May 2024

Reforming Capacity Markets: How to Incorporate the Flexibility of Residential Consumers?

OIES Paper: EL55

Dimitra Apostolopoulou, Research Fellow, OIES
Rahmatallah Poudineh, Head of Electricity Research Programme, OIES

ISBN 978-1-78467-245-4

Abstract

Capacity markets as a mechanism of capacity remuneration, have been widely applied in the United States, Australia, Europe, and United Kingdom. In these markets, generators and other resources make offers based on their unavoidable costs to keep them available, and the transmission system operator determines the demand based on forecasted peak load. However, capacity markets have limitations, mainly because they are one-sided since the system operator is the only buyer, with the government deciding on the required reliability levels, and not consumers who are the ones actually experiencing the risk of outage. Moreover, capacity markets have been tailored for conventional resources whereas renewables, storage, and demand response resources have different characteristics. Demand response is specifically very important because it constitutes a potential alternative to building new generation capacity. At times of generation scarcity there are consumers who might be willing to reduce their demand in return for financial compensation.

This paper analyses the capacity market and proposes an approach to incorporate the flexibility of residential consumers in this market. We introduce and assess the potential of non-linear pricing schemes, specifically priority pricing contracts, as mechanisms to enhance the implicit residential demand response. We suggest a model to integrate priority pricing contracts via an aggregator as explicit demand response in capacity markets, advocating for participation on the 'demand' side rather than the 'supply' side to prevent market distortion and better align with consumer reliability preferences. To incorporate these contracts into the capacity demand curve, we examine the correlation between capacity and reliability, establishing that under ideal conditions, the marginal cost of outages aligns with the marginal cost of capacity, thus linking capacity to the value of lost load and loss of load expectations. This connection informs the design of the capacity demand curve using data from priority pricing contracts. We demonstrated this approach using the 2027-28 UK capacity market, and also introduced a refined capacity product for trading in capacity markets, designed to encourage investment in demand response resources and incorporating specifics about location, flexibility, and adjustments to the calculation of firm capacity profile. Lastly, we propose a potential business model for the aggregator that leverages residential demand response that does not create distortions in the retail market.

Contents

Figures

Tables

1. Introduction

The decarbonization of power systems has led to the integration of a high number of renewable resources – thus changing significantly the generation portfolio in many countries. For example, in the United Kingdom (UK) solar photovoltaic (PV) capacity totals 15 GW, and generated a record of 5.5 TWh of energy between April to June 2023 (DESNZ 2023). However, renewable resources are highly intermittent and variable, and lead to net load variations of increased uncertainty. To maintain system reliability under peak load as well as in contingency cases like unexpected outages, there is a need for sufficient energy, sufficient capacity, and flexibility. To further explain the three points above, electric devices consume energy that is supplied to them by the electricity system, and is measured in Watt-hour (Wh). The electricity system needs the required potential to meet demand under peak conditions; this is what we call capacity that is measured in Watts (W). A newer concept in power system operations is flexibility or 'ramp rate' which is the capacity amount that may be provided in a given timeframe, measured in Watt per hour (W/h). The latter has attracted more attention due to operational issues when renewable generation increases and decreases its output in a short time. An example of such a situation is the 'duck curve' phenomenon in California (CAISO 2016). When the Sun sets, the contribution of PV resources drops suddenly and there is an increased need for the quick ramp-up of energy production, or ramp down of demand. Many researchers have studied the adequacy of energy, capacity, and flexibility, collectively referred to as 'resource adequacy' in the context of increasingly frequent and severe extreme weather events and the ongoing transformation of power systems around the world.

Resource adequacy has become more challenging in the 'energy transition' because of market failures that do not promote adequate investments. Electricity markets are currently designed to dispatch generation based on economic merit order. In this regard, under current market designs zero carbon resources lead to low future electricity prices, as these will be cleared first. Moreover, units with high marginal costs, like combined cycle gas turbine (CCGT) units, will most likely not be cleared (or cleared very infrequently) in the energy market (Papavasiliou 2020). Furthermore, regulators impose price caps in energy markets to address potential market power issues and this reduces payments for all types of generation, the so-called 'missing money' problem (Neuhoff & De Vries 2004). As a result, generators are faced with insufficient market revenues to support adequate new investments, which in turn affects system resource adequacy. The refinement of scarcity pricing in the energy-only markets alleviates this situation since it uplifts balancing prices when the system is tight, and thus rewards flexible resources for providing much-needed energy to the system when it is under stress (Hogan 2013). Another solution to the missing money problem is the introduction of capacity remuneration mechanisms (CRMs). These mechanisms make payments to installed capacity based upon various factors, including location, availability during peaks, and the total capacity relative to the need, based on reliability criteria. CRMs compensate resources for making generation capacity available for utilization, regardless of the extent to which it is actually operated. This provides an additional revenue stream to generators and incentivizes investment in additional generation capacity, thereby ensuring resource adequacy. There are several types of CRMs; they can be technology neutral or not, centralized or decentralized, volume- or price-based. The commonality of all CRM designs is that they reduce the risks for new investments by offering resource providers supplementary income on top of their earnings from selling electricity in the market, thus ensuring there are no system adequacy concerns at times of stress. In the US, the earliest such mechanisms date back to the late 1990s. In recent years, several European countries have also started implementing different kinds of CRMs (Bublitz et al. 2019). In this paper we focus on capacity markets as a form of CRM, with such mechanisms widely used in many jurisdictions like the US, Australia, Europe, and the UK.

However, capacity markets have several shortcomings, mainly that a central agency makes decisions on behalf of consumers relating to the reliability needs and capacity margins of the system, whereas consumers who are actually experiencing outage risks cannot affect this decision, since reliability is considered to be a public good. Moreover, historically, capacity markets have typically targeted more conventional resources and have received criticism that even though they are supposed to be technology-neutral they fail to do so due to their regulatory framework. For example, participation rules do not favour the participation of distributed energy resources, such as electricity storage or demand response (DR) and in fact provide hidden subsidies to operators of conventional power plants. For example, in the UK, the four year ahead capacity auction with delivery year of 2026-27 cleared 19 GW of gas-fuelled generation out of 43 GW total cleared capacity[1].

However, an intuitive alternative to building more capacity would be to encourage the development and use of DR resources. Indeed, in times of scarcity a transmission system operator (TSO) can identify consumers who are willing to reduce their demand for financial compensation (Lambin 2020). The European Network Transmission System Operator for Electricity (ENTSO-E) stresses that DR 'often has a high capacity value relative to its energy value in many countries' (ENTSO-E 2015). In this regard, the European Commission (EC) has officially required full technology neutrality in all types of CRMs in Europe (EC 2013). In the same vein, in the US the Federal Energy Regulatory Commission (FERC) with Order No. 2222 enables distributed energy resources participation in electricity markets (including capacity markets) (FERC 2021). This is in line with other policy mandates for encouraging DR participation at the wholesale market level. The Energy Policy Act of 2005 states, 'It is the policy of the United States that unnecessary barriers to DR participation in energy, capacity, and ancillary service markets shall be eliminated' (US 2005). In the UK, the minimum capacity threshold was reduced from 2 MW to 1 MW, allowing smaller entities to participate in the capacity market. The primary rationale behind this was to align capacity markets with the other energy markets; however, this also benefited smaller providers who have indicated their desire to operate independently of aggregators (ESO 2021). However, while most CRMs in Europe and the US generally allow the participation of DR units, the rules applied to each mechanism differ substantially. This is related to the fact that unlike conventional power plants DR cannot provide full power output throughout scarcity periods of whatever length due to their technical characteristics. The rules defined for DR participation in a given CRM have a strong impact on the competitiveness of these technologies.

In this paper we discuss how DR resources can participate in wholesale markets as well as the associated system benefits. Even though the system's benefits, as well as those for consumers, from DR have been broadly demonstrated and documented (see for example, O'Connell et al. 2014, Dupuy & Linvill 2019), we notice that residential DR – with great potential with the use of heating, refrigerators, air conditioning, electric vehicles (EVs) and others – has not yet been fully developed. This is due to the several barriers, sourcing from market, political, social and technological factors, to name a few, that residential consumers need to overcome. On the bright side, conditions today favour the development of DR in the sense that some of these barriers can be addressed with the appropriate mechanisms. In particular, we propose the use of non-linear pricing and more specifically priority pricing contracts in the residential sector to incentivize its participation in DR. Priority pricing contracts offer consumers a menu of reliability levels with different prices. For example, let us consider three pairs: (i) cheap power that can be interrupted frequently; (ii) power that can be interrupted in emergency situations; and (iii) expensive power that cannot be interrupted. One main advantage of priority pricing contracts are their simplicity that makes them appropriate for residential DR programs. These contracts are between consumers and an aggregator who then participates in capacity markets. We propose the participation of the aggregator in capacity markets as a demand resource so that consumers express their willingness to pay for capacity. As such, one of the key shortcomings of a capacity market is addressed. To this end, we analyze how reliability concepts, say, the value of lost load (VOLL) and Expected Energy Unserved (EEU), are related with capacity and how optimal capacity investments may be determined. In order though for capacity markets to welcome such initiatives there is a need for some changes in the capacity product definition. In this regard, we propose a refined capacity market product that is expanded to include location as well as flexibility requirements. Last, we develop a business model for an aggregator and analyze how priority pricing contracts interact with the retail market so that potential market distortion issues are resolved. To summarise, the contributions of this paper are outlined as follows:

[1] It has also the rule that resources which are currently receiving, or will receive, support under Contracts for Difference (CfDs), Final Investment Decision Enabling Regime (FIDeR), Feed in Tariffs (FiT), and Renewables Obligation (RO) are not eligible for participation in the capacity market (ESO 2023b)

1) Integration of priority pricing contracts via an aggregator as explicit demand response in capacity markets, participation on the ‘demand’ side and construction of the capacity demand curve
2) Refinement of capacity market products to encourage investment in demand response resources and incorporate specifics about location, flexibility, and adjustments to the calculation of firm capacity profile
3) End-to-end business model for an aggregator incentivising residential demand response participation in capacity markets

The remainder of the paper is organized as follows. In Section 2 we provide an overview of CRMs around the world and their basic principles; we then focus on capacity markets and their shortcomings. Next, in Section 3 we provide an overview of the benefits of DR participation in wholesale markets, comment on the barriers that DR face, and how some of these can be lifted with appropriate pricing schemes like priority pricing contracts. In Section 4, we discuss how capacity markets need to be adapted to incorporate residential DR on the ‘demand’ side and how the capacity product exchanged can be refined so that it represents the needs of low-carbon future power systems paradigm with increased need for flexibility coming from DR. In Section 5, we propose an end-to-end business model for an aggregator implementing the participation of DR in capacity markets through priority pricing contracts. Lastly, in Section 6, we summarise the results and provide concluding remarks.

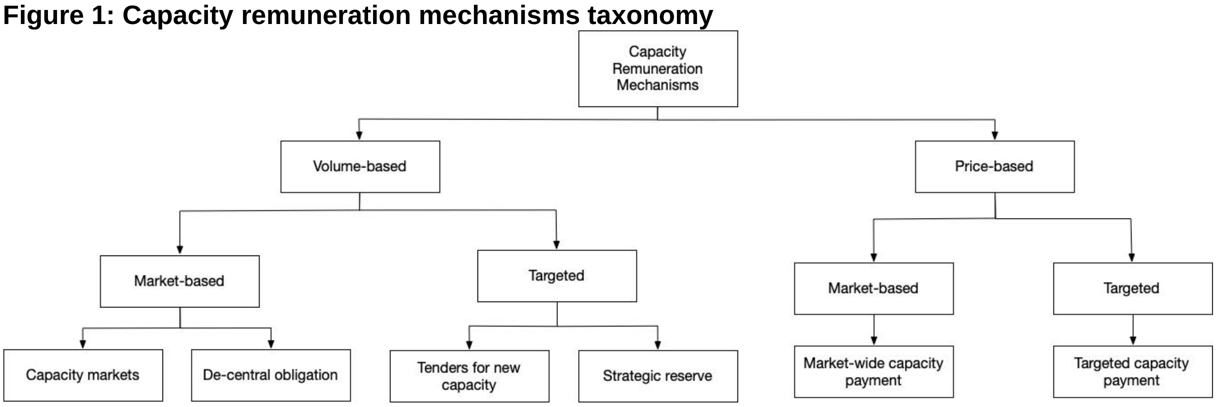
2. Overview of Capacity Markets

In this section, we categorize the CRMs and discuss their basic principles of operation. We then focus on capacity markets and discuss their shortcomings in terms of incentivizing new investments in appropriate locations and with specific characteristics.

2.1 Capacity Remuneration Mechanisms

In Section 1 we discussed the need for CRMs to guarantee new investment targets and ensure that all generators could recover their capital costs. Initially, some countries including some European countries did not see the value in CRMs; however, this has changed in recent years since short-term, marginal cost-based prices usually cover the short-term costs rather than the capital costs (Keay & Robinson 2019). In this regard, a number of studies have investigated how markets can be redesigned to incentivise a socially optimal generation portfolio. A taxonomy of different CRMs that have been introduced to address this missing money problem is depicted in Figure 1 (ACER 2018).

Figure 1: Capacity remuneration mechanisms taxonomy



Source: adapted from ACER 2018

These are broadly categorized into volume-based mechanisms, where a specific capacity sufficient to guarantee the desired level of system adequacy is set and price-based mechanisms, where the amount of the procured capacity is steered by setting a target price. Both categories can also be subdivided into market-wide and targeted approaches. Whereas market-wide mechanisms provide support to all capacity in the market, targeted mechanisms aim at supporting only a subset, for instance, newly built capacity or capacity expected to be required additionally to that already provided by the market. More specifically, six different types of mechanisms can be differentiated:

1) Capacity markets: The total amount of required capacity is set by a central body and procured through a central bidding process so that the market determines the price. Two common variants of the central buyer mechanism include the forward capacity market (see Cramton & Stoft 2005) and reliability options[2] (Oren 2005). Such mechanisms are found in the UK, Ireland, Italy, Poland, ISO New England (ISONE), MISO, NYISO, and PJM Interconnection (PJM).

2) De-central obligation: This refers to the obligation of each load-serving entity to secure the total capacity it needs to meet its consumers' demand. In contrast to the central buyer model, there is no central bidding process. Instead, individual contracts between load serving entities and capacity providers are negotiated. Such mechanisms are found in France, Australia, CAISO, and SPP.

3) Tenders for new capacity: Financial support is provided to capacity providers to build the required additional capacity, for example by financing the construction of new capacity, or offer long-term power purchase agreements (PPAs). Such mechanisms are found in Bulgaria and Croatia.

4) Strategic reserve: A certain amount of additional capacity is contracted and held in reserve outside the energy-only market. The reserve capacity is only operated if specific conditions are met, as in a shortage of capacity in the spot market, or a price settlement above a certain threshold. Mechanisms like these are found in Germany, Belgium, Sweden, and Finland.

5) Market-wide capacity payment: A capacity price is determined centrally based on estimates of the level of capacity required to meet reliability criteria. This is then paid to all capacity providers in the market.

6) Targeted capacity payment: A central body sets a fixed price paid only to selected technology-type resources. Such mechanisms are found in Spain and Portugal.

CRMs are typically designed to maintain generation adequacy and ultimately avoid shortage situations by offering capacity providers income on top of their earnings in energy and ancillary services markets. Although mechanisms may vary substantially in the way the required capacity and the corresponding capacity prices are determined, all types of CRMs should in theory lead to similar outcomes. Price-based mechanisms, such as the targeted capacity payment approach used in Spain, face the problem that if payments are too low there is no guarantee that investments will occur, while if payments are too high, excess investments can happen; both scenarios are equally inefficient. CRM approaches that only support new plants may create distortions in the energy market since they depress pricing, thus exacerbating the need to pay existing plants to avoid their closure. In Europe the most popular CRMs are strategic reserves and capacity markets, as seen in Table 1 where an overview of implemented CRMs is presented (adapted from Bublitz et al. 2019).

[2] Reliability options can be introduced in capacity markets by requiring every generator that receives a payment for capacity to sell a reliability option for the same amount of capacity. The buyer of the option contract has the right to buy equivalent electricity on the wholesale market at a predefined 'strike price'. Reliability options are in essence risk sharing arrangements between load serving entities and capacity providers. Reliability options penalise generators that remain unavailable during a period when the spot price is above the strike price.

Table 1: Types of CRMs around the world and participation of technologies

Type	Market area	Thermal power plants	Variable renewable resources	Demand side management	Interconnections
Capacity market	Colombia	X	X		
	Ireland	X	X	X	X
	Italy	X		X	X
	Poland	X	X	X	X
	Belgium	X	X	X	X
	UK	X	X	X	X
	US - ISONE	X	X	X	X
	US – MISO	X	X	X	X
	US – NYISO	X	X	X	X
	US - PJM	X	X	X	X
De-central obligation	Australia – SWIS	X	X	X	
	France	X	X	X	X
	US – CAISO	X	X	X	X
	US – SPP	X	X	X	X
Strategic reserve	Germany	X		X	
	Sweden	X		X	
Targeted capacity payment	Spain	X			

Source: Bublitz et al. 2019

The main advantage of capacity markets, when well designed, is that they almost guarantee that the desired level of reliability is achieved. Furthermore, competitive bidding reduces prices and can encourage innovation, especially when bidding is open to new and existing capacity, both from the demand and the supply sides. One measure of the success of capacity markets is their increasingly widespread use. For instance, they have been central to almost all North American liberalized markets (with the exception of Texas), many Latin American countries (notably Colombia, Brazil, Peru, Chile), and a growing number of European countries, including the UK, Belgium, Italy, Ireland and Poland.

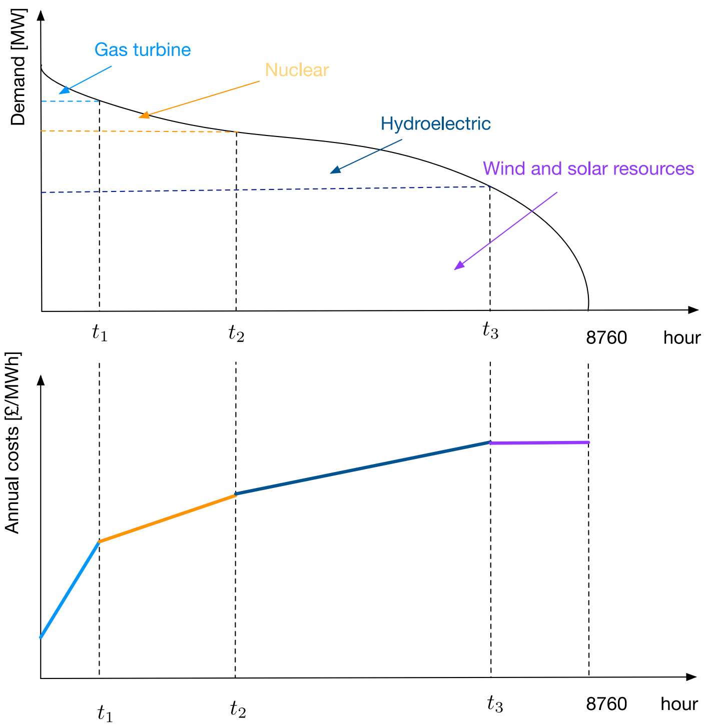
2.2 Shortcomings in Capacity Market Design

Capacity markets are designed to create the investment incentives that will provide revenue sufficiency to a least-cost set of generators or demand-side resources to meet resource adequacy goals. In contrast with a forward contract for energy (say, through a PPA), capacity markets only require the availability to deliver energy, but they do not specify the price at which the energy will be purchased. Capacity awards, therefore, provide an option-like payment to generators, and in return, consumers are less exposed to high energy prices due to the additional capacity that enters the market. Energy and capacity prices are linked because the implementation of capacity markets will tend to suppress the price spikes in the energy market that would have otherwise served as price signals for additional capacity. Unlike energy, capacity is inherently procured ahead of time, so the capacity resources will be able to contribute to system reliability during critical periods.

All capacity markets have some similar elements. Demand is based on a peak demand forecast that determines the need for future-installed capacity. Capacity market participants who are willing to supply power submit capacity offers. The market is cleared with a competitive auction where TSOs order capacity offers from the lowest to the highest, then use the capacity demand curve to determine the market clearing price of capacity.

Capacity market supply is determined by resources offering into the capacity market. Capacity offers are based on the unavoidable costs to keep each resource available, for example, ongoing maintenance costs of existing resources, or the capital investment costs of new resources or planned improvements. Since most resources will recover a significant portion of their capital costs through the energy market, capacity market offers are also reduced based on the resource's expected energy and ancillary services revenues. Peaking plants will tend to offer the highest costs in the capacity market since they are only expected to operate a limited number of hours per year, and will not receive as much energy market revenue as baseload resources do. An example of such a case is depicted in Figure 2; hydroelectric, nuclear, wind and solar resources are used the majority of the time and gas units only for a small fraction, as seen in the load duration curve. However, as seen in the curve depicting a linear approximation of their investment and production costs, the gas units are the most expensive.

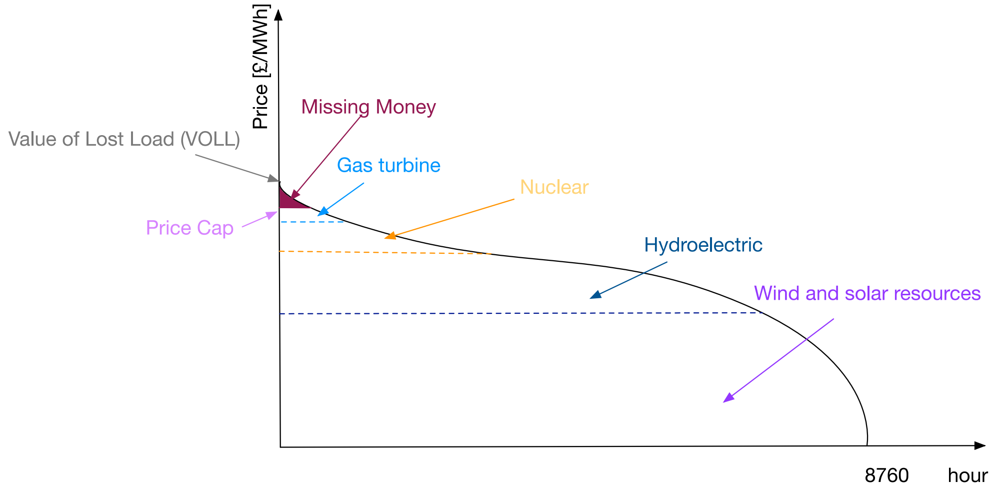
Figure 2: Load duration curve and approximation of investment and production costs for different technologies



Source: authors

This is exacerbated by the introduction of price caps to avoid the exercise of market power as seen in Figure 3 that mainly affects the peaking units.

Figure 3: Missing money problem



Source: authors

Capital and operational costs can also be affected by state subsidies (for example, a subsidized plant can offer capacity at near zero-cost to clear the market, which has led to price suppression and some rule changes within capacity markets). To make sure resources are able in practice to produce power or reduce demand when needed, several markets have added various 'pay-for-performance' capacity market mechanisms. For example, some markets (PJM, ISONE, NYISO) either pay resources a bonus or charge resources a penalty for each hour they do not meet their compliance obligation during certain compliance hours (like when the system is under shortage of capacity). Most markets employ capacity zones to reflect regional or more granular capacity requirements.

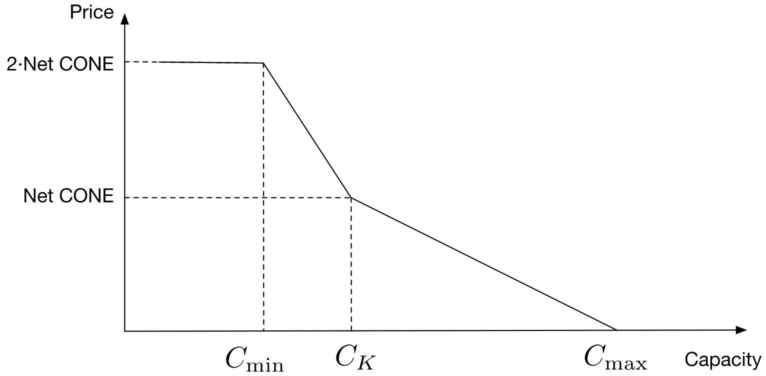
Conventional generators participate in capacity markets with an outage-adjusted capacity factor, the so-called equivalent forced outage rate on demand (EFORd)[3] multiplied by the resource's nameplate capacity (Cramton & Stoft 2005). Like conventional generators, renewable generator capacities are also derated of power capacity for participation in capacity markets. Based on the expected availability during system peak, wind is usually credited at about 20 percent and solar about 60 percent of their nameplate capacity. However, many TSOs have implemented enhanced mechanisms for renewable capacity market participation because, unlike conventional generators, unforced capacity (UCAP) calculations do not accurately capture the amount of system capacity provided by resources with correlated output. That is, UCAP will overestimate the capacity provided by renewable power generation since all solar or wind generation will have low output at the same time, and in addition, the typical change in renewable power output throughout the day can shift the timing of critical periods when the system is short on capacity. Most markets have implemented or are considering an approach called effective load-carrying capability (ELCC) to quantify the incremental contribution of renewables to capacity and resource adequacy needs. ELCC methods broadly consist of performing Monte Carlo simulations to estimate the marginal contribution of additional capacity given typical weather patterns and the market's existing resource portfolio. ELCC calculations are typically performed for each individual resource and, therefore, will vary by location. The effects of implementing ELCCs over previous heuristic methods can have a huge impact on investment decisions (see Dent et al. 2010).

Capacity market demand is different from energy market demand because it is determined administratively by the TSO. As such, capacity markets are essentially one-sided since they do not bring the buyers and sellers together; instead it is the TSO which is the single buyer (see Keay & Robinson 2019, Billimoria & Poudineh 2019). In this regard, reliability is seen as a public good, the desired value of which is set by the TSO or government, and not a private responsibility – which creates capacity market distortions. The framework proposed in this paper tries to address this problem through the

[3] EFORd refers to a metric used to gauge the likelihood that a power generation unit will be unavailable because of unexpected breakdowns or reductions in its power output when it is needed to produce electricity.

construction of the capacity demand curve where information from the consumers' value of reliability is used, as further explained in Section 4. The TSO calculates the required installed capacity, which is the capacity needed to meet forecasted peak demand – plus a capacity reserve margin. Next, the TSO calculates a price cap to anchor demand that is based on the cost of new entry (CONE) for a typical peaking plant, usually the cost of a new gas-fired power plant. The CONE represents how much investors are willing to pay to add new capacity. The TSO also calculates net CONE, which is the CONE minus energy and ancillary services market revenues. Net CONE estimates the missing money for the representative plant in the market. The TSO then uses a methodology to determine the downward sloping demand curve for capacity, which determines how much capacity the TSO will procure at each price point. The demand curve is the key element in the capacity market design. A proper demand curve should reflect the true requirement of generation supply and incentivize sufficient capacity investments. Originally, the implementation of capacity markets relied on inelastic demand curves for capacity, which resulted in a 'bipolar' behaviour of capacity prices depending on whether the system was temporarily short or long of the desired capacity target. The idea of a downward-sloping demand curve, which is nowadays typically employed in capacity markets, is to reduce the volatility of the capacity payments while aiming to cover, on average, the cost of building capacity and also allowing for a certain degree of mitigation on the exercise of market power in the capacity auction. The downward slope is based on the fact that people are willing to pay less for capacity once they have reached a desired reliability level. A prototypical design of a demand curve for capacity is discussed in (Cramton & Stoft 2005), and is presented in Figure 4.

Figure 4: Capacity market demand curve



Source: Cramton & Stoft 2005

In this figure three capacity levels are determined and associated with a price and reliability level, these are: minimum capacity (Cmin), capacity at the kink in the demand curve (CK), and maximum capacity (Cmax). The minimum capacity corresponds to the amount of capacity needed to keep loss-of-load events to, for example, 3 hours per year. More details on how these points are determined may be found in (Cramton & Stoft 2005). Among countries that implement a centralised capacity market, there is a consensus in utilizing a downward sloping demand curve, and the net CONE and target capacity level are key parameters in the definition of the demand curve. The fine balance that one attempts to strike when calibrating these demand curves is to secure adequate investment in capacity while ensuring that the procured capacity is not excessive, especially given the uncertain conditions that unfold in the energy market after the capacity market is concluded. The intersection of supply and demand determines the market clearing price. If the shape of the demand curve is wrong, the TSO will procure too much or too little capacity.

A first question is whether capacity markets are well-suited for incentivizing investments in renewable or carbon-free resources such as hydroelectric or nuclear. Various generation technologies depend on capacity revenues to varying degrees based on their competitiveness in the energy market. For example, some generators may require a large capital investment cost in order to produce electricity more efficiently and at a lower marginal cost. These resources can rely on the energy market to recover most of their investment costs. Conversely, other resources may have very low investment costs, but are accordingly less efficient and have more expensive marginal operating costs. These resources would not be very profitable in the energy market due to their high costs, but their low capital investment costs may make them attractive for meeting resource adequacy needs. In (Mays et al. 2019) the authors conclude that capacity markets, although nominally technology neutral, favour investment in resources with high marginal costs and low capital costs due to differences in the risks associated with energy and capacity market revenues. Capacity market incentives may, therefore, work against investment in resources with low marginal costs and high capital costs, such as renewable wind and solar resources.

Capacity market reforms have the potential to correct the issues discussed above and to better align investment incentives in renewable resources with system capacity needs. For example, very recently on January 8th, 2024 Ofgem sought to obtain views from people with an interest in the functioning of the capacity market as part of the Ten-Year Review of the Capacity Market Rules (Ofgem 2024a). The proposed changes focus on enabling battery operators to address issues around degradation, greater flexibility to allow low-carbon projects with longer-build times to access the scheme, new longer-term agreement options for low-carbon technology and measures to support the growth of the residential demand response sector. The locational contribution of renewable capacity is a major area for potential reforms. Because renewable power generation is correlated with geography and can affect the timing of peak net load, additional capacity investments can better improve resource adequacy if they are less correlated, or even negatively correlated with the existing renewable capacity mix. The authors in (Bothwell & Hobbs 2017) show how capacity markets can support more efficient investment incentives by using the location and type of renewable resources to calculate each capacity resource's marginal contribution to system resource adequacy. Today's capacity markets include zonal definitions that reduce capacity price signals to a rough approximation. However, the zonal definitions often follow preexisting regulatory or service area boundaries that do not reflect differences in resource variability. How to design zonal and system demand curves that appropriately capture the interaction between zonal and system capacities while keeping the curves in a simple form becomes the major technical challenge. Usually assuming statistical independence of generator outages and load, the desired reserve margin would be calculated by convolving generator outages and loads, considering unit nameplate capacities, forced outage rates, and load distributions. However, these simple probabilistic methods do not capture the increased uncertainty introduced by intermittent renewable generators whose outputs, like load, depend on weather patterns and cannot be modelled as independent.

Conventional generators remain available essentially year-round, with the exception of planned maintenance or (typically rare) unplanned forced outages, and this corresponds well with annual capacity payments that compensate the reliability benefit provided by the resource's consistent availability. In contrast, renewable resources are weather-dependent, with geographic and temporal correlations among separate resources. Demand-side resources also vary in their availability throughout the day. The capacity contribution of renewable, storage and demand-side resources is, therefore, not as straightforward as calculating an outage-adjusted capacity factor times the resource's nameplate capacity. For example, capacity markets in Colombia due to their regulation favour conventional thermal and hydroelectric power plants instead of DR resources; and faced adequacy problems in the el Niño hydro shortages in 2015-16. In the UK a parallel auction was held to enable DR resources as a transitional arrangement in the two years preceding full capacity market delivery in 2018-19. However, still in UK capacity markets only a small percentage of DR participates; namely less than 7 per cent in the T-1 auction for the delivery year 2023/24. One notable debate is on whether DR cleared in the capacity market should be required to offer into wholesale energy markets like generation (i.e., must-offer requirement). For generation, the must-offer requirement aims to mitigate the risk of resources withholding to drive up energy prices. These examples show, that there still exists no consensus about the role of such resources in capacity markets. While it is generally agreed that these technologies have some kind of capacity value, the specific rules of participation in capacity markets may hinder them from being competitive against conventional resources. Secondary trading in capacity markets is necessary when for example a capacity market unit is down for maintenance or, in the case of a new build asset, there are construction delays which means it will not be operational when required. It can also provide real-time information about how much capacity is worth; as such they might discourage unnecessary future investment in generation. This refers to the trade in capacity contractual obligations that owners no longer want, or are unable to fulfil, to other suitable providers. Currently, secondary trading in the capacity market is performed on a one-to-one basis, meaning capacity market providers have no way of gaining exposure to the entire market and finding a buyer can be time-consuming and complex. Full transparency into who would like to buy and sell contracts, and streamlining the trading process, would promote economic efficiency. Over the past two delivery years, the majority of secondary trades were made by DR resources.

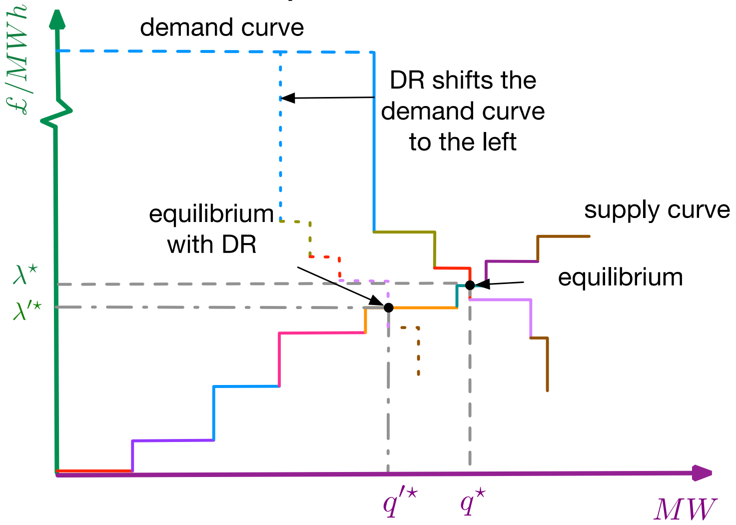
3. Residential Demand Participation in Capacity Markets

In this section, we provide an overview of the ways DR can participate in wholesale markets, consisting of energy, ancillary services and capacity markets, and investigate the barriers that residential consumers face when participating in DR programs with a focus on capacity markets. This allows us to understand which barriers need to be lifted to enable residential DR resources. Next, we discuss how dynamic pricing can lead to increasing DR participation, and in particular focus on priority pricing contracts which have several advantages compared with other dynamic tariff designs.

3.1 Overview of Demand Response Participation in Wholesale Electricity Markets

The literature provides various definitions of DR, but a clear common theme is that DR reflects electricity demand that changes based on a signal (Warren 2014). For example, FERC defines DR as 'changes in electric usage by demand-side resources from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized' (Murthy Balijepalli et al. 2011). Information sent to consumers in terms of, say, their consumption patterns, electricity prices, or the consequences of their consumption for the power system, can contribute to consumers adjusting their demand. More specifically, DR is usually based on the following mechanisms: price-based programs (or implicit demand response), which use price signals and tariffs to incentivize consumers to shift consumption; and incentive-based programs (or explicit demand response), which make direct payments to consumers who shift demand as part of a demand-side response programs. We therefore might say that reducing costs is the main reason for implicit DR (reducing or shifting demand) and earning revenues is the main reason for explicit DR (sales into markets). Advances in modelling and information technology capabilities have made DR an attractive option to increase power system flexibility. This will consequently allow a more efficient use of system assets and resources. The flexibility provided by DR can be used to meet the fluctuations of renewable generation and facilitate a higher penetration than could be achieved by relying on conventional generation alone (O'Connell et al. 2014). This is most effective in systems operating with market-based DR mechanisms as even a relatively minor DR will tend to displace the most expensive peaking units, reducing the system marginal cost and resulting in substantial welfare gains, as depicted in Figure 5. As it is seen in the Figure, the market clearing price without DR (λ*) is higher or equal than the market clearing price with DR (λ'*), i.e., λ'*≤ λ*. The market cleared quantity without DR (q*) is higher than the market cleared quantity with DR (q'*), i.e., q'* < q*.

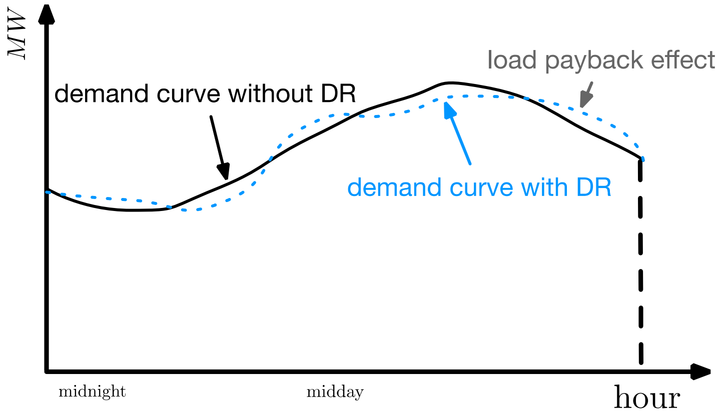
Figure 5: Impacts of DR on the market equilibrium



Source: authors

DR resources usually defer their demand to a time later or earlier than the critical period. This is the so-called 'load payback effect' as depicted in Figure 6.

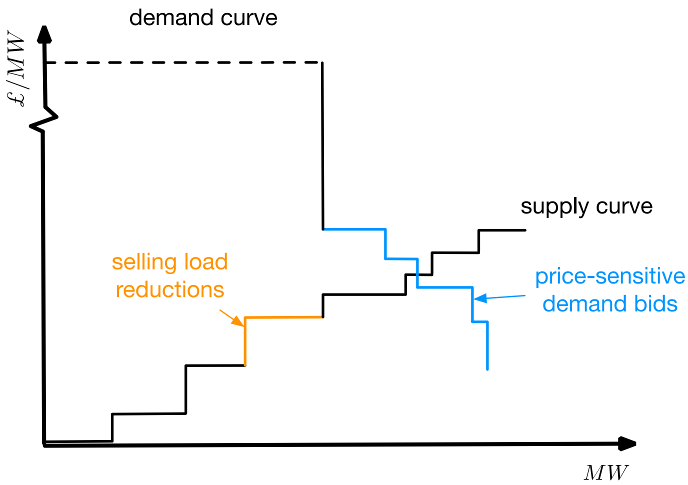
Figure 6: Impacts of the DR on the cleared power demand over a one day period



Source: authors

DR has both temporal and geographical diversity that can be used, among other things, for more economic use of interconnections, reductions in generation capacity requirements, transmission and distribution network congestion management, and increased economic efficiency. For example, according to a study made for the UK in 2008 the value of DR for reducing the generation capacity requirement is about £250-400/kW, and its value for congestion management that leads to deferred transmission network reinforcement is about £300/MW/km (Strbac 2008). Jurisdictions around the world have experimented with numerous ways of enabling more DR and to integrate it directly into wholesale markets. DR can participate in wholesale electricity markets in the following three ways: (i) energy markets; (ii) ancillary service markets; and (iii) capacity markets (Brown et al. 2015). Wholesale energy market participation of DR takes on three forms: simply purchasing less when the price is high (price-responsive load); submitting price-sensitive demand bids in day-ahead markets (but few or no loads do this in real-time markets[4]); and selling load reductions from a hypothetical baseline back into the energy market as supply (most often from resources that are also earning revenues in the capacity market). These forms of participation are depicted in Figure 7.

Figure 7: DR participation in energy markets under three different ways



Note: The price responsive load is not directly apparent to the system operator, but it still reduces demand when the prices are high

Source: authors

The volumes of energy provided are trivial in normal periods, since energy prices will usually be below the willingness-to-pay for almost all loads. Ancillary services markets now admit DR in numerous

[4] This is due to the high costs of complying with dispatch requirements compared with few benefits

jurisdictions, particularly for contingency reserves and regulation[5]. They are treated as 'supply' just like generators (Brown et al. 2019). Participation continues to be quite high in PJM and ERCOT.

Capacity market participation and emergency reserve mechanisms resemble traditional, vertically integrated utilities' interruptible load programs. They enable the system operators to 'dispatch' load reductions to keep the lights on when supplies are limited. Like other capacity or emergency resources, demand resources' load reductions are treated and paid as supply. This has enabled a flourishing industry of specialised third parties that develop, aggregate, and sell load reductions. Capacity revenues are their biggest sources of revenue, but they can sell energy and ancillary services from the same assets. As such DR mostly participate in capacity markets to earn revenues; however, the value of DR from the system perspective is also on their capacity compared to their energy. The value of DR in terms of capacity from a system perspective is demonstrated through participating in regulation, load following, ramping, and flexibility; thus replacing the need for additional capacity from generators. There are other markets too, notably including local energy and congestion markets. Piclo, for instance, provides platforms where DR can compete with other sources of flexibility. This could be done either for short-term energy supply or long-term capacity.

ENTSO-E stated that DR often has a high capacity value relative to its energy value in many countries (ENTSO-E 2015). Indeed, the cost of acquiring and maintaining generating capacity is a significant component of the total costs. DR is capable of reducing the need for generation capacity investments in order to ensure resource adequacy. The ability of flexible demand to balance wind fluctuations and reduce peak demand through demand shifting reduces the need for investment in expensive and often inefficient peaking and flexible plant such as CCGT units. There is strong policy interest in opening capacity markets to demand-side resources; for instance, in the US, ISONE and PJM already allow DR to directly compete against generation and other capacity types in their capacity markets. Loads that are curtailable by PJM are allowed to sell capacity into PJM's capacity market to help meet resource adequacy requirements. Participants must offer into the energy market, although most offer energy at only a very high 'strike price', such that they are curtailable only in emergencies. Capacity payments account for the vast majority of DR's revenue in PJM, at more than $500 million per year. In ISONE, all active demand capacity resources can receive obligations and compensation comparable to generating resources in the capacity market and, since 2018, are subject to the same performance penalties. The UK is one of the few European countries whose market and regulatory framework enable the participation of the commercial and public sector DR in both balancing and capacity markets (Cardoso et al. 2020). However, in the aforementioned examples, we notice that DR is coming from mainly industrial customers rather than residential. The advantage of having large industrial consumers provide DR is that their change in consumption patterns significantly affects the electricity system as a whole. Aggregation is cost effective and reliable in large volumes due to the tradability of their load flexibility on balancing and reserve markets.

3.2 Barriers faced by Residential Demand Participation

In (Good et al. 2017), the barriers of residential DR participation in power systems operations and markets are outlined and categorized in Table 2.

Table 2: Barriers' categorization for DR resources

Economic	Social	Technological	Political	Structural
Market inefficiency **Market barriers**	Organizational Behavioural	Sensing Computing Communication Technology standardisation Technology skills	Taxes Energy markets regulation End-user price regulation Unclear policy	Baselines Product definition Complexity

Source: Good et al. 2017

[5] These include services necessary to balance the grid, e.g., dynamic regulation and slow and quick reserves (ESO 2023a)

In terms of economic and market structures barriers, one of the greatest barriers for DR is the lack of appropriate market mechanisms in current market structures. Currently, DR is primarily employed for the provision of emergency contingency support and ancillary services, with limited participation in the day-ahead market. This participation occurs in the form of direct market bidding as well as contracts between individual market stakeholders. The restrictive nature of these markets and contracts often requires that DR is planned many hours ahead, or that substantial advance notice is required before the demand is adjusted in emergency scenarios. For example, in a capacity market, a particular load may be capable of adjusting demand instantaneously, but regulations may require an advance warning of three hours, making the effective switch-on time of this resource three hours. Long lead times between the conclusion of the capacity contract and the start of the delivery obligation for the successful bidders may also hinder the participation of distributed energy resources, especially some demand response units that may not be able to commit for too long in advance of the delivery period. In Dupuy & Linvill (2019) the authors document several barriers to participation for demand-side resources in capacity markets, including complex market rules, limited geographical and high minimum aggregation requirements, as well as 365-days/year availability requirements. For example, capacity markets in Ireland, Italy, and Poland do not have a time-limited availability period, that is to say capacity resources must remain available all year round. Moreover, the minimum eligible capacity is 2 MW and 10 MW in Poland and Ireland, respectively; and both Poland and Ireland have multi-year capacity contracts up to 17 and 10 years in length, respectively (ACER 2023*a*). Moreover, capacity markets are too blunt in their ability to signal locational and temporal needs. For example, demand response may be effective for alleviating transmission congestion or as a substitute for generator ramping flexibility, but these services are difficult to value in a capacity market construct. For example, PJM has system-wide price zone with 12 subzones but with no direct mapping to PJM load zones. Capacity markets may also generally undervalue DR due to the inclusion of price caps, given that price caps primarily affect resources that are infrequently dispatched, such as DR. Such limitations, as well as stringent telemetry and performance standards, prevent demand from participating effectively in the capacity market.

Furthermore, the capacity product definition is based on firm capacity and calculated based on effective load carrying capability (ELCC), as further elaborated in Section 2.2. However, the quantification of the DR capacity value is not a trivial exercise, not least because there is no reliable ‘counter-factual’, as in there is no way of knowing what the equilibrium levels of electricity demand would have been in the absence of DR. To overcome this barrier, it is necessary to establish a baseline and determine the load reduction DR achieves. For DR, it involves estimating the difference of real-time load during a DR event and the ‘counter-factual’ baseline established from loads in recent similar days and right before the DR event (Liu 2017). Furthermore, while DR is very different from renewable generation, they do share some common characteristics. For example, like wind and solar generation, which are highly reliant on an uncontrollable meteorological phenomenon, load-shifting is dependent on an underlying resource which is largely driven by variable consumer demand behaviour although largely predictable if aggregated. In terms of further market barriers these include access to capital, uncertainty and hidden costs. For example, the initial technological investments of the installation of smart meters, in-home displays, and other devices for enabling DR participation in capacity market are costly. Moreover, the revenue streams are uncertain. Bidding into capacity markets means direct competition against other resources – depending on the offers of other resources, there is uncertainty of whether DR clear the auction and the level of payment. Since the implementation of the PJM capacity market on 1 June, 2007, the capacity market has been the primary source of DR revenue. In the first nine months of 2023, total DR revenue decreased by $217.0 million, or 61.8 percent, from the $351.4 million of the first nine months of 2022 to $134.3 million, primarily due to a decrease in capacity market prices and revenue (Monitoring Analytics 2023). Hidden costs refer to negotiation and enforcement transaction costs, which can be large for residential customers.

In terms of social barriers, consumers’ perceptions are important. System operators recognise that demand is a valuable resource, but that consumers may withdraw from demand response programs if the inconvenience of participating becomes too high. Human nature is a further issue which compounds the problem of market design for demand response. While large generators typically exhibit economically rational behaviour through their profit maximizing objective, smaller customers do not show the same rationality in their consumption decisions. End-users, particularly in the residential sector, have many different priorities, and minimizing their electricity bill may not always be at the forefront of their concerns. In contrast, the profit-driven objectives of generators means that their behaviour fits established economic models. Consequently, enough information can be drawn from their bidding behaviour for their supply curve to be revealed. The corresponding demand curve is much more difficult to extract from demand behaviour due to its dependence on many different and time-variable external factors, ranging from the weather to whether the consumer cooks dinner using an electric oven or a gas cooker. Empirical studies have demonstrated some of the ways in which consumer demand does not fit the conventional economic model. Energy is a special kind of commodity that fails to arouse much interest among the vast majority of small end-users, making it difficult to understand variations in price and the need to consume flexibly (Kahma & Matschoss 2017). A comparison between buying a new car and paying for electricity is made in (Kim & Shcherbakova 2011). Both of these actions account for approximately the same proportion of annual household expenditure (when considering annualised car payments), but significantly more thought is put into the car purchase. As the authors mention, this is because payment for electricity is a passive action which occurs at regular intervals, so does not require substantial consideration from the consumer. This lack of interest results in a low response to price changes. Increasing consumer awareness in this manner can increase their flexibility to price signals.

As for technological barriers, many of the monitoring and communications technologies required for widespread demand response are currently available. The deployment of smart meters and information and communication technology (ICT) infrastructure enables a paradigm shift in the way electricity systems are operated, transforming traditionally passive end-users into active market players (see He et al. 2013). The central remaining technological obstacle is the development of standards and protocols so that all components of this complex system are harmonised, and efficient communication can be achieved across the system.

As per political barriers, end-user price regulation does not allow end-users to face full variation of wholesale energy-related markets or to effectively communicate user preferences in terms of reliability and be able to make reliability a product that can be differentiated and traded. This is related to current regulatory and tariff structures, particularly for residential customers. Until now, time-based pricing has been applied mostly to incentivise large industrial users, leaving the approach unclear for residential customers (ACER 2023*b*). However, in future electricity systems the generation-follows-demand perspective will be increasingly replaced by a demand-follows-generation paradigm. This transition requires that electricity consumers receive real- time reflective information regarding the electricity prices through dynamic pricing. Typical billing components on electricity bills are the energy costs, network charges, and government obligations. The latter include environmental and social tariffs to support the decarbonization of the electricity sector and enclose capacity market payments (if any). The larger the component of the energy costs in the electricity bill, the more room there is for cost savings for DR. For example, in the UK the energy component has varied from 30-55 per cent over recent years, mostly at the higher end due to the recent energy price shock. It has now reduced to 30 per cent (Ofgem 2024*b*). Traditional electricity charging with uniform rate structures does not further differentiate between consumers and prosumers. However, in a smart-grid environment, where demand participation is fostered, end-user tariffs can incorporate customer categorization, time and location (see Picciariello et al. 2015). For example, in Norway where there is a high share of EVs and heat pumps, more than 90 per cent of customers have chosen to be exposed to price signals with dynamic electricity price contracts. Customers are metered at an hourly resolution and are directly exposed to the day-ahead prices which enables implicit demand response and can be supported by (out of market) automation (ACER 2023*a*).

3.3 Priority Pricing Contracts for Residential Demand Response

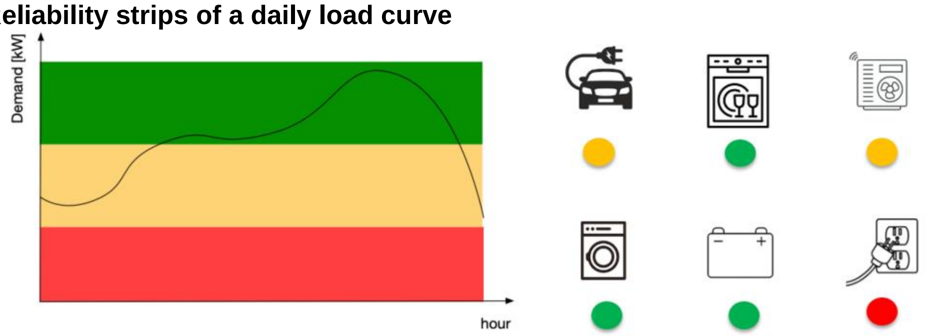
Consumers may attach different value to the guarantee of continuity of supply. This is especially the case when considering consumption for non-essential uses more than what is necessary to cover basic needs (lighting, cooking and heating). Therefore, for implicit DR, schemes have been proposed where consumers can choose their preferred level of continuity of electricity supply and pay for it. Those choosing for a higher continuity of supply will pay higher charges than those who are willing to have their continuity of supply interrupted at times of scarcity. Implicit DR schemes span two extremes – price-based and quantity-based methods (see Borenstein et al. 2002). In quantity-based methods, such as direct load control, an aggregator can modify a certain amount of the load. An example of such a scheme is to control thermostatic loads by modifying the temperature levels for load following (Mathieu et al. 2013). In terms of price-based methods, real-time pricing represents the benchmark of economic efficiency. Under real-time pricing, demand is adjusted in real-time in response to prices that reflect the instantaneous needs of the system. However, procuring electricity in real-time cannot be easily accommodated by residential consumers due to the high information overload and complexity. As such, to increase the participation of residential consumers in DR there is a need for either simpler approaches or retention of complex pricing signals that are met by aggregators who reward their consumers for allowing remote control or messaging to influence energy use.

Time-of-use (TOU) pricing is a simple, widely adopted pricing scheme where a day is typically divided into a peak period and an off-peak period[6]. Under TOU pricing consumers reduce their consumption at peak periods and potentially shift it to less expensive off-peak periods. TOU signals are more frequently used for consumers connected at the distribution level compared with transmission level: 21 out of 27 countries (over 75 per cent) in the EU apply static TOU charges in distribution tariffs and about one third in transmission tariffs (ACER 2023b). However, TOU tariffs may not be effective in future power systems, because the peak or off-peak periods are difficult to specify as renewable resources are intermittent and uncertain, thus making periods of reduced load potential candidates for peak periods. Future low-carbon systems would require dynamic peak and off-peak periods. Another approach to simple pricing schemes is critical peak pricing (CPP), which is closely related to TOU pricing, and includes the addition of certain critical hours that can be announced to consumers by the utility with short notice. Both TOU and CPP tariffs attempt to reduce peak demand; more details on the aforementioned designs may be found in (Borenstein et al. 2002).

We now introduce the use of priority pricing contracts as a means to promote the uptake of residential DR (Chao & Wilson 1987). The idea behind this pricing scheme is to bridge the gap between retail and wholesale prices but maintain, at the same time, the simplicity, independence and privacy of residential consumers. Priority pricing theory is based on the differentiated continuity of supply and prices. Consumers have different valuations for different reliability levels. Retailers are not aware of which consumers are characterized by higher valuation. However, by facing different prices for different reliability levels consumers reveal their valuation for reliability. This results in higher efficiency in both normal and tight system operations. Several examples of differentiated service exist in non-electricity industries. For instance, telecommunication companies offer different plans for data with different monthly subscription fees and network speeds. Transportation and postal systems offer express and regular services. Thus, customers have experience with choosing between alternative service contracts.

The idea behind the implementation of priority service pricing is to offer residential consumers a menu of reliability and price pairs (see Papalexopoulos et al. 2013). For example, we may define three pairs that the consumers may choose from. We use a colour coding system to differentiate the different levels of reliability: (i) *green* indicates cheap power that can be interrupted frequently; (ii) *yellow* indicates power that can be interrupted under emergency conditions; and (iii) *red* indicates expensive power that cannot be interrupted. As depicted in Figure 8, load is divided into three strips that correspond to different levels of reliability. An example of what appliances are associated with a given reliability level is given in Figure 8. Note that the consumer can change the colour code of an appliance at their desire.

Figure 8: Reliability strips of a daily load curve



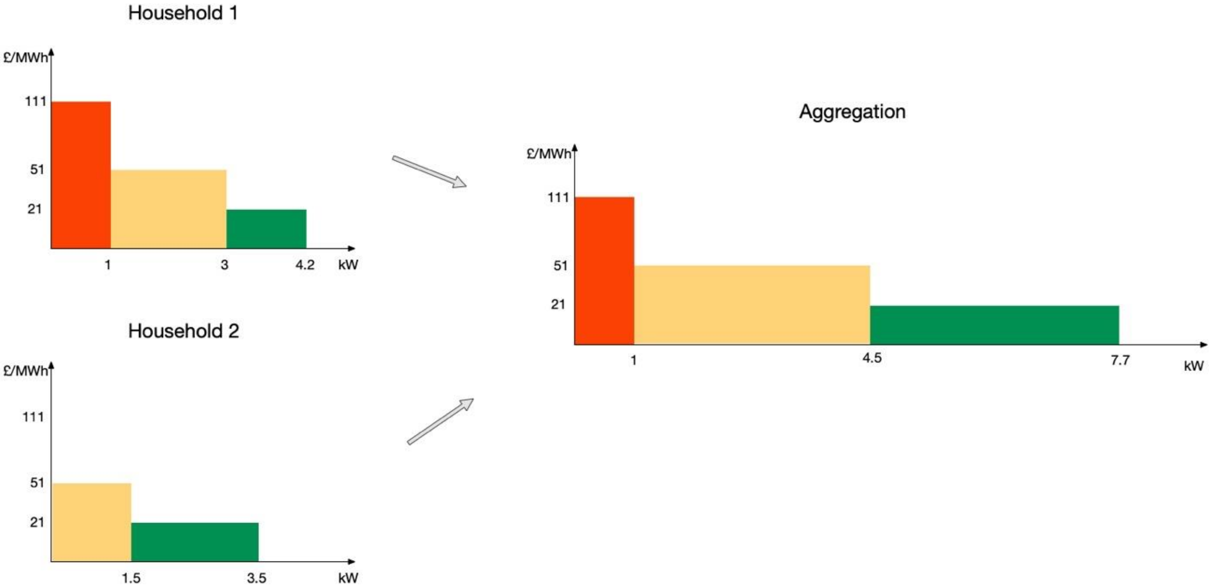
Source: authors

An example of a priority pricing menu consists of £111/MWh for 100 per cent reliability (red), £51/MWh for 90 per cent reliability (yellow), and £21/MWh for 70 per cent reliability (green)[7]. The red colour

[6] These can also vary seasonally.
[7] The 90 per cent reliability level means that 90 per cent of the time that energy is requested it is actually provided.

absolutely guarantees service, which is why it is the most expensive. Next the yellow colour has a reliability of 90 per cent and thus is priced less than red, with yellow corresponding to load that may be interrupted when the system is stressed. Green colour has the lowest price due to it being highly unreliable, for example it may be only available when there is abundant renewable production. An aggregator offers priority pricing contracts to consumers and then aggregates them. An example of such an operation for two households is demonstrated in Figure 9. Household 1 has a typical inflexible load, a typical flexible load, for example, a washing machine, dishwasher and tumble dryer, and a rooftop PV. Household 2 has all that and a battery and an EV. As such, household 2 has greater room to offer flexibility. This is reflected in the choice of contracts from the two households. Household 1 has a great amount of the red colour reliability and a very small amount of the green colour reliability. On the other hand, household 2 has a large amount of the green reliability level and zero amount of the red reliability level. Their aggregation is also shown, and this can be scaled up to a high number of households.

Figure 9: Example of priority pricing contracts



Note: Red corresponds to 100%, yellow to 90% and green to 70% reliability levels

Source: authors

This approach presents consumers with the benefit of reduced electricity bills through the use of priority pricing. However, as seen in this simple example, the benefits of priority pricing are better magnified for future consumers that have additional sources of flexibility. We provide an additional example to demonstrate the potential cost savings under different type of consumers. Let us assume we have the priority pricing contracts presented in Table 3.

Table 3: Priority pricing contracts with 70%, 90% and 100% reliability levels

Reliability [%]	Priority charge [£/MWh]
70	21
90	51
100	111

We define four types of consumers as shown in Table 4. Type 1 can be seen as a traditional consumer and Type 4 as the future consumer which will be more common in future low-carbon power systems by 2050 (ESO 2023*a*).

Table 4: Consumer types definition

Consumer Type	1	2	3	4
PV	No	Yes	Yes	Yes
EV, electric heat pump	No	No	Yes	Yes
Battery	No	No	No	Yes

Then, each consumer would make the following selections from the priority pricing menu given in Table 3 as shown in Table 5[8]. The cost calculation for priority service charges for one day, e.g., Type 1 is as follows: $0.1 \cdot 21 \cdot 24 \ \cdot 30 + 0.05 \cdot 51 \cdot 24 \cdot 30 \cdot 0.9 + 0.2 \cdot 111 \cdot 24 \cdot 30 \cdot 0.7 = £18.7/day$.

Some key take away points from this simple example:

- Consumers that put less value on reliability can take greater advantage of the green and yellow reliability contracts and see greater savings in their bills.
- Consumers that have power resources on site (most extensively, Type 4s), choose to self-consume as much as possible and buy less energy from the grid.
- Most current consumers that are modelled as Type 1 can still benefit from priority pricing contracts and see a 33 per cent reduction[9] in their bills.
- All consumers see reductions in their electricity bill without increasing their discomfort levels due to the existence of flexible resources such as EVs, electric heat pumps, and batteries.

In this simple example we do not consider the initial investment costs needed for the purchase and installation of these resources.

Table 5: Examples of priority pricing contracts for each consumers' type

Consumer Type	1	2	3	4
Priority Pricing Contracts				
Green	0.1	0.1	0.15	0.15
Yellow	0.05	0.15	0.05	0.03
Red	0.2	0.01	0	0
Cost [£/day]	18.7	6.8	3.2	2.6

Note: The quantities in the three reliability colours are in kW and the cost in £/day.

Priority pricing contracts are fully defined by a menu of (i) priority charge, which is paid in advance; (ii) a service charge, paid as energy is provided; and (iii) reliability level (Gérard & Papavasiliou 2022). The above example focuses on the service charges of the contracts.

Priority pricing at the household level may be implemented either manually or automatically through a home energy router by tagging plugs with colours that correspond to reliability levels in the home. In this way households preserve control on their household consumption. The control and communication technology needs for implementing priority service pricing contracts are a means of tagging plugs according to reliability levels, an ability to monitor different reliability levels in near real-time intervals, and an energy router that can receive control signals from an aggregator. In offering priority service contracts to residential consumers, the aggregator commits to a certain level of reliability for each service option. This level of reliability must be respected on average over an extended period of service

[8] These are fictitious numbers but their relative importance shows that consumers with more flexibility resources choose greater amounts of lower reliability contracts.

[9] This calculated as if all consumption was at the 100 per cent reliability level.

(for example, annually). By selecting plans, consumers reveal their valuation for power. These valuations can be aggregated and bid into the wholesale electricity market and the capacity market; this is explained in Section 4.2. More details on the responsibilities and actions of an aggregator to promote implicit residential DR through priority pricing contracts are discussed in Section 5 where a potential business model for an aggregator is presented.

Priority pricing contracts demonstrate some advantages compared with other contract types. We first compare it with the two extremes, namely, the real-time pricing, and the quantity-based direct load control. Priority pricing can be viewed as a compromise between price- and quantity-based methods that attempts to combine the best of both worlds. Under real-time pricing residential consumers face fundamental challenges in assessing their real-time valuation for power. However, with priority pricing contracts these complexities are passed on to, for example, the home energy router. Moreover, grid operators are usually reluctant to offer price-based DR schemes since consumer responses to changes in price are uncertain and hard to predict, leading to uncertain load variations. With priority service contracts demand is predictable since an aggregator controls the interruption of colours which are in turn represented by a certain amount of kilowatts. On the other hand, direct load control is perceived as being overly intrusive. Consumers prefer to maintain control over their consumption. In priority service, as discussed above this is achieved by the colour tagging system where the consumer prioritizes a certain device over another but can also swap them anytime they wish. As a result, as demonstrated in Table 6, consumers keep control of their load and their energy volume risk is medium since they are expecting a certain level of unavailability of service, based on their contract type. Another important weakness of direct load control relates to a lack of privacy. In priority pricing contracts the aggregator is only aware of the aggregate kilowatts assigned to different service levels, as opposed to knowing which devices specifically constitute those kilowatts at any given balancing interval.

A comparison of several contract types based on six criteria is given in Table 6.

Table 6: Comparison of different types of tariffs

Contract type	Signal volatility	Price Risk	Economic Gain	Volume risk	Complexity	Intrusiveness
Flat Tariff	Static	None	None	None	None	None
TOU pricing	Static	Low	Limited	None	Low	None
CPP	Static	Low	Limited	None	Low	None
Priority pricing	Dynamic	Medium	Medium	Medium	Medium	None
Real-time pricing	Dynamic	High	High Potential	None	High	None
Direct load control	Pre-defined	None	Varying	High	None	High

In summary, the priority service pricing scheme addresses some of the barriers to DR, as outlined in Table 2, in terms of behavioural matters, like privacy concerns, consumer engagement, product definition, and complexity. For example, the use of automation, such as the colour system described above, enables DR by making operational decisions which the user is unable or unwilling to make themselves. This also reduces the level of inconvenience for the user as they have in effect delegated decision-making, further enabling DR. Another advantage is that it includes the ‘opt-out’ function as elaborated on previously. The consumers enrol in an electricity service but preserve control of their household consumption. This is a very important characteristic of priority pricing contracts since, while consumers are willing to pay for technical support services, they are likely to demand significant compensation to participate in automated demand response programs that involve remote monitoring and control of electricity usage (Richter & Pollitt 2018). Moreover, the authors of the paper studied the acceptance rate for example of contracts that combine a fixed compensation payment with a transaction-based fee component, namely a payment per £1 saved in the electricity bill. They find that a higher fixed monthly compensation can partly make up for a higher fee to expected savings ratios, for example, a compensation of £4 and a fee-to-expected-savings ratio of 0.5, had a 35 per cent rate acceptance rate compared with a 24 per cent acceptance rate based on an example contract providing a £2.19 fixed monthly compensation with a 0.33 fee-to-expected-savings ratio. This can be considered when designing the priority pricing menu, since there is flexibility in determining the service (variable) and priority (fixed) charges. By using the results of (Richter & Pollitt 2018) we might argue that a contract with a higher fixed price and a lower variable price might have a higher acceptance rate, since customers prefer lower uncertainty in prices.

Challenges in time-varying rates are mainly insufficient evidence of benefits, potential customer dissatisfaction, and their effects on sensitive or disadvantaged customers. When moving to time-varying rates, the costs are allocated in a more efficient and fair manner. However, there are consumers that experience a bill raise as a result. They need time to change their consumption patterns and manage their consumption in response to a new time varying rate to reduce their bill. Moreover, special attention needs to be dedicated to the needs of customers with medical disabilities, customers who are unemployed and low-income customers in general in a transition to time varying rates. There are some successful examples of time varying tariffs. As an example, Google sells Wifi-connected smart thermostats that let users adjust the temperature of their homes from afar, and in exchange for a monetary incentive lets utilities reduce the energy use when the grid is strained, but they are still at their infancy. As in any tariff structure change, for priority pricing contracts to be successful, there is a need to study the effects on customers electricity bills; understand the likely response of customers through analysing prior pilot studies; and invest in customer outreach and education, for instance, explain why tariffs are changing and how they will work. Barriers to large-scale implementation of residential DR include effects on sensitive or disadvantaged customers, that may be addressed by, say, offering rebates to low-income customers and carrying out energy efficiency improvements in their houses. Other options include providing bill protection for the first years as consumers get acquainted with the new contracts, that is to say ensuring that customer bills will not rise but they will be able to keep the savings, with those protections being phased out gradually over time. Further protections could be added for sensitive customers in the form of financial assistance for a limited period of time (Faruqui & Tang 2023).

4. Proposed Capacity Market Reform

As discussed in Section 2.2 some of the shortcomings of the current capacity market design involves that it is a one-sided structure, its reliability levels are set by a central entity, and that it does not incentivize the participation of resources arising from future power systems, for example, DR or storage resources. In this regard, in this section we address the 'capacity product definition' issue with a two-fold approach: (i) an alternative methodology of determining the capacity demand curve, considering consumers' willingness to pay for capacity; and (ii) the specification of a refined product to be traded in capacity markets that involves appropriate locational, temporal and flexibility signals.

4.1 Demand Response as a 'Demand' Resource

DR resources that participate nowadays in capacity markets do so from the supply side, since, as discussed in Section 2.2, the TSO is the only buyer in the capacity market. This is also in line with the proposing of harmonized market rules for both DR and capacity, in compliance with the EU Energy Efficiency Directive (EU 2012) so that DR and traditional generation are on a level-playing field. As in traditional generation where an outage-adjusted capacity factor is used, it is common to only consider a portion of the DR bid capacity as qualifying for remuneration, since there is some uncertainty regarding the amount of demand response capacity that can be provided in real-time. For example, if the aggregator is using a price-based DR paradigm, the response of consumers to changes in price can be highly uncertain. When aggregators sell a firm supply in the capacity mechanism by using DR they make a formal commitment. As such, they need to demonstrate that the demand reduction is credible, which makes the definition of a baseline necessary. This is translated in the existence of robust measurement and verification techniques to ensure that transitional DR programs have the desired behaviour.

If DR resources are to continue competing directly with generation capacity resources in the capacity markets, they need to also adhere to the same regulations in order to obtain a functional market design – which is currently not the case. For example, DR resources do not have a must-offer requirement into the day-ahead energy market and are able to offer above a large price per MWh without providing a fuel cost policy, or any rationale for the offer. The must-offer requirement that traditional generating units must meet is that resources with a capacity supply obligation are required to offer into both day-ahead and real-time energy markets during all hours, at a MW amount equal to or greater than their capacity supply obligation amount, whenever the resources are physically available. However, DR capacity resources considered 'emergency only' are not subject to the must-offer requirements, and are not economically dispatched unless they similarly offer into the energy markets. For example, in (Lambin 2020) the author elaborates on the need to include derating factors in DR participation in capacity markets in the absence of a must-offer requirement, since some DR bid very high prices in the energy market; thus, they never get dispatched, making their capacity value less than other DR resources. Other modifications to current regulations include that DR resources need to have telemetry requirements similar to other capacity performance resources, provide a location, and should be dispatched per location to enhance the effectiveness of demand resources and be held accountable for failures to dispatch, like traditional resources.

The more intuitive approach is to let DR resources participate as "demand" resources through the capacity demand curve. Capacity demand curves are supposed to reflect consumers' willingness to pay for capacity. Capacity by its name is the ability to provide a certain amount of energy. It does not require the actual production of energy, and is different from an energy product. An immediate question is who wants the capacity product and how much they are willing to pay for it. Consumers need energy for heating, lights, and so forth, and they buy energy from energy markets. Energy has to be produced upon the existence of capacity. However, consumers are not directly interested in capacity. What they actually like to have besides the energy is the non-interruption of service, namely, reliable supply. To obtain such a quality of service, one can choose to sign a reliability contract with a service provider, have a certain amount of energy stored, or buy a back-up generator. Also, consumers may have different levels of reliability needs, and they should be able to specify that. It can be seen that reliability, instead of capacity, is indeed what consumers value. Traditionally, mostly due to political reasons, all energy consumers connected to the power grid share the same level of reliability. Certainly, some loads are less elastic than others, and some end-uses are more suited to adopting and responding to dynamic pricing than others. But the fact is that the VOLL for many end-use services, especially large new transport and heat loads, varies widely across time and location. Whilst the VOLL to interrupt lighting may be quite high, the VOLL to interrupt EV charging and shift it to a different time is usually very low. However, with the use of priority pricing contracts, consumers are able to specify their different levels of reliability needs and their willingness to pay for them. Priority pricing contracts incentivize residential implicit DR; then, an aggregator or utility aggregates all these contracts and participates in the capacity market in the form of explicit demand.

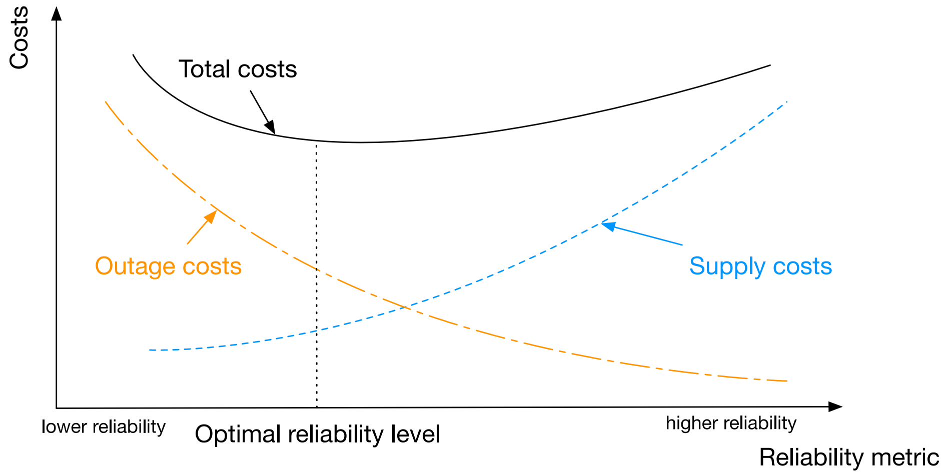
Constructing the capacity demand curve based on consumers' choices regarding their reliability levels has several benefits. As a result, the reliability level implied by the capacity demand curve requirements and the cost for the required capacity levels are optimal. Moreover, the bids of such costumers include the locational information which, as established in Section 4.3, is important, since capacities in different locations have different impacts on reliability, thus leading to a non-uniform VOLL. A modification of the demand curve based on customer choices is investigated in (Astier & Lambin 2019). Under this approach, customers who wish to avoid capacity payments would reduce their load during expected high-load hours, not limited to a small number of peak hours. Capacity costs would be assigned to utilities and by utilities to customers, based on actual load on the system during these hours. Customers wishing to avoid high energy prices would reduce their load during high-price hours. Customers would pay for what they actually use, as measured by meters, rather than relying on flawed measurement and verification methods. No measurement and verification estimates are required. No promises of future reductions, which can only be verified by inaccurate and biased measurement and verification methods, are required. Thus, by moving DR's participation to the demand side of the capacity market we avoid the use of baselines, the need for other customers to fund supply-side compensation, and distortions from over-compensating demand response at the full energy prices.

The submission of capacity demand bids by DR informs the TSO of the level of capacity participating, the months and hours of participation, and the price threshold at which load would be reduced. The TSO then would reduce the load forecast used in the capacity market based on the designated reductions. In traditional capacity demand curve construction, VOLL is considered to be uniform for all locations. Moreover, estimating VOLL is notoriously difficult and can cover a wide range of values. As such, both problems are solved by DR directly submitting demand bids in the capacity market. Furthermore, traditionally, capacity demand curves were designed to over-provide so that the risk of an interruption to supply is low. The TSO who constructs the capacity demand curve is not actually experiencing the risk of interruption. So decisions about risk have been taken without reference to consumers, while the price consequences have been passed on to them, since capacity markets' costs are passed on to consumers (Keay & Robinson 2019).

4.2 Capacity Demand Curve Construction with Priority Pricing Contracts

The economic basis for selecting a reliability criterion is depicted in Figure 10 where the optimal value of reliability corresponds to the least value of the supply and outage costs, namely, total costs. At this point the marginal cost of additional capacity is equal to the marginal benefit of additional capacity. Our control variable is the capacity installed, which we shall call C_N. As C_N grows the supply costs are higher and the outage costs are lower and vice versa (Gross 2012).

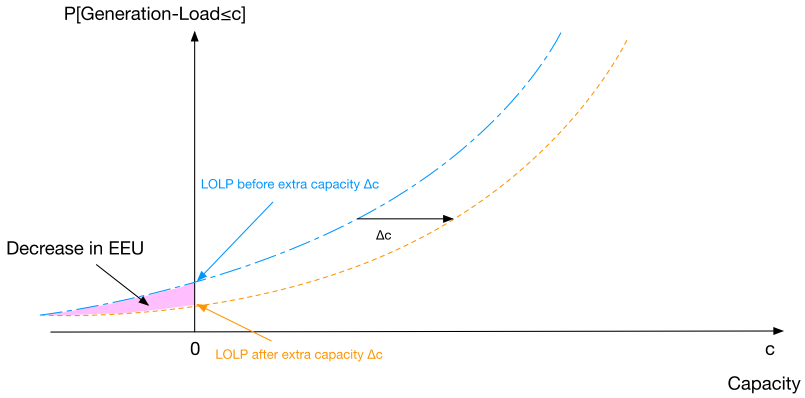
Figure 10: Optimal investment decision problem



Source: authors

Let us assume s to be the marginal cost of capacity. Let us denote the optimal capacity value $C_N^* = \sum_{i=1}^{N} C_i$, where N the number of generating units and C_i the capacity of the ith unit. We have that $s = \frac{\partial EEU}{\partial C_N} \cdot q$, where q is the marginal outage costs. The fraction $\frac{\partial EEU}{\partial C_N}$ demonstrates how the EEU changes with a change in capacity. An illustration of how EEU changes with capacity is demonstrated in Figure 11 with the shaded pink area with an addition in capacity of Δc. In the same figure the change of Loss of Load Probability (LOLP) is also depicted.

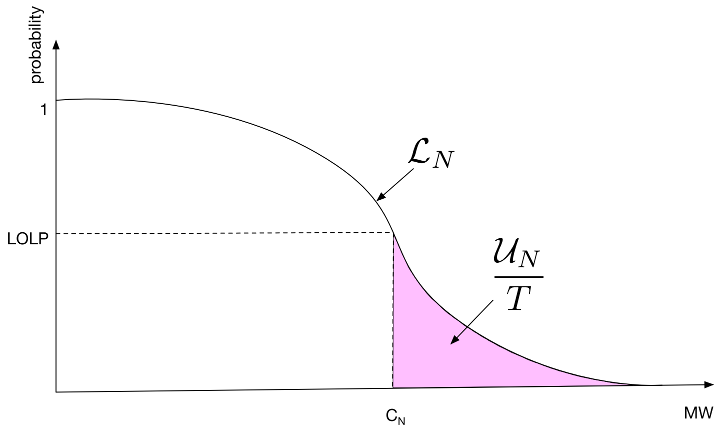
Figure 11: Modification in EEU with a capacity increase of Δc



Source: authors

In order to calculate the EEU changes with a change in capacity, i.e., $\frac{\partial EEU}{\partial C_N}$, we use the inverted equivalent load duration curve (IELDC) $\mathcal{L}_{\mathcal{N}}$ depicted in Figure 12. The IELDC gives the probability of the load supplied by the N generating units not being served, i.e., being greater than the available capacity. As such, the IELDC) $\mathcal{L}_{\mathcal{N}}(C_N)$ at capacity C_N is equal to the LOLP; thus it provides reliability information of the capability of units to serve the load. For a time period T, the expected energy unserved $\mathcal{U}_{\mathcal{N}} = T\int_{C_N}^{\infty} L_N(x)dx$; thus $\frac{\partial EEU}{\partial C_N} = T \cdot \mathcal{L}_{\mathcal{N}}(C_N) = T \cdot LOLP$.

Figure 12: Inverted equivalent load duration curve and expected energy unserved



Source: authors

Thus, if we go back to the marginal costs optimality conditions we have that

$$s = \frac{\partial EEU}{\partial C_N} \cdot q = T \cdot LOLP \cdot q = LOLE \cdot q,$$

where LOLE is the loss of load expectation. This is usually written as: Net CONE = LOLE · V OLL.

An alternative way of calculating the capacity demand curve is to use the VOLL that may be inferred by the priority pricing contracts, as explained in previous sections. In such a paradigm the value of reliability is directly used to derive capacity demand curves. Each consumer decides on their subscription to maximise their surplus, meaning the benefit minus the payment that needs to be submitted to the producer in order to secure this service. The benefit is a function of the VOLL, the levels of reliability, and the amount of energy purchased at each reliability level. This information is used to determine the VOLL of each costumer. For example, let us assume that consumers have access to the set of contracts given in Table 7.

Table 7: Priority pricing contracts with 70%, 99.7% and 100% reliability levels

Reliability [%]	Priority Charge [£/MWh]
70	21
99.7	154
100	157

Let us assume Consumer 1 purchases the following: 0.3kW of 100 per cent, 0.1kW of 99.7 per cent and 0.15 kW of 70 per cent. Consumer 2 purchases: 0.5kW of 100 per cent, 0.05kW of 99.7 per cent and 0 kW of 70 per cent. In this case $VOLL_{C_2} > VOLL_{C_1}$ since Consumer 2 chooses a higher portion of the highest reliability contracts. In other words, consumers with higher valuation select more reliable plans and pay more. By the consumers' selection of contracts and quantities their VOLL can be inferred. The relationship between contract prices and consumer valuation is given by:

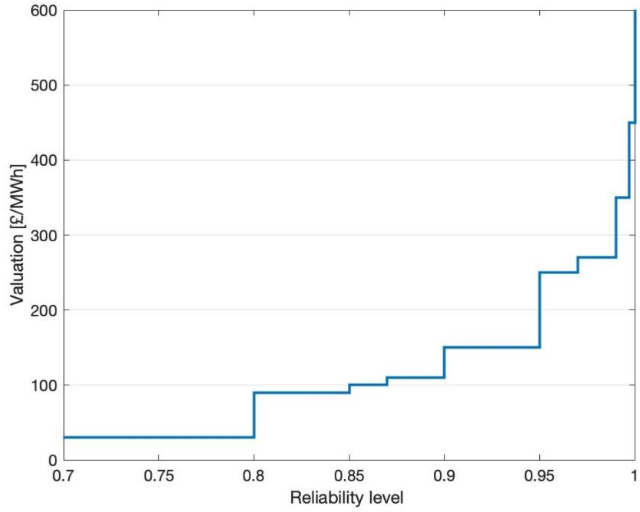
$$\mathrm{p_i = v_0 \cdot r_0 + \sum_{k=1}^{i} v_i(r_i - r_{i-1}),}$$

where p_i is the priority charge of contract with reliability level r_i and v_i is the valuation of consumer that buys a contract with the corresponding reliability level. In Table 8, we provide an example of priority pricing contracts with 10 reliability levels and plot in Figure 13 the valuation of consumers for each reliability level.

Table 8: Priority pricing contracts with 10 reliability levels

Reliability [%]	Priority Charge [£/MWh]
70	21
80	30
85	35
87	37
90	42
95	54
97	60
99	67
99.7	70
100	72

Figure 13: Reliability levels as a function of consumer valuation



When consumers select priority pricing contracts they demonstrate their valuation for each reliability level. By considering the reliability levels of Table 7 the consumer valuation may be divided into three components that correspond to the three reliability levels: $VOLL^{100\%}$, $VOLL^{99.7\%}$, and $VOLL^{70\%}$. In Table 9 we provide an example of priority pricing contracts for the four Type of consumers of Section 3.3.

Table 9: Consumer types and their valuation for reliability

Consumer Type						**1**	**2**	**3**	**4**
Priority Pricing Contract									
Reliability level	70%	Price [£/MWh]	21	Consumer valuation	30	0.1	0.1	0.15	0.15
	99.7 %		154		450	0.05	0.15	0.05	0.03
	100%		157		600	0.2	0.01	0	0
Average Consumer VOLL [£/MWh]						416	294	135	100

The average consumer VOLL for Type is calculated as follows:

$$\text{Average Consumer VOLL} = \frac{30\cdot 0.1+450\cdot 0.05+600\cdot 0.2}{0.1+0.05+0.2} = £416/MWh.$$

The others are calculated in a similar manner.

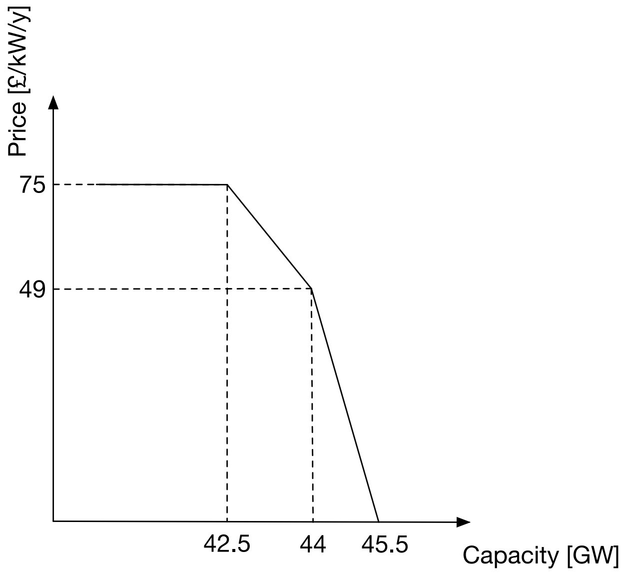
The three reliability levels of Table 7 correspond to different levels of loss of load expectations $LOLE^{100\%} = 3$h/y[10], $LOLE^{99.7\%} = 29$h/y and $LOLE^{70\%} = 2630$h/y. Each LOLE corresponds to a capacity level, for example, for the UK system $LOLE^{100\%}$ corresponds to 44 GW of capacity and $LOLE^{99.7\%}$ to 40 GW, respectively (ESO 2022). These values are calculated based on Monte Carlo simulation studies. Let us assume we have data for the priority pricing contracts of J customers. Then we can construct the capacity demand curve as follows: the value of capacity demand is $D^r = VOLL^r \cdot LOLE^r$, where r is the reliability reliability level and takes values in our example of 100 per cent, 99.7 per cent, and 70 per cent, in Table 10.

Table 10: Value of capacity demand for different reliability levels

Reliability Level (r) [%]	100	99.7	70
$LOLE^r$ [h/y]	3	29	2630
$VOLL^r$[£/MWh]	600	450	30
D^r [£/kw/y]	1.8	13.2	79
Percentage of Load [%]	51	28	21

We apply the aforementioned procedure in the recent T-4 capacity market for delivery in 2027-28 in the UK. The actual capacity demand curve used is depicted in Figure 14 based on the data available in the auction report (ESO 2024). The market cleared at £65/kW/y with a total of 43 GW capacity procured, less than the 44 GW target capacity. Indicatively, out of the 43 GW less than 3 per cent corresponds to DR resources. The Net CONE used as seen in Figure 14 was £49/kW/y with a reliability standard of 3 hour per year LOLE.

Figure 14: Capacity demand curve of T-4 capacity market for delivery in 2027-28



Source: authors

[10] The loss of load expectation of 100 per cent reliability is considered to be 3h/y instead of zero as per industry standards.

By using the relationship derived above, i.e., $Net\ CONE = LOLE \cdot VOLL$, we calculate the implied VOLL in the UK capacity market, which is $VOLL = \frac{£49/KW/y}{3h/y} = £16{,}333/MWh$. However, if we consider that residential consumers have indicated their VOLL with priority pricing contracts then we have a modified capacity demand curve. An example of such a curve is depicted in Figure 15. This example is based on the assumption that the UK power system comprises small and medium-sized enterprises (SMEs) and residential consumers whose respective VOLL are based on (London Economics, 2013) for Winter weekend peak (shown in Table 11). Under the assumption of 24 per cent and 76 per cent split between residential and SMEs the load-share weighted average across domestic and SME users for winter, peak, weekday is £16,333/MWh. This number coincides with the implied VOLL in the UK capacity market for delivery in 2027-28.

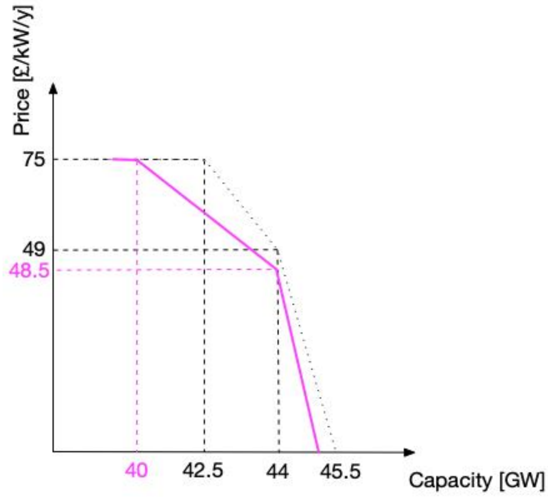
Table 11: Great Britain VOLL as a function of consumer groups

Consumer	VOLL [£/MWh]
Residential	10,289
SMEs	35,488

Source: London Economics, 2013

If we go back to the example of Table 9 there are consumers that are willing to pay less than what the TSO assumes, as shown in Table 11. In particular, the four Type of consumers have an average VOLL of 416, 294, 135, and 100 expressed in £/MWh, respectively (as seen in Table 9). If we assume that these consumers comprise 1.8 per cent of the residential population, then consumers might be willing to pay less that what the TSO assumes for desired capacity for LOLE of 3h/y, that is to say, £48.5/kW/y, compared with £49/kW/y[11]. Moreover, the minimum capacity assumed by the TSO is 42.5 GW. However, consumers are willing to experience reduced reliability levels as indicated by the priority pricing contracts which correspond to lower capacity levels. For example, $\mathrm{LOLE}^{99.7\%} = 29\mathrm{h/y}$ corresponds to 40 GW of capacity (ESO 2022). If more reliability levels are included in priority pricing contracts (as in Table 8) then the capacity demand curve would have more breakpoints corresponding to the different LOLEs.

Figure 15: Updated capacity demand curve of T-4 capacity market for delivery in 2027-28 taking into account VOLL of consumers



The capacity cost will be distributed to all consumers since the long-term security of supply benefits all consumers. Any other arrangement would create an evident situation of free-riding, because some users

[11] The new implied VOLL is $\frac{48.5£/kw/y}{3h/y} = 16{,}167£/MWh = 0.24 \cdot 35{,}488 + 0.742 \cdot 10{,}289 + 0.018 \cdot (30 + 450 + 600)$.

will be taking advantage of the system adequacy without paying the associated costs. Consumers that value capacity less, as indicated by the VOLL and capacity demand curves, benefit from the existing capacity but pay less for the energy they consume as well as they receive benefit for the capacity value of their demand response resources.

4.3 Refined Capacity Product for Enabling Demand Response

Drawing from the conclusions of Section 2.2 with respect to the shortcomings of capacity markets; a refined capacity product is defined as follows: Capacity product = {Location, Capacity profile, Flexibility}; we elaborate on each element below.

Location: In capacity market literature, reliability is often used interchangeably with resource adequacy. These two terms, however, are different. Resource adequacy is the ability of generation resources to meet the aggregated demand. The term originated from the conventional planning process of the vertically integrated industry. In such an environment, a utility company carries out the resource adequacy process on the aggregated system level without taking into account transmission, because any transmission security issues are handled in the subsequent transmission planning process. This can be done because the company owns both generation and transmission. On the other hand, *reliability* refers to a system's ability to serve electricity on a nearly continuous basis. Such a definition necessarily encapsulates both resource adequacy and transmission security. In summary, the capacity market should address the reliability requirement, which is a broader concept than just resource adequacy. The differentiation of these two concepts is important since the reliability objective justifies the need for the location attribute of capacity product.

The location attribute is captured through the definition of capacity zones by the TSO. Ideally, the location of a capacity resource should be specified at the more granular level to reflect the different reliability impacts of different locations. However, such granularity brings computational challenges to the solution of a capacity market. In addition, an aggregate DR resource may spread across several locations, but only reveals itself as a single capacity profile without specifying its actual distribution among buses. This led to the adoption of zonal models in some existing capacity markets where each zone is an aggregation of buses without considering internal transmission. In practice, the definition of zones is often a compromise of many factors, for example, geographic and regulatory boundaries, problem scalability, and similar electric features of buses in a zone. The introduction of capacity zones can be viewed as an approximation to the location attribute of a resource's reliability impact. For example, in PJM, locational constraints that represent transmission facility limitations or voltage limitations are included in their capacity market to quantify the locational value of capacity within their region.

Capacity Profile: Capacity profiles are probabilistic in nature. As a result, they are inherently difficult to measure and price. To overcome this difficulty, the TSO designates a deterministic capacity value to a resource to represent its capacity profile. Depending on the size, technology and maintenance of generators, resources may have differing abilities in providing a certain amount of power when called on, as in a 100 MW thermal resource with a 0.05 EFORd, or a wind resource with a varying output from 0 MW to 100 MW. For instance, a wind resource may be qualified at 50 MW based on its evenly distributed capacity profile between 0 and 100 MW. Note that such representation is only approximate since it does not distinguish the above resource from a thermal resource with a 50 MW unforced capacity. Such ability, or 'capacity profile,' obviously affects a resource's contribution to the system reliability. The use of the ELCC by many TSOs is promising. However, outage risks might not be uniform throughout the year so, for example, DR resources based on air conditioning load control are better supported by capacity obligations defined for shorter time intervals. In this regard, capacity market products that split into seasons, months and times of day would send more granular signals about when capacity is needed, and would also provide incentives for resources like DR to contribute when they are needed. The clearing of these individual capacity products would be co-optimized to ensure that the total capacity procurement is cost-effective while meeting reliability standards.

Flexibility: In present capacity market models, the flexibility requirement is not included explicitly, which leads to fewer incentives for flexible resource investment as flexible generation is generally more expensive than less flexible generation. In the UK, the underlying reason for redesigning the existing capacity market mechanism is to encourage investment in and the deployment of flexible resources. For instance, in February 2018, the capacity market auctions witnessed record low prices due to a very high participation by conventional generators, which were seeking additional revenue, while a very small amount of the more flexible resources was contracted. Although the clearing price of the capacity market was low, the need to incentivize flexibility was high.

Another potential element to be included in the capacity product definition is resource carbon intensity. However, it is not clear how this will interact with other subsidies offered to renewable resources, for instance, contracts for difference, and it is beyond the scope of this paper to review. Capacity markets have carbon emissions limits that resources must meet to enter the auction. Yet, there are certain gaps in carbon policies that allow, for example, a high percentage of capacity market agreements to be awarded to carbon-emitting gas generators smaller than 20MW that are exempt from the UK emissions trading scheme. There have been efforts to enforce new, lower emissions limits in the capacity market which will kick in for new plants from 1 October, 2034. This means all new oil and gas plants receiving long-term agreements through the capacity market will be obliged to lower emissions, through decarbonizing their capacity by introducing carbon capture, hydrogen and other low-carbon methods into their generation mix, and by reducing running hours (Department for Business & Strategy 2023). If the level of renewable energy support is defined centrally, for example in the form of feed-in tariffs or premiums, the aim is to ensure sufficient profitability for renewable energy investments. Any extra revenues will therefore result in windfall profits for such investments and an unnecessary burden on end-users. Reducing the level of remuneration for renewables in accordance with the amount of revenues received from capacity mechanisms, as in the French, Irish, and Italian policy mixes, appears to be a viable solution. In such cases, system reliability requirements can potentially determine the technological mix of renewables in the system (Kozlova & Overland 2022).

5. Aggregator Business Model

In this section, we discuss the relationship of the aggregator, providing priority pricing contracts to consumers and participating in capacity markets, with the consumers as well as with the other market entities. To this end, we present an analytical framework for a possible aggregator business model.

5.1 Aggregator Relationship with Consumers

With priority service, the aggregator can always provide the qualified capacity with a considerable level of certainty, since consumers engage in capacity subscription. A remaining risk for the aggregator is to match the offered reliability to the that of the priority service menu. Once a household is subscribed, a home energy router allocates particular devices within the house to different colours by ensuring that the mean power over time does not exceed the subscribed amount of kilowatts for that particular colour. In offering priority service contracts to residential consumers, the aggregator commits to a certain level of reliability for each service option. This level of reliability must be respected on average over an extended period (say, annually) even if certain periods may be characterised by fluctuations around this average. Another challenge for the aggregator is to design the menu of reliability and price pairs to offer. If reliability levels are priced too low, then all consumers would select the highest value reliability. This would be undesirable since the aggregator will not be able to provide such reliability at the selected price. If reliability levels are priced too high, then consumers will not enrol. Aggregate statistical information about the population is needed to design the menu to avoid the aforementioned problems (Mou et al. 2021). Then the priority and service charges can be tuned accordingly to various reliability levels.

We propose a business model and economic paradigm for a utility or aggregator that bridges the gap between wholesale markets and retail service. Aggregators incentivise participation of residential consumers through the priority service contracts and then participate in wholesale markets. In this paper, we show how they can participate in capacity markets – but a similar approach can be followed for energy and ancillary services markets. (Strauss and Oren 1993) propose the use of priority pricing contracts in the context of pooling them together with intermittent generation resources to transfer the risk from generators to load. In order to analyze the viability of such a business model we look at three aspects: value proposition, value creation and delivery and value capture (Hamwi et al. 2021).

In our case the value proposition is the capacity value of DR. For value creation a key mechanism is 'aggregation', which offers the opportunity for small-scale energy customers to exploit their potential. The isolated contribution of an individual residential DR to the power system is negligible, and the effects of small-sized consumers in the electricity market are inefficient. Therefore, using a mechanism that increases the efficiency of the resources is a prerequisite. The aggregator, as a market intermediary is responsible for many activities, such as participant registration and communication infrastructure, load-data transfer, data security, and participants' remunerations. In terms of resource availability, this is associated with behavioural patterns of residential consumers.

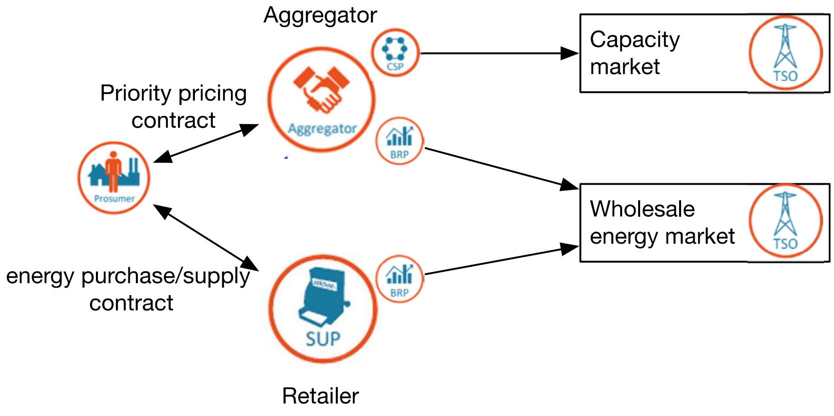
Value capture comprises cost structure and revenue model. The cost considered is mainly the initial participant cost. This consists of the activation cost, and includes the costs of investing in the enabling technology and establishing a response plan. This main cost can be categorised into two types: transaction cost, and intervention cost. The market transaction cost includes the costs of collecting information regarding products and customers, managing contracts, and procedures for external transactions. The DR transaction cost represents the costs of spending time identifying potential resources that can adjust their electricity consumption, understanding their electricity consumption patterns, assessing their suitability for participation, selecting the appropriate flexibility product, and evaluating the cost and benefits of each customer. These are affected by customers' preferences and the physical characteristics of devices deployed (as explained in Section 3.3) with the various preferences in priority pricing contracts and available resources. The information interaction between aggregators and customers is carried out in an automated manner through smart grid infrastructures, including control and communication devices, like smart meters with load control capabilities. The installation and maintenance of these devices are the obligation of aggregators. The transaction cost is high when the aggregator manages and aggregates a large number of customers. The intervention cost is related to behavioural adaptation, and is based on the fact that consumers are traditionally unpredictable, and are not used to dynamically and temporally adapting their consumption processes. The main economic challenge in DR operations is in generating a sufficient income to cover the expenses. In DR, the captured value is shared by both the aggregator and customers. The aggregator generates revenue, and customers receive remuneration for their participation. Aggregators provide financial incentives to customers to encourage them to actively participate in DR. Economic incentives come in many forms, such as extra compensation or discounts on electricity charges. This revenue model is based on electricity bill savings and capacity payment, or payment reductions as is the case of implicit demand response.

5.2 Aggregator Relationship with other Market Entities

Retailers are independent entities that, besides bridging the gap between the wholesale market and customers, are responsible for load forecasting; risk assessment from load and price fluctuation; and designing tariffs (Lu et al. 2020). Retailers are the best candidates to be aggregators due to their existing strong interaction with consumers as they have energy supply contracts with them. However, retailers are not responsible for the installation of control and communication infrastructures for the customers. In the current power system under EU regulations, the retailer acquires the amount of energy the consumer is expected to consume and informs the TSO about the consumer's projected consumption for the coming hours or next day. This is usually referred to as 'scheduled demand' or the 'firm program'. Moreover, the retailer takes responsibility for imbalances between the retailer's forecast of the flexible consumer's demand and the actual consumption that occurs in real time with its balancing responsible party (BRP). When the aggregator is an independent entity it has its own BRP. The problem in the case of independent aggregator is that the forecasted load and purchased energy by the retailer will be different due to the priority pricing contracts offered to the prosumers by the aggregator. This might create retail market distortion and financial penalties to retailers if not addressed appropriately. More specifically, the retailer bears the full risk of any imbalances in its supply portfolio. In addition to this, the consumer's consumption falls short of the amount of energy procured by the retailer. In the absence of any further arrangements, the retailer may not invoice this difference and thus, may not recover the full electricity procurement cost. In this regard, we propose as a possible solution using the separated win-win model as discussed in (Alba et al. 2021). Just to note there is extended literature on how independent aggregators can participate in wholesale markets and their responsibilities to other market participants (see Carreiro et al. 2017, He et al. 2013). In the business model where the aggregator is a different entity to the retailer, the consumer has two contracts, one with his retailer for supplying energy, and one with his aggregator for the priority pricing contracts. Aggregator and retailer are different companies, each with different balancing responsibilities. In the win-win model, the balancing responsibility is transferred from the retailer to the independent aggregator before day-ahead gate closure. Thus, the independent aggregator takes responsibility for its selected consumers and for their imbalances over the next 24 hours, allowing it to adjust the portfolio in all energy markets. Before day-ahead gate closure, the independent aggregator communicates to the TSO the aggregated load curve of its customers who at the same time have an energy contract with any retailer. The independent aggregator has its own BRP and is responsible for the imbalance of the program communicated to the TSO. The retailer purchases the energy of all its customers in the day-ahead market. The retailer remains responsible for procuring the energy for all customers, including those that will be managed by an independent aggregator. The TSO transfers the aggregated load profile previously defined by the aggregator at the day-ahead market price. This transfer is not a real one, but an accounting one for the upcoming settlement through which the independent aggregator pays for the energy at the market price before selling it in the markets. The retailer, therefore, is already compensated for such energy. The aggregator offers demand services in the other markets where it detects an opportunity. The aggregator receives the income for the services provided to the system. Measurement deviations from its program are settled by the aggregator BRP. Since the retailer has to bill consumers, the revenue for the energy consumed, as measured by smart meters each hour, is transferred by the aggregator to the retailer at the day-head market price, in addition to the average cost of imbalances incurred by the retailer for the energy sold to customers – for which the aggregator is responsible. This addition is important since retailers usually internalize in their offers any imbalances their customers may have. A retailer's customers, therefore, pay an amount of money to their retailer for system imbalances. In this model, the aggregator is responsible for the customer's imbalances and therefore they must pay the system for such imbalances. However, since the retailer is billing such customers, retailers (instead of aggregators who bear the cost) are compensated for customer's imbalances. These benefits received by the retailer should somehow be transferred to the aggregator since they are the agents assuming the whole imbalances of these customers. The aggregator only buys the energy in those hours where it activates the demand management service. However, it is responsible for the imbalances for the DR customers in other hours, solving the rebound effect and, thus, ensuring that imbalances do not penalize retailers. In this model, each agent is responsible for his own participation in the market. On the one hand, the retailer takes responsibility for the imbalances between the flexible consumer's forecast and actual consumption that may occur in real-time only for those customers who are not also managed by an independent aggregator.

On the other hand, the independent aggregator is responsible for the imbalances between the flexible consumer's forecast and actual consumption only for those customers previously communicated to the TSO. This procedure is depicted in Figure 17.

Figure 17: Aggregator and market entities interaction model

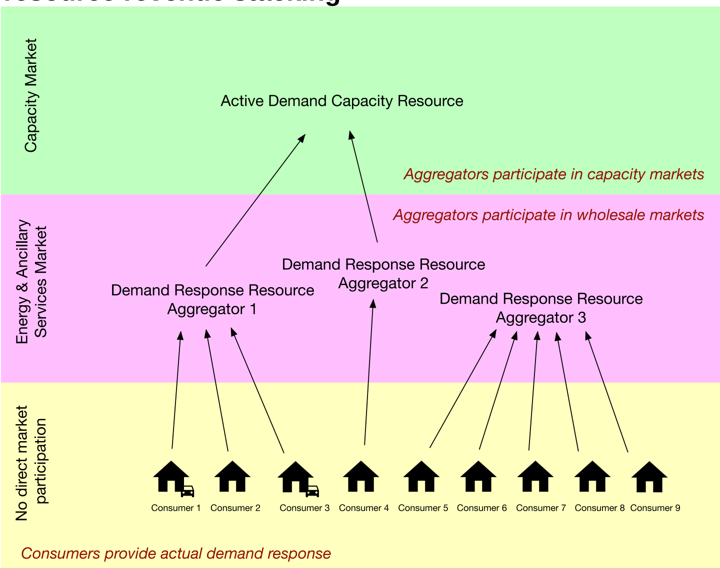


Source: authors

As explained above, the TSO acts as the buyer of capacity in the form of DR offered by the aggregators, whose operation rationality needs to be verified by the TSO, including the activation, control and scheduling scheme. The verification can be more competently done by the distribution system operator (DSO), who has direct access to the authentication demand from customers within its jurisdiction area. The DSO operates the network to ensure the efficient and reliable power delivery to customers. It has access to the consumption data of the customers measured by the advanced metering infrastructure, which could also be fetched by the aggregators who will utilise it for settlement. In this regard, aggregators do not need to invest in separate communication infrastructure and could use the one belonging to the DSO. As such, aggregator and DSOs are linked in this manner.

Using the framework presented here an aggregator does not have to limit its activity to the capacity market but can stack its revenues by participating in energy and ancillary services markets – see Figure 18.

Figure 18: Demand resource revenue stacking



Source: authors

In revenue stacking there are still some challenges, for example, DR is not allowed to participate in wholesale energy markets in the UK (this will change at the end of 2024 with the adoption of new rules). It is also interesting to investigate the relative importance of DR in the various markets from an aggregator's perspective as well as from a system perspective, and from the view that some DR products may be more suitable for capacity markets, and others for energy and ancillary markets.

6. Concluding Remarks

Energy-only markets rely on high price signals during periods of scarcity to incentivize investment. However, these signals can be uncertain and insufficient to cover the fixed costs of new generation. In this regard, capacity remuneration mechanisms have been introduced in many countries to compensate generators for making generation capacity available for utilization. In this paper, we focused on capacity markets as a form of remuneration mechanism and analyzed their shortcomings. This was principally that a central agency decides on the reliability level that needs to be met by the capacity cleared in capacity markets. However, the consumers are the ones that are actually experiencing outage risk. Moreover, we demonstrated that the current capacity market regulatory framework is more tailored for conventional generating resources; thus, leaves less room for storage and demand response resources. However, instead of building new generating units to meet reliability criteria, demand response resources could be used. During times of generation scarcity there exist consumers who are willing to reduce their demand for monetary compensation. Although commercial demand response has been more prominent, there has not been great engagement from residential demand response.

In this regard, we discussed the benefits of demand response in wholesale markets; in terms of social welfare maximisation, economic use of interconnection, reductions in generation capacity requirements, transmission and distribution network congestion management, and increased economic efficiency. Next, we analysed the barriers that prohibit greater demand response participation. These span from economic, social, technological, political, and structural. The main barriers include the market product traded and consumers' perception. Residential consumers tend to withdraw from demand response programs if the inconvenience of participating becomes too high. However, consumers may attach different value to the guarantee of continuity of supply. This is especially the case when considering consumption for non- essential uses in excess of what is necessary to cover basic needs (lighting, cooking and heating). In this regard, we proposed in this paper the use of priority pricing contracts so that consumers can demonstrate their value for reliability by subscribing in specific contracts. We analysed the use of priority pricing contracts under different consumer types and how much each one can benefit from such contracts. We compared priority pricing contracts to other implicit demand response contracts and showed that their benefits include: privacy, control over consumption, simplicity and modularity. Thus it might be able to address some of the barriers faced by residential demand response in terms of privacy concerns, consumer engagement, product definition, and complexity. We also provided guidelines on the actions that need to be taken so that consumers are more receptive to priority pricing contracts through education campaigns and bill protection for sensitive or disadvantaged consumers.

We then proposed a framework to incorporate priority pricing contracts through an aggregator as explicit demand response in capacity markets. We argued that the aggregator should participate through the 'demand' instead of the 'supply' side in the capacity market to avoid market distortion and be able to reflect desired consumers' reliability levels. In order to input priority pricing contracts in the capacity demand curve we studied the relationship of capacity and reliability. Under optimal conditions the marginal cost of outages is equal to the marginal cost of capacity. This can be translated to a relationship connecting capacity with the value of lost load and the loss of load expectations. The information that can be inferred by priority pricing contracts, in terms of value of lost load and loss of load expectations can be incorporated in the construction of the capacity demand curve. We illustrated the proposed framework in the 2027-28 delivery UK capacity market. Moreover, we proposed a refined capacity product to be traded in capacity markets so that it incentivizes investments in demand response resources. The refined product included information about location, flexibility and modification of the firm capacity profile calculation.

Last, we proposed a business model for an aggregator that provides priority pricing contracts to consumers and participates in capacity markets. We analysed their relationship with consumers and how they can build their value proposition, value creation and delivery and value capture. We also studied the aggregator relationship with other market entities. We focused on their relationship with retailers and what information-responsibilities need to be exchanged so that penalties due to inconsistencies in actual and projected load are avoided.

References

ACER (2018), 'Annual report on the results of monitoring the internal electricity and natural gas markets in 2017 - electricity wholesale markets' volume, Technical report.

ACER (2023*a*), 'Demand response and other distributed energy resources: what barriers are holding them back?', Technical report.

ACER (2023*b*), 'Report on electricity transmission and distribution tariff methodologies in Europe', Technical report.

Alba, J. J., Vereda, C., Barquín, J. & Moreda, E. (2021), – 'Market design and regulation to encourage demand aggregation and participation in European energy markets', Chapter 17 *in* F. Sioshansi, ed., *Variable Generation, Flexible Demand*, Academic Press, 393–410.

Monitoring Analytics (2023), 'PJM state of the market – 2023', Technical report.
https://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2023.shtml

Astier, N. & Lambin, X. (2019), 'Ensuring capacity adequacy in liberalised electricity markets', *The Energy Journal* 40(3), 227–242.

Billimoria, F. & Poudineh, R. (2019), 'Market design for resource adequacy: A reliability insurance overlay on energy-only electricity markets', *Utilities Policy* 60, 100935.
www.sciencedirect.com/science/article/pii/S0957178719300177

Borenstein, S., Jaske, M. & Rosenfeld, A. (2002), 'Dynamic pricing, advanced metering, and demand response in electricity markets', Technical report, Center for the Study of Energy Markets, UC Berkeley. https://escholarship.org/uc/item/11w8d6m4

Bothwell, C. & Hobbs, B. F. (2017), 'Crediting wind and solar renewables in electricity capacity markets: The effects of alternative definitions upon market efficiency', *The Energy Journal* 38, 173–188.

Brown, T., Newell, S. A., Oates, D. L. & Spees, K. (2015), 'International review of demand response mechanisms', Technical report, Australian Energy Market Commission.

Brown, T., Newell, S., Spees, K. & Wang, C. (2019), 'International review of demand response mechanisms in wholesale markets', Technical report, The Brattle Group.

Bublitz, A., Keles, D., Zimmermann, F., Fraunholz, C. & Fichtner, W. (2019), 'A survey on electricity market design: Insights from theory and real-world implementations of capacity remuneration mechanisms', *Energy Economics* 80, 1059–1078.

CAISO (2016), 'What the duck curve tells us about managing a green grid?', Technical report.

Cardoso, C. A., Torriti, J. & Lorincz, M. (2020), 'Making demand side response happen: A review of barriers in commercial and public organisations', *Energy Research & Social Science* 64, 101443.

Carreiro A. M., Jorge H. M., Antunes C. H. (2017). 'Energy management systems aggregators: a literature survey',*Renew Sustain Energy Rev* 2017;73:1160e72.

Chao, H.-P. & Wilson, R. (1987), 'Priority service: Pricing, investment, and market organization', *The American Economic Review* 77(5), 899–916. www.jstor.org/stable/1810216

Cramton, P. & Stoft, S. (2005), 'A capacity market that makes sense', *The Electricity Journal* 18(7), 43–54.

Dent, C., Keane, A. & Bialek, J. (2010), 'Simplified methods for renewable generation capacity credit calculation: A critical review', *IEEE PES General Meeting*, 1–8.

Department for Business, E. & Strategy, I. (2023), Capacity market 2023 consultation: Strengthening security of supply and alignment with net zero, Technical report.

DESNZ (2023), 'Energy trends UK, April to June 2023', Technical report.

Dupuy, M. & Linvill, C. (2019), 'Implementing demand response 2.0: Progress toward full potential in the United States', *The Electricity Journal* 32(7), 106622. Special Issue: Energy Optimization is the Key to Affordable, Reliable Decarbonization.

European Commission (EC) (2013), 'Generation adequacy in the internal electricity market - guidance on public interventions', Technical report.

ENTSO-E (2015), 'Market design for demand side response', ESO (2023*a*), *Future energy scenarios*, Technical report.

ESO, National Grid (2021), 'Capacity market: 2021 consultation on improvements', Technical report.

ESO, National Grid (2022), 'National grid ESO electricity capacity report', Technical report.

ESO, National Grid (2023*a*), 'Operability Strategy Report', Technical report.

ESO, National Grid (2023*b*), 'Final auction report 2022 four year ahead capacity auction (t-4) delivery year 2026/27', Technical report.

ESO, National Grid (2024), 'Provisional auction report 2023 four-year ahead capacity auction (t-4) delivery year 2027/28', Technical report.

EU (2012), 'Directive 2012/27/EU of the European Parliament and of the council', Technical report.

Faruqui, A. & Tang, Z. (2023), 'Time-varying rates are moving from the periphery to the mainstream of electricity pricing for residential customers in the United States', Technical report, The Brattle Group.

FERC (2021), 'FERC order no. 2222: Facilitating participation in electricity markets by distributed energy resources', Technical report.

Gérard, C. & Papavasiliou, A. (2022), 'The role of service charges in the application of priority service pricing', *Energy Systems* 13(4), 1099–1128.

Good, N., Ellis, K. A. & Mancarella, P. (2017), 'Review and classification of barriers and enablers of demand response in the smart grid', *Renewable and Sustainable Energy Reviews* 72, 57–72.

Gross, G. (2012), Electricity resource planning, Technical report, University of Illinois at Urbana- Champaign.

Hamwi, M., Lizarralde, I. & Legardeur, J. (2021), 'Demand response business model canvas: A tool for flexibility creation in the electricity markets', *Journal of Cleaner Production* 282, 124539.

He, X., Keyaerts, N., Azevedo, I., Meeus, L., Hancher, L. & Glachant, J.-M. (2013), 'How to engage consumers in demand response: A contract perspective', *Utilities Policy* 27, 108–122.

Hogan, W. W. (2013), 'Electricity scarcity pricing through operating reserves', *Economics of Energy & Environmental Policy* Volume 2 (Number 2) .

Kahma, N. & Matschoss, K. (2017), 'The rejection of innovations? Rethinking technology diffusion and the non-use of smart energy services in Finland', *Energy Research & Social Science* 34, 27–36. www.sciencedirect.com/science/article/pii/S2214629617301445

Keay, M. & Robinson, D. (2019), 'The limits of auctions: reflections on the role of central purchaser auctions for long-term commitments in electricity systems', Technical report, Oxford Institute for Energy Studies.

Kim, J.-H. & Shcherbakova, A. (2011), 'Common failures of demand response', *Energy* 36(2), 873–880.

Kozlova, M. & Overland, I. (2022), 'Combining capacity mechanisms and renewable energy support: A review of the international experience', *Renewable and Sustainable Energy Reviews* 155, 111878.

Lambin, X. (2020), 'Integration of demand response in electricity market capacity mechanisms', *Utilities Policy* 64, 101033. www.sciencedirect.com/science/article/pii/S095717872030028X

Liu, Y. (2017), 'Demand response and energy efficiency in the capacity resource procurement: Case studies of forward capacity markets in ISO New England, PJM and Great Britain', *Energy Policy* 100, 271– 282.

London Economics (2013), 'The Value of Lost Load (VOLL) for Electricity in Great Britain: Final Report for Ofgem and DECC', July 2013, 1-225.

Lu, X., Li, K., Xu, H., Wang, F., Zhou, Z. & Zhang, Y. (2020), 'Fundamentals and business model for resource aggregator of demand response in electricity markets', *Energy* 204, 117885. www.sciencedirect.com/science/article/pii/S0360544220309920

Mathieu, J. L., Kamgarpour, M., Lygeros, J. & Callaway, D. S. (2013), 'Energy arbitrage with thermo- statically controlled loads', *2013 Control Conference (ECC)*, 2519–2526.

Mays, J., Morton, D. P. & O'Neill, R. P. (2019), 'Asymmetric risk and fuel neutrality in electricity capacity markets', *Nature Energy* 4(11), 948–956.

Mou, Y., Gerard, C., Papavasiliou, A. & Chevalier, P. (2021), 'Designing menus for multilevel demand subscription', *HICSS*.

Murthy Balijepalli, V. S. K., Pradhan, V., Khaparde, S. A. & Shereef, R. M. (2011), 'Review of demand response under smart grid paradigm', *ISGT2011-India*, 236–243.

Neuhoff, K. & De Vries, L. (2004), 'Insufficient incentives for investment in electricity generations', *Utilities Policy* 12(4), 253–267. Infrastructure Regulation and Investment for the Long-Term. www.sciencedirect.com/science/article/pii/S0957178704000682

O'Connell, N., Pinson, P., Madsen, H. & O'Malley, M. (2014), 'Benefits and challenges of electrical demand response: A critical review', *Renewable and Sustainable Energy Reviews* 39, 686–699.

Ofgem (2024a), Ten-year review of the capacity market rules, Technical report.

Ofgem (2024b), Energy and data research. Chart.

Oren, S. S. (2005), 'Generation adequacy via call options obligations: Safe passage to the promised land', *The Electricity Journal* 18(9), 28–42.

Papalexopoulos, A., Beal, J. & Florek, S. (2013), 'Precise mass-market energy demand management through stochastic distributed computing', *IEEE Transactions on Smart Grid* 4(4), 2017–2027.

Papavasiliou, A. (2020), 'Scarcity pricing and the missing European market for real-time reserve capacity', *The Electricity Journal* 33(10), 106863. www.sciencedirect.com/science/article/pii/S104061902030155X

Picciariello, A., Vergara, C., Reneses, J., Frías, P. & Söder L. (2015), 'Electricity distribution tariffs and distributed generation: Quantifying cross-subsidies from consumers to prosumers', *Utilities Policy* 37, 23–33.

Richter, L.-L. & Pollitt, M. G. (2018), 'Which smart electricity service contracts will consumers accept? The demand for compensation in a platform market', *Energy Economics* 72, 436–450.

Strauss, T. & Oren, S. (1993), 'Priority pricing of interruptible electric service with an early notification option', *The Energy Journal* 14(2), 175–196.

Strbac, G. (2008), ‘Demand side management: Benefits and challenges’, *Energy Policy* 36(12), 4419–4426. US (2005), Energy policy act of 2005, Technical report.

Warren, P. (2014), ‘A review of demand-side management policy in the UK’, *Renewable and Sustainable Energy Reviews* 29, 941–951.