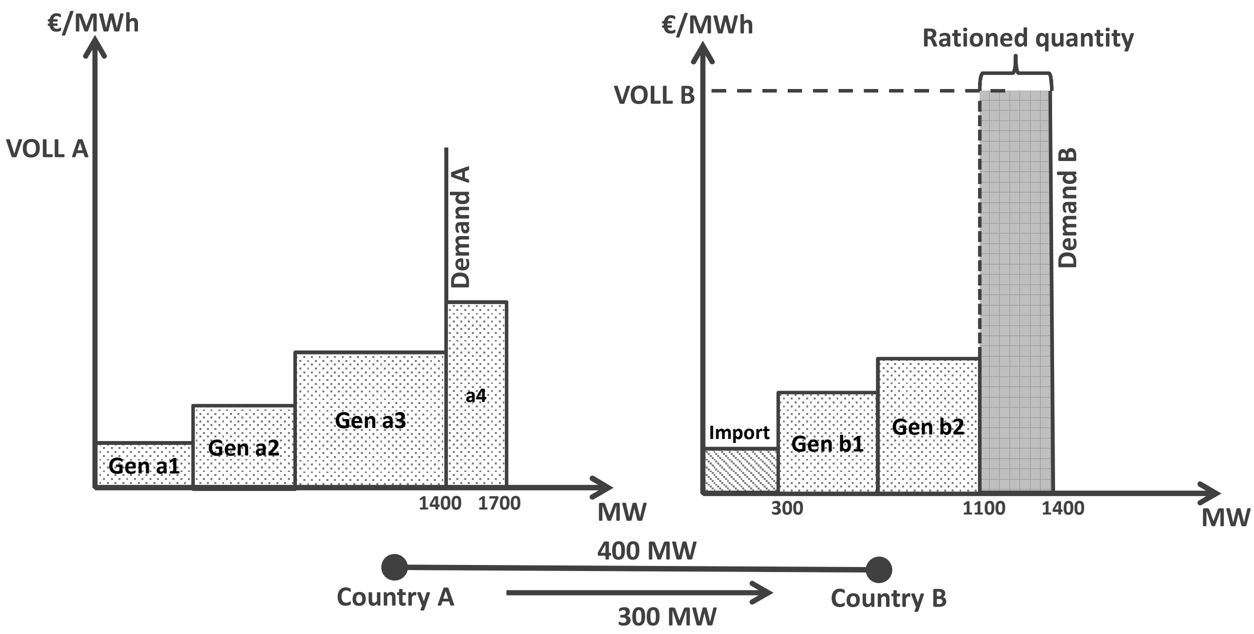
Note that – even if used with priority to meet load in country A - additional capacity made available by the capacity remuneration scheme introduced in country A benefits country B for all demand realizations such that curtailment in B is smaller than it would be otherwise (after load in A is served). However:

- If congestion occurs between A and B, consumers in country A, who bear the cost of the capacity remuneration scheme, appropriate (part of) the value created by that capacity in country B through congestion rent, provided electricity is correctly priced in both countries. This is illustrated in Figure 8, where we assume that demand in country A is 300 MW less than in the previous example. We also assume that 50% of the congestion rent is allocated entirely to consumers in country A, typically as a negative component of the transmission tariff[50].
- If no congestion between A and B occurred, i.e. if interconnection capacity in our example were larger, the same energy price would result in both countries. Generators in country A would then profit from additional demand by country B; additional profits would then reduce the need for revenues to be provided by the capacity remuneration scheme.

In either case consumers in B contribute to remunerate country A's generation capacity, by increasing its utilization at times and to the extent that it is not used to match country A's loads. Interpreting this situation as an instance of *free-riding* is not obvious, since both country A and country B benefit and to some extent pay for the additional capacity built under the remuneration scheme.

[50] Note that a rent equal VoLL- B - VoLL-A could result even if the A to B interconnection is not fully utilized, in case the excess capacity in Country A is smaller than the interconnection capacity. In this situation the obvious allocation of the rent would be entirely to consumers in country A.

Figure 8 - Power flows between interconnected countries with scarcity



Cross-border capacity transactions

Implementation of coordinated pan-European capacity procurement may take place under the same arrangements as cross-border electricity trading. A first option is coupling multiple national capacity markets, as energy markets are for example in central Western Europe. Demand[51] and supply for capacity in all systems would then be simultaneously cleared, subject to transmission security constraints. As for electricity trading, with market coupling there is no direct commitment between buyers and sellers of capacity. Consider for example a market outcome in which country A imports 5000 MW capacity from country B. Generators selling capacity in B do not commit directly to buyers of capacity in country A – for example load serving entities – to deliver capacity. On the contrary, each seller commits to the system operator of the network where it connects, while the commitment to maintain 5000 MWs flow from country B to country A in scarcity situations is between the two countries' system operators. In practice implementation of cross-border capacity arrangements requires a complex set of institutional arrangements, similar to those for cross border energy trading. These include in particular

[51] Expressed either by the national system operators or, in a decentralised system, by parties bearing the capacity obligation, typically the load serving entities.

agreements necessary to coordinate assessment by the system operators of cross-border flows which would occur in case of scarcity.

Second, cross-border exchanges of capacity commitments may be bilateral. In this model a generator in country B supplying 100 MW capacity to, for example, a load serving entity in country A commits to deliver 100 MW in scarcity situations in A's system. Execution of this agreement requires that, during each scarcity hour in country A:

- the generator produces 100 MWh and
- contry A and country B's system operators implement a 100 MWh flow to country A, even if that requires curtailing load in country B.

Note that under both arrangements for cross-border capacity trading:

- execution of cross-border capacity transactions requires coordination among system operators
- capacity commitments have no physical content in non-scarcity conditions, when the demand for energy and ancillary services in country A can be met at some price by existing capacity. This happens because when available capacity exceeds demand for energy and ancillary services, there is no competition between country A and country B's consumers for capacity[52] or, in other words, capacity has no value.

The same implementation issues for bilateral cross-border energy trading need to be addressed in order to support cross-border bilateral capacity commitments.

Finally, a subtle implementation issue arises when capacity remuneration interacts with cross-border bilateral energy trading[53]. Consider again the two-country system illustrated in Figure 5 and assume:

- VOLL A = VOLL B so that energy prices do not drive cross-border flows in case of scarcity in the both countries, like in our example;

[52] There is indeed competition for energy and ancillary services, resulting in positive clearing prices in the corresponding markets.
[53] The author is in debt to Wilhelm Suessenbacher at E-Control for pointing out this case.

- a baseload 400 MW bilateral contract has been signed by a generator in A and a buyer in B and corresponding “physical” transmission rights have been purchased;
- The generator is also enrolled in the capacity remuneration scheme implemented in country A.

According to the usual design of bilateral energy contracts, system operators should maintain a constant 400 MW flow from country A to country B matched by the generator’s injection in A and the buyer’s withdrawal in B[54].

However, the generator’s capacity commitment to country A’s system operator establishes that capacity must be available to match country A’s load (or operating reserve requirements). Therefore, in the event of scarcity, the flow to country B should not take place.

Our characterisation of the bilateral contract makes it incompatible with the capacity commitment assumed by the generator. A possible interpretation is that by entering the two transactions the generator sells the same thing – 400 MW delivered at times of scarcity – to two different buyers, country A’s system operator and the counterparty to the bilateral contract.

However, exports via bilateral transactions can be made compatible with the sale of capacity to the country where the generator is located by modifying the seller’s obligation in the event of scarcity in country A. In case of scarcity in country A, then:

- no power flow from A to B would B implemented, consistently with country A system operator’s entitlement on the generator’s capacity;
- the generator’s would be dispatched to match load in A and compensated according the (scarcity) pricing rule in country A;
- the buyer in the bilateral contract would receive from the generator financial compensation 400 MW*VOLL B for each scarcity hour in country A;

[54] In fact the cross border flow would have to be implemented irrespective of the generator’s injections and of the buyer’s withdrawals.

- the buyer in the bilateral contract would face disconnection, or alternatively pay VOLL B for power withdrawn during scarcity hours, according to rules implemented in his country.

Note that the outcome of this modified bilateral contract would be the same as if the cross-border energy transaction were implemented through a combination of: i) simultaneous buy and sell deals in coupled national spot markets and ii) a derivative contract[55] between buyer in B and generator in A.

Our example highlights that enforcement of a country's rights on generation capacity located abroad may require cross-border energy flows to depart, in case of scarcity, from those corresponding to energy transactions. This may appear in contrast with art. 4 (3) and art. 4 (4) of Directive 2005/89/EC on security of electricity supply, stating a non-discrimination obligation between cross-border contracts and national contracts. In fact our example suggests that, provided cross border energy transactions are properly characterized, no discrimination take takes place. In particular a generator exporting electricity is committed (and commits the involved system operators) to deliver electricity to the foreign buyer in all but scarcity situations. On the contrary, a generator exporting capacity commits to delivering energy only in scarcity situations. Energy and the capacity contracts set different and non-overlapping obligations. Therefore, no discrimination takes place if cross border flows corresponding to energy contracts, but not backed by capacity commitments, are curtailed in case of scarcity.

3.1.2 Capacity remuneration mechanisms and cross-border competition

One argument, sometimes surfacing in the policy debate is that capacity remuneration schemes, by increasing the generators' income, induces more aggressive offer behaviour in the energy and ancillary service markets, resulting in lower prices in the country where remuneration is implemented. In fact, as shown in the previous chapter, capacity support mechanisms can be designed in such a way that they do not induce an inefficient offer strategy in the energy and ancillary service markets[56], provided revenues from the capacity remuneration scheme do not depend on the revenues obtained by generators in the energy and ancillary service markets.

However, capacity remuneration measures may impact on energy and ancillary service prices,

[55] In the industry jargon a two-way "contract for differences".

[56] In fact, a mechanism based on reliability options (see section 2.3) may bring the generators' profit-maximizing offer strategy closer to the competitive one, by reducing incentives to exercise market power.

and therefore possibly on cross-border flows, by modifying the level and composition of installed capacity in the country where a remuneration scheme is implemented.

Ceteris paribus, a capacity remuneration measure makes generation investment in the country where it is implemented more attractive, because of the income-increasing and risk-reducing features of the scheme. However, major factors mitigate the distortion in the generators' siting decisions caused by unilateral or asymmetric capacity support. First, if cross-border interconnection is limited, the expected return on the generation investment depends on the country where the plant is located. There is not much point building excess generation capacity in a country implementing a capacity remuneration scheme if, because of congestion, selling energy and ancillary services to the neighbouring country (which does not implement a capacity remuneration scheme) is impossible[57].

Second, remuneration schemes pursuing a quantitative target typically set their capacity target based on the energy and reserve requirement originating from the consumers connected to the country's network, who will bear the cost of capacity remuneration. As a consequence, capacity in excess of the country's needs does not benefit from the remuneration scheme[58].

A more general (and political) assessment of the issues related to cross-country competition would lead to questioning the merit of a setting in which some countries let the market mechanism drive the development of their generation fleet while other interconnected countries put the system operator and/or the regulator in charge of deciding on generation capacity investment and transfer part of the commercial risk that would otherwise fall on the investors onto the consumers, through the capacity remuneration scheme. However such an assessment of generation capacity remuneration schemes is beyond the scope of this paper.

Russia-Finland and Ireland-UK cross-border issues

Distortions of short term cross-border trade have been identified by ACER across the Ireland-UK and Russia-Finland borders[59]. In both cases a distortion would be caused by the structure of the fee raised to cover the capacity remuneration mechanism's cost. In particular, a per-MWh fee

[57] A related, even though possibly temporary, argument is that imperfect coordination of reserve and balancing markets appears to limit cross-border trading these products.

[58] It may be argued that also a capacity payment scheme attracting capacity massively in excess of the payers' needs would sooner or later be withdrawn and would therefore be difficult to credibly commit to by the regulator.

[59] ACER: Capacity remuneration mechanisms and the internal market for electricity, 30 July 2013; http://www.acer.europa.eu/Official_documents/Acts_of_the_Agency/Publication/CRMs%20and%20the%20IEM%20Report%20130730.pdf

adds to the variable cost of the exporting generators, which may result in less than efficient exports by the country where the capacity remuneration scheme is implemented.

Notice that this distortion is not specific to capacity remuneration schemes. Distortions with the same nature result, for example, if different per-MWh transmission charges are implemented in the importing and exporting countries.

3.1.3 Discussion

Our analysis shows that, even in the context of liberalized and integrated wholesale power markets, different countries may still pursue different reliability (or capacity) targets. Different countries may also express different values for consumption of electricity by consumers whose demand is not price-dependent (or VoLLs).

However, enforcing each system operator's rights over generation capacity requires coordination in an interconnected system. In particular, in the event of scarcity, cross border flows consistent with each system operator's rights on generation capacity must be implemented. These flows, which ultimately determine the allocation of black-outs among interconnected systems, may even be different from those corresponding to the clearing prices of energy and ancillary service markets. Well-designed capacity remuneration schemes do not distort the profit maximizing competitive generator offer strategy in the energy and ancillary service markets.

A longer-run effect of asymmetric or unilateral capacity remuneration schemes on cross-border competition results from the measure's impact on the country's generation fleet. The scope for this effect depends on the level of interconnection, and more generally on the degree of integration of the neighbouring markets, as well as on the design of the capacity remuneration measure.

3. 2. Capacity remuneration mechanisms and flexibility of the generation fleet

The increasing share of renewable generation capacity, in particular solar and wind, is creating concerns that the market may not provide enough flexible generation resources to offset intermittency of renewable sources.

In a frictionless setting, the usual market mechanism addresses not only the level of installed capacity, as we discussed in section 1.1, but also its composition. In particular, the increasing need for flexibility results in greater trading close to real-time and larger demand for balancing services by the system operator.

Since more flexible generators only, capable of varying output quickly, can engage in extensive trading close to real time and supply balancing services at short notice, the price-increase caused by greater demand for dynamic services primarily benefits flexible generators. As a result the relative profitability of more flexible units increases, which attracts investment in flexible resources.

Issues that may justify specific remuneration for flexible capacity are no different from those discussed above with reference to the achievement of an overall target for installed capacity. Remuneration measures specifically targeting flexible generation capacity would be motivated by the assessment that flaws in short term energy and/or ancillary service markets prevent prices from getting very high when flexible capacity is scarce.

The discussion on the merit of capacity remuneration schemes developed in Chapter 2 carries over, *mutatis mutandis*, to remuneration measures selectively targeting flexible generation capacity. In the event support to flexible generation capacity is deemed necessary, the mechanisms analysed in Chapter 2 can be adapted to impact only on the more flexible subset of the generation fleet. In particular, capacity payments may be differentiated according to the dynamic performances of the capacity supplied and specific volume requirements may be set for flexible capacity. Implementing a measure to remunerate flexibility designed around the strategic reserve concept is perhaps less straightforward, as more flexible capacity is often provided by relatively new units or hydro-electric generators, which are often likely to be infra-marginal. As a consequence, implementing strategic flexible reserve might involve keeping capacity with variable cost lower than spot-market clearing prices regularly out-of-the market.

3. 3. Capacity remuneration mechanisms and demand side response

Flexible loads, capable and willing to give up consumption in case of high prices, may reduce the system capacity requirement.

The straightforward way to exploit load flexibility can be characterized in these terms:

- flexible load is exempted from any capacity procurement obligation. In case a centralized capacity remuneration system is implemented, where the system operator procures capacity on behalf of consumers, flexible load be exempted from capacity charges;
- the system operator does not include flexible load in the assessment of the system's capacity requirement;
- flexible load is required to make consumption price-dependent by bidding in the energy and/or ancillary service markets in a way such that load will (voluntarily) not be dispatched for some energy price[60].

From this perspective, flexible load does not sell capacity. It just does not contribute to the system's capacity requirement. That happens because flexible load does not "use" capacity when its value, embedded in the energy and ancillary service prices, rises above the consumer's availability to pay.

Regulatory or market imperfections may justify paying consumers for making their load flexible, i.e. for reducing their contribution to the system's capacity requirement. Such imperfections may relate to the structure of energy prices and of the charges levied to consumers in order to cover balancing costs and/or the cost of generation capacity remuneration. Those charges may be averaged, across time or consumers, in a way such that the savings to a consumer from not consuming at a certain time do not match – and in particular are smaller than – the corresponding cost avoided to the system. In that situation, flexible load may be exposed to the correct price signals by allowing consumers to participate in the market as suppliers of capacity[61].

Other imperfections may relate to the inability of consumers to assess the future value of flexibility. We illustrate this possibility through a simple example. Consider a power system featuring:

- 100 MW inflexible demand in year 2013
- expectation of additional 10 MW inflexible demand from 2014 onwards

[60] The content of the bids and offers that flexible load would be required to submit depends on the market design.
[61] In order to avoid double payments, in this scenario flexible loads would be charged the (averaged) capacity-related charges like inflexible loads and allowed to collect the marginal value of capacity by becoming suppliers of capacity.

- Marginal cost for capacity, for a requirement corresponding to 100 MW inflexible load: 20.000 €/MW_year
- Marginal cost for capacity, for a requirement corresponding to 105 MW inflexible load: 35.000 €/MW_year
- Marginal cost for capacity, for a requirement corresponding to 110 MW inflexible load: 45.000 €/MW_year
- A simple capacity payment is implemented and paid for through a charge on inflexible loads equal to the marginal cost of capacity
- For simplicity ther is no discounting of future revenues or savings.

Consider now an existing consumer that:

- could make 5 MW load permanently flexible at a one-off cost of 350.000 €.
- does not know or does not believe that the charge for capacity will rise to 45.000 €/MW_year in case its investment in flexibility is not carried out and to 35.000 €/MW_year if it is.

The following table summarizes our assumptions over the 2013-2016 time horizon.

Year	**2013**	**2014**	**2015**	**2016**	**…**
Inflexible load without the investment in flexibility (MW)	100	110	110	110	
Marginal cost for capacity without investment in flexibility (€/MW_year)	20K	45K	45K	45K	
Inflexible load with investment in flexibility (MW)	100	105	105	105	
Marginal capacity cost with investment in flexibility €/MW_year	20K	35K	35K	35K	

Note that the 350.000 € investment would be profitable for the consumer, as it would allow 5*35.000*3 = 525.000 € savings over the 2014-2016 period, since the consumer would not pay 35.000€, the per MW capacity payment, on 5 MW of flexible capacity for 3 years

The investment in flexibility would be rational also in system-wide perspective, as it would avoid the cost of building 5 MW of additional generation capacity, at cost 45.000 €/MW_year or 675.000 € over the three year period.

However, because of either lack of information or risk aversion the consumer prefers to keep paying capacity charges, rather than investing in flexibility.

In that scenario, some mechanism allowing load to commit to be flexible over the 2014-2016 timeframe, in exchange for a payment, could deliver the efficient outcome. If, for example, the consumer trusts that the capacity charge will never fall below the price observed in 2013 (20.000 €/MW_year), he would be keen on investing in flexibility in exchange for a one-off 50.000 € payment, computed as the difference between the 350.000 € investment cost and the consumer's expected savings in capacity charges (20.000*5*3).

Such a mechanism would operate according to broadly the same logic as a capacity remuneration scheme. It remains to be seen whether (and why) the market could not be relied upon to achieve the same outcome. For example the supplier of electricity to that customer would probably be interested in offering the following deal:

- the supplier pays the 350.000 € one-off cost of making the load flexible;
- the consumer commits to paying the supplier 15.000 €/MW_year for 3 years on the 5 MW load that was made flexible.

This agreement leaves the consumer better off, since he does not bear the cost of becoming flexible and saves 5.000 €/MW_year on his (prudently) estimated capacity charge for the following 3 years. The supplier is also better-off, since flexible loads are not subject to capacity charge. Therefore, in exchange for investing 350.000 € and foregoing 150.000 revenues from the consumer in 3 years[62] , the supplier will save 35000*5*3 = 525.000€ over the three years, with a net benefit of 25.000 €.

[62] The difference between the capacity fee 35.000 €/MW_year and the 15.000 €/MW that the consumer keeps paying

Appendix: The EC Consultation paper on generation adequacy, capacity mechanisms and the internal market in electricity, 15 November 2012

Our analysis provides insights into the issues raised by the EC Consultation paper on generation adequacy, capacity mechanisms and the internal market in electricity of 15 November 2012 ("the Consultation paper"). In this Appendix we relate the results of our analysis which are relevant to the issues raised in the Consultation paper.

A first set of questions in the Consultation paper concerns the assessment of the need for generation capacity remuneration schemes in Europe. These include Question 1, Questions 7-10 and Question 12.

An empirical assessment of the current and future demand/supply balance is beyond the scope of our project. However, our analysis highlights that assessing whether market dynamics are such that adequate investment in generation capacity is attracted and therefore whether a capacity remuneration mechanism is necessary is, in all but very extreme situations of persistent scarcity, very hard. We argue, first, that evidence of, respectively, scarce or excess capacity at a certain point in time does not by itself prove that a capacity remuneration mechanism is or is not necessary. Building generation capacity takes a long time and major demand as well as supply shocks can and do happen. Therefore prolonged situations of excess or scarce supply are not exceptional in the power industry.

Second, assessing the opportunity for a capacity remuneration scheme requires figuring out the supply conditions that would be produced by the measure. Such counterfactual scenario is likely to be hard to identify and highly uncertain.

Finally, in most European countries central elements of the electricity wholesale and ancillary service market design are still being developed, also within the context of the Target Model implementation. As a consequence the need for a capacity remuneration mechanism must be assessed with reference to a somewhat untested market design.

Question 2 in the Consultation paper investigates whether subsidies to specific energy sources, such as renewables, coal and nuclear may undermine investments needed to ensure generation adequacy.

Our analysis suggests that politicization of the development of (part of) the generation fleet creates a political risk that increases the rate of return required by the market to invest in generation capacity and may lead to delaying investment until persistent scarcity events reassure investors on the profitability of additional capacity. Public authorities should aim at minimizing that risk, by making the evolution of the subsidized share of generation capacity predictable with certainty by the investors in non-subsidized capacity. More generally we argue that credibility of all facets of public intervention in the industry is crucial to attract investment in power generation capacity.

Questions 3-5 and question 11 investigate the impact of greater integration of the European national power markets on generation adequacy. Although a quantitative assessment, and ultimately a judgment, on the impact of full implementation of the Target model is beyond the scope of our work, our analysis makes it clear that capacity remuneration mechanisms should not be regarded as a substitute for a sound market design and a predictable regulatory environment. Policy attention should therefore focus on eliminating any market design flaws or unnecessary regulatory uncertainty, independently of capacity support measures. In particular our analysis emphasizes the need for pan-European coordination of the Value of lost load, the administrative price for electricity in case of demand rationing and of operations of the national power systems in the event of scarcity.

Question 6 asks how public authorities should address potentially different consumer preferences in relation to security of supply. Our analysis highlights that given the available technology, selective disconnection in case of scarcity of (all but large) consumers based on their willingness to pay is impossible. All (the many) consumers connected to the same network branch must be disconnected together.

In principle a different VoLL may be assessed, and used to prioritize disconnection, for each group of consumers connected to the same network branch. However, extracting from consumers the information on their valuation for reliability is a very difficult exercise. In addition, aggregating the different consumers' valuations into a single value (for the network branch) would be an arbitrary exercise, as it is probable that there is no consistent valuation even by individuals. Finally, a policy of prioritizing disconnections based on differentiated VoLL might not be politically acceptable.

Our analysis shows, however, that even in a market setting different countries may assess different VoLLs, provided system operators coordinate to manage cross-border flows consistently, in case of scarcity.

The large scale deployment of smart-metering systems, expected to take place in Europe by 2020, will enable most consumers to plan and adjust power consumption in response to spot and real-time prices. It remains to be seen whether the net benefits of such behaviour is positive for a such a large share of the consumers that a significant portion of the market's demand will become flexible, removing one of the main reasons for capacity adequacy concerns.

Questions 13 and 17 investigate the role of capacity support mechanisms in attracting investment in flexible generation capacity. Our analysis suggests that the reasons for concern on adequacy of flexible capacity have conceptually the same nature as those relevant for overall capacity adequacy. In case supporting flexible generation capacity is deemed necessary, some mechanisms designed to address overall capacity adequacy can be adapted to impact on the flexible subset of the generation fleet.

Questions 14-16 address the features and the relative merits of alternative capacity support mechanism. On this matter we refer the reader to Chapter 2 of the report.