Scarcity Pricing and the Missing European Market for Real-Time Reserve Capacity

Anthony Papavasiliou

Center for Operations Research and Econometrics

Université catholique de Louvain

Voie du Roman Pays 34, Office b.114

B-1348 Louvain-la-Neuve

Belgium

Contents

1. Introduction

The electric power industry is undergoing a radical transformation as we are increasingly replacing conventional and controllable thermal technologies by renewable energy resources. Renewable resources are characterized by low marginal costs, limited controllability, and a large degree of short-term forecast uncertainty. This implies that the value of the electricity market is shifting from energy to reserve services, and that the role of real-time operations in reflecting the level of scarcity in the system is becoming increasingly important.

In response to this evolution, the European balancing market is adapting. European transmission system operators and market operators are increasing their coordination by putting in place cross-border markets for trading energy[1], transmission capacity, and reserves. However, although the trading of *energy* is becoming increasingly coordinated near real time in Europe, the trading of *reserve capacity* in real time is simply inexistent. This is a fundamental shortcoming in a system where renewable resources are shifting value from energy to reserves.

This paper argues for a set of modifications to the pricing of energy and reserves that aims at implementing a European real-time market for reserve capacity. The goal of the proposal is to put a mechanism in place that can value the availability of reserve capacity in a way that is reflective of system scarcity, while remaining compatible with European legislation. The mechanism aims at respecting the interaction between the value of energy and reserve capacity.

The proposal that we put forth in this paper is the outcome of a series of market analysis and design studies that commenced in 2014, in collaboration with the Belgian regulatory authority for electricity and gas (CREG) and the Belgian transmission system operator (ELIA) (Papavasiliou & Smeers, 2017) (Papavasiliou, et al., 2018), (Papavasiliou, et al., 2019). The paper thus provides an account of the evolutions that have taken place towards implementing scarcity pricing since the commencement of the analysis, the status of the discussion, and a collection of considerations that are emerging as we are moving forward.

Belgium provides an interesting context for the analysis of scarcity pricing, due to the ongoing debate regarding the implementation of scarcity pricing versus a capacity remuneration mechanism. Although this paper is not concerned with capacity markets, it is important to state that the dilemma between implementing scarcity pricing or capacity markets is false. Scarcity pricing does not preclude the implementation of a capacity remuneration mechanism, and can co-exist with a capacity market. On the other hand, precedence is relevant: before rejecting the possibility that scarcity pricing can mobilize investment, it is important to give it a chance to function properly. The current European real-time market is exclusively an *energy* balancing market, and it does not trade reserve in real time. Consequently, the existing balancing market design seems to be seriously handicapped in terms of supporting scarcity pricing.

[1] The launch of integrated market clearing platforms provides evidence of this evolution. These include the platforms for the activation of replacement reserve (Trans European Replacement Reserves Exchange, abbreviated "TERRE"), tertiary reserve (Manual Activated Reserves Initiative, abbreviated "MARI") (ENTSO-E, 2017), secondary reserve (Platform for the International Coordination of Automated Frequency Restoration and Stable System Operation, abbreviated "PICASSO"), and primary control (The International Grid Control Cooperation, abbreviated "IGCC").

2. European Balancing Markets

Two fundamental attributes of the US Standard Market Design are coordination and the anchoring of forward markets against real-time prices. Coordination allows markets to trade multiple products at prices that are consistent with the complex engineering constraints[2] that govern the interdependencies between these products. One fundamental such constraint stipulates that the total amount of capacity of a generator places a limit on the *joint* provision of energy and reserve capacity. This implies an equilibrium relation that couples energy and reserve prices: for a resource that splits its capacity between the two markets, profit margins in these markets should be equal. This implies that energy and reserve prices follow each other in lock step, with the marginal cost of the generator that is offering in both markets separating the two. The multi-settlement design, on the other hand, places the real-time market at the center stage of the electricity market. By settling real-time deviations (of energy and reserve capacity respectively) against the real-time price (of energy and reserve capacity respectively), the electricity market allows agents to hedge risks through forward (e.g. day-ahead) trades while maintaining their incentive to deploy their resources efficiently in real time. Given the fact that the integration of renewable resources in power systems is introducing an enormous degree of uncertainty in real-time operations (including uncertainty in flow patterns and the availability of excess reserve capacity), the role of the real-time market has become central in forming accurate price signals. These price signals ideally communicate the need for investing in resources with specific capabilities (e.g. the ability to respond rapidly) at the right location.

The European market differs from this design in two fundamental respects: the trading of energy and reserve is not co-optimized, and there is a weak consistency between day-ahead and real-time market design. Both of these elements complicate price formation. In some countries, reserves are auctioned before the day-ahead energy market (e.g. Belgium, Germany, the Netherlands), whereas in other countries reserves are auctioned after the day-ahead market (e.g. Spain, Italy). Therefore, the aforementioned equilibrium between day-ahead energy and reserve prices relies on the ability of asset owners to anticipate the opportunity cost of trading their generation capacity in one auction versus the other. The weak consistency between day-ahead and real-time markets implies a host of challenges. The one that is of interest in the present paper is a fundamental difficulty for valuing reserve accurately, which undermines an effective implementation of scarcity pricing, as we argue in the sequel.

The functional segmentation between energy and reserve markets in Europe is exemplified best by the separate consideration of so-called "balancing service providers" (abbreviated hereafter as BSPs) and "balancing responsible parties" (abbreviated hereafter as BRPs). BSPs are essentially resources that can offer reserves, and BRPs are assets that may consume or produce power, but are not able to offer reserve services. In the day-ahead market, BSPs participate in reserve auctions that are organized by national TSOs. By selling a certain amount of capacity to these reserve auctions, BSPs commit to bid this capacity into the balancing market. The day-ahead reserve auctions are operated by TSOs, who act as buyers of the reserve capacity. In real time, BSPs are activated upward or downward by TSOs in balancing auctions, in order to cover BRP imbalances. BRP imbalances correspond to deviations between BRP forward energy trades and their real-time net consumption. The essential economic difference between BSPs and BRPs, therefore, is that the former function as price-elastic suppliers/consumers of energy in case of upward/downward activation, whereas the latter function as price-inelastic consumers/suppliers, in case of positive/negative imbalance. Each

[2] We ignore transmission capacity throughout the paper.

BSP must be attributed to a unique BRP portfolio, as required in article 18(4).d of the European Balancing Guideline (European Commission, 2017), which we refer to hereafter as the EBGL.

There are various functional details that characterize the European balancing market, and which are relevant for the implementation of scarcity pricing. The balancing market interval is 15 minutes. In future balancing platforms, BSPs will be paid a uniform clearing price. Each balancing platform will be run separately, and will produce a different balancing price, at a different time step. For example, the MARI platform for activating tertiary balancing energy, which will go live in December 2021, is planned to clear every 15 minutes. Instead, the PICASSO platform for activating secondary balancing energy, which will go live in December 2021, is planned to clear every few seconds. Free bids are allowed to participate in the balancing auction, as eligible reserve capacity that has not been committed in forward reserve markets.

From an engineering standpoint, the differences between BSPs and BRPs are essential: the TSO can rely on the former for reserve, whereas it cannot rely on the latter. And there are numerous significant functional distinctions between different types of reserves, in the European system, as in any system. However, the functional separation of BSPs and BRPs does not negate the fact that the two are essentially trading energy in the balancing market. The functional distinction between BSPs and BRPs has unfortunately led to two notable distortions in European market design. On the one hand, BRPs and BSPs face a different real-time price for trading energy, with BSPs being settled at a co-called balancing price, and BRPs being settled at a so-called imbalance price. On the other hand, whereas Europe operates a real-time energy market (the so-called balancing market, which settles *energy* imbalances), there is no provision for a real-time reserve market (i.e. there is no settlement of *reserve* imbalances). The question arises: what is the logic of settling energy imbalances, but not doing the same for reserve imbalances?

3. Scarcity Pricing

3.1 Scarcity Pricing Based on ORDC

Scarcity pricing refers to the notion of increasing energy prices above the marginal cost of the marginal unit under conditions where the system is short on generation capacity. Scarcity pricing generates profits for generating resources that serve towards covering the capital costs of these units. Scarcity pricing is therefore essential for attracting investment in a market.

In a market with price-responsive demand, scarcity pricing is an automatic consequence of demand-side price elasticity. Concretely, during periods of scarcity in generation capacity, it is the valuation of demand that sets the equilibrium price for energy. Whenever this valuation exceeds the marginal cost of the most costly unit in the system, equilibrium prices create profits for all technologies in the system, including the most expensive ones, thereby supporting investment. In a market without price-responsive demand, the market clearing price is determined administratively, and acts as a proxy for the average valuation of the consumers who are being curtailed involuntarily. Involuntary load shedding can be a relatively infrequent occurrence, and can therefore lead to unpredictable profit margins that make it especially risky to invest in peaking technologies. Since peaking technologies often coincide with those technologies that are capable of offering reserves to the system (for example, combined cycle gas turbines (CCGTs)), the profit margins generated from selling energy alone can become highly unpredictable. This effect is exacerbated in a regime of large-scale renewable integration, where these margins from the energy market are squeezed due to the near-zero marginal cost of renewable technologies. Ironically, it is exactly the integration of renewable resources that also increases the need for reserves that can be provided by technologies with high marginal costs, such as CCGTs. Consequently, therefore, we have a counter-intuitive effect whereby the more renewables we integrate into the system, the more we need technologies such as CCGTs, but the more slim and unpredictable the profits of these technologies become from selling *energy*. This is not surprising: integrating renewable resources shifts the value of the market from energy to reserve services. Therefore, the design of the market needs to respond to this evolution.

The goal of operating reserve demand curves (Hogan, 2005) (Hogan, 2013) is to respond to this evolution by introducing elasticity to the procurement of reserve capacity. Traditionally, reserve capacity has been procured in electricity markets through fixed reserve requirements. These requirements may vary daily. For example, the requirement for tertiary reserve capacity in Belgium is adapted to the degree of uncertainty in the system, a practice which is referred to as dynamic reserve dimensioning. Dynamic reserve dimensioning was launched in Belgium in February of 2020 (De-Vos, et al., 2019), (Federal Public Service Economy, 2019). Nevertheless, even if dynamic, a fixed reserve requirement corresponds effectively to price-inelastic demand for reserve capacity. Due to the lack of price elasticity, fixed reserve requirements inherit the undesirable properties of an inelastic demand for energy that were mentioned previously, namely relatively unpredictable price spikes.

Operating reserve demand curves introduce price elasticity. The valuation of operating reserve demand curves is linked to the amount of reserve capacity that is available in the system. Concretely, the valuation as a function of reserve capacity that has been proposed in (Hogan, 2013) can be stated as follows for a unique reserve product:

$$V^R(R) = \left(VOLL - \widehat{MC}\right) \cdot LOLP(R). \tag{1}$$

for $R \geq X$, where X is a minimum threshold of reserve that the system should carry before resorting to involuntary load shedding. In this expression, $VOLL$ corresponds to an estimate of the value of lost load, $\widehat{MC}$ corresponds to a proxy of the system lambda for the balancing interval in question, R is the amount of available reserve capacity currently available in the system, and $LOLP$ is the loss of load probability given the available amount of reserve capacity, R.

There is a formal justification for this specific choice of demand function (Hogan, 2013), which links the valuation of reserve to the loss of load probability. It is worth noting that the connection between scarcity pricing and loss of load probability is a desideratum that has been expressed explicitly in the guidelines of the European Commission for the implementation of scarcity pricing in Belgium (European Commission, 2020). The intuition behind this specific expression for a demand function is that the incremental value of reserve capacity diminishes, as the amount of reserve capacity headroom that is available in the system increases. When the system is tight, increments of reserve capacity are valuable, because they significantly reduce the loss of load probability. When the system is comfortable, increments of reserve capacity have a negligible value.

Recall that reserve and energy equilibrium prices are coupled by the fact that reserve providers who find themselves splitting their limited capacity between both markets should be indifferent between the profit margin of the energy market and the market for reserve. Operating reserve demand curves introduce price elasticity in the market for reserve capacity. Since the price of energy follows the price of reserve capacity in lock step, the price elasticity of the operating reserve demand curve that reduces the volatility of reserve prices can result in energy prices that are also less volatile.

Concretely, in a market with ORDC, one expects to see more frequent but less violent energy price spikes, which are conducive towards a favorable economic environment for investing in resources that can offer reserve services to the system. This is the answer that ORDC offers to the shift of value from energy to reserve in electricity markets that are dominated by renewable resources.

3.2 Back-Propagation of Scarcity Prices

Scarcity pricing through ORDC is fundamentally a real-time market mechanism. The resources that benefit most from the design are those that can respond rapidly to real-time conditions. If energy prices spike due to scarcity, fast-moving units that are offering reserve can be activated, and can pocket the profits of responding to tight system conditions. The mechanism thus has a built-in pay-for-performance attribute. Although the consistency between day-ahead and real-time demand curves is an important aspect of the design (Hogan & Pope, 2019), implementing the mechanism in real time is a *necessary* condition for delivering its intended benefits. It is pointless to introduce this demand curve in the day-ahead market, without following through with a real-time market for reserve capacity. A day-ahead ORDC can potentially affect the market equilibrium, however the pay-for-performance attribute of the mechanism is cancelled unless the reserve imbalances are settled in real time.

In a nutshell, the impact of introducing ORDC is that it sets a real-time price for reserve and uplifts real-time energy prices. The price for reserve is given by expression (1) above, which we will refer to as an ORDC adder. Note that an explicit co-optimization of energy and reserves is not a necessary condition for applying the adder (ERCOT, 2014).

The real-time market for reserve capacity is an essential element of the ORDC design, due to the back-propagation of real-time reserve prices to forward (e.g. day-ahead) reserve markets. A forward price for reserve capacity, in turn, attracts investments into reserves. When the reserve capacity of the system is systematically inadequate, the loss of load probability of the system increases, and this is reflected in real-time and consequently day-ahead reserve prices, thereby attracting reserves. Conversely, when abundant reserve is available in the market, the loss of load probability reflects this and the scarcity adder dissipates to zero both in real time and in the day ahead. This self-correcting behavior of ORDC creates a stable market environment with less risk for investing into reserves. The mechanics of the back-propagation of the ORDC adder to forward reserve markets are the standard mechanics of a two-settlement system. In selling a MW of reserve capacity in the day-ahead reserve market, agents engage in a financial commitment of buying that MW back at real-time prices. For risk-neutral agents, the asking price of selling that MW day ahead is therefore anchored around the average real-time price of reserve capacity.

3.3 Undermining the Back-Propagation of Scarcity Prices

In a market without real-time trading of reserve capacity, such as the European market, this backpropagation is undermined. In the European design, the only commitment implied by selling reserve in the day-ahead market is that this reserve should be bid into the balancing auction. The (day-ahead) price of reserve is therefore driven by the opportunity cost of using that capacity for self-balancing the portfolio of a BRP. As an analytical model[3] of the European BSP/BRP design without fixed (startup or minimum load) costs suggests (Papavasiliou, 2020), this opportunity cost is typically zero. Concretely, therefore, agents lose nothing in selling reserve capacity in the day-ahead market, and therefore they have no clear incentive to ask for anything in return. This undermines the

[3] https://perso.uclouvain.be/anthony.papavasiliou/public_html/AnalyticalV3.pdf

formation of a day-ahead reserve price signal that is driven by the level of scarcity in the system. In practice, reserve suppliers in European markets do ask for something in return, for example recovering the investment costs or fixed startup/min load costs of their units for providing the reserve. But this is not *directly* connected to the short-term level of scarcity in the system. This observation is consistent with the fact that the Belgian market has been carrying a strategic reserve of CCGT units since 2014. The presence of strategic reserve in Belgium is an indication that the Belgian system needs these units, but is not generating sufficient profit for them to remain economically viable in the Belgian market.

4. Implementing Scarcity Pricing in a European Design

In light of the fact that Europe does not operate a real-time market for reserve capacity, an inconsistent application of the principles of the scarcity pricing design would amount to adding the scarcity component to the energy price alone. The distinction between balancing and imbalance prices in Europe allows for further creative combinations: one may envision applying the adder to the imbalance price, but not the balancing price. This allows for a plurality of options that are debated in the implementation of scarcity pricing in Belgium, and which are discussed in turn below. These options can be analyzed (Papavasiliou, 2020) by considering the optimal strategy of a fringe market participant who is stacked up against truthfully bidding competitors. The analysis ignores many of the realistic aspects of electricity markets, but focuses on the interactions between balancing and imbalance prices, and the effects of operating a real-time energy market without a real-time market for reserve. Arguably, a sound market design should at least pass the test of back-propagating scarcity prices in an idealized model. Failure to do so would indicate fundamental problems in more complex realistic settings.

4.1 Scarcity Pricing with a Real-Time Market for Reserve

Consider a market agent who owns reserve capacity, and also manages a portfolio of assets that impose an imbalance that is out of the control of the agent in real time. In European parlance, this is a BRP which also has a BSP attached to its portfolio.

In a market that does not distinguish BRPs from BSPs, and simply applies a unique real-time balancing price to both, the agent faces the following cash flows when trading *energy*:

$$\lambda^B \cdot qa - \lambda^B \cdot (Imb - ai) - C \cdot (qa + ai) \tag{2}$$

λ^B: The balancing price (system lambda).

C: The marginal cost of the reserve resource of the agent.

qa: The amount of reserve capacity that is activated for balancing by the system operator.

Imb: The imbalance in the portfolio of the agent, which is beyond the control of the agent. A positive imbalance corresponds to a net demand for power.

ai: An active imbalance that the agent may choose to effect on its portfolio by dispatching its reserve asset upward or downward, without being asked to do so by the system operator. Positive ai corresponds to an upward activation, i.e. a negative contribution to imbalance.

The first term in the expression above corresponds to the settlement of balancing energy provided by the reserve asset, whereas the second term settles imbalance energy. In a market that does not distinguish balancing prices from imbalance prices, the two are settled at a unique real-time price.

The implementation of scarcity pricing based on ORDC introduces a scarcity adder, which effectively amounts to the real-time price of reserve capacity. This has two effects: (i) Due to the no-arbitrage relation between energy and reserve capacity, this adder uplifts the energy price. (ii) Reserve imbalances need to be settled against the real-time price of reserve capacity. The cash flows of the agent in question therefore change as follows:

$$\begin{aligned}(\lambda^B + \lambda^R) \cdot qa - (\lambda^B + \lambda^R) \cdot (Imb - ai) - C \cdot (qa + ai) \\ + \lambda^R \cdot (P^+ - qa - ai) - \lambda^R \cdot qa^R\end{aligned} \quad (3)$$

λ^R: The scarcity adder (see equation (1)).

P^+: The capacity of the reserve.

qa^R: The reserve traded in the day-ahead market.

The last two terms in this expression implement the real-time market for reserve capacity, and essentially constitute a settlement of *reserve* imbalances. On the one hand, the agent is paid a real-time price for reserve for any amount of reserve capacity that it makes available in real time. On the other hand, the agent must buy back the amount of reserve capacity that it has sold to the system operator in the day-ahead market at the real-time price of reserve. It is exactly the presence of this last term in the settlement that permits the back-propagation of scarcity prices to the day-ahead reserve market.

4.2 The Existing Basic European Design

The absence of a real-time market for reserve capacity in Europe implies that the existing settlement of European balancing markets follows closely equation (2), with the only difference that imbalances are settled at an imbalance price that can be different from the balancing price.

$$\lambda^B \cdot qa - \lambda^I \cdot (Imb - ai) - C \cdot (qa + ai) \quad (4)$$

λ^I: The imbalance price.

In European parlance, the first term is payable to BSPs and is referred to as a balancing settlement, whereas the second term applies to BRPs and is referred to as an imbalance settlement. The balancing activation in Belgium has been paid as bid in the past, however the integrated platforms for aFRR and mFRR will be moving to a uniform price for balancing, which is the assumption in the ensuing discussion.

Assuming that an agent is a fringe supplier in the balancing market (Papavasiliou, 2020), the agent has no interest in misrepresenting its marginal cost to the balancing auction. Overstating its cost would increase its chances of not being profitably activated when the system is short, whereas understating its cost would increase its chances of being activated at a financial loss. When deciding whether or not to self-balance its portfolio (by choosing a non-zero active imbalance ai), the agent is trading off the potential profits from the activation of its balancing capacity in the balancing market against the benefits of recuperating an imbalance price. When balancing prices and imbalance prices are aligned ($\lambda^B = \lambda^I$), the latter option is clearly not advantageous, because the agent will sometimes be paid an imbalance price below its marginal cost, whereas if the agent bids truthfully in the balancing auction it is only activated when the balancing price is profitable for the agent.

This is a positive result: when the balancing price is equal to the imbalance price, $\lambda^B = \lambda^I$, a fringe agent has an interest to bid its entire capacity to the balancing auction at its true marginal cost. This maximizes the degrees of freedom of the system operator, and promotes price discovery and operational efficiency in real time. The problem is that, since the agent has an interest in bidding its balancing capacity in the balancing auction anyway, it has no opportunity cost associated to offering this reserve capacity in the day-ahead reserve market. This latter effect undermines the back-propagation of scarcity prices based on ORDC, and it is exactly this problem that is corrected by putting in place a real-time market for reserve capacity.

4.3 The Belgian Alpha Components

In an attempt to transfer increasing balancing responsibility to BRPs for managing their imbalances within their portfolios, the Belgian system operator has put in place an imbalance price that is characterized by penalties when the system experiences large imbalances. In the past, these penalties would penalize BRPs for exacerbating the balancing problems of the system (ELIA, 2015), they now apply regardless of the balance of the BRP portfolio (ELIA, 2019), just as long as the system imbalance is very long or very short. Concretely, the imbalance price in Belgium can be expressed as follows:

$$\lambda^I = \lambda^B + \alpha^U \cdot \mathbb{I}[Imb^s > UI] - \alpha^L \cdot \mathbb{I}[Imb^s < LI] \quad (5)$$

$\mathbb{I}[x]$: An indicator function that is equal to 1 when condition x is true, and zero otherwise.

Imb^s: Total system imbalance.

α^U, α^L: An alpha component which increases (respectively decreases) the imbalance price when the system is short (respectively long).

UI, LI: The threshold beyond which the system is considered to be short (respectively long).

It has been argued by the Belgian system operator that the alpha components urge BRPs to reduce their imbalances. Insofar as other aspects of balancing are concerned (e.g. improved forecasting), there is a clear incentive generated by these components to avoid being on the wrong side of the imbalance price. Insofar as deploying balancing capacity in order to self-balance a portfolio, in a system with independent and symmetric imbalances there is no interest for a fringe supplier (Papavasiliou, 2020) to use its flexible capacity for achieving balancing within its perimeter. The intuition behind this result is that, for symmetric independent imbalances, the average upward and downward penalties cancel each other out, and the design reverts back to that of the previous paragraph. This is actually a good thing: self-balancing within a BRP's perimeter undermines operational efficiency, since resources are not pooled in the balancing auction, where price discovery and efficient allocation of resources can take place. On the other hand, the problem of back-propagation of scarcity prices persists. Since agents have an interest in bidding their entire reserve capacity in balancing auctions anyway, they face no opportunity cost in selling this reserve capacity in the day-ahead reserve auction.

4.4 Scarcity Adders Limited to Imbalance Prices

An alternative that has been considered for the possible implementation of scarcity pricing is to limit its application to the imbalance price alone[4]:

"*In particular, I do not understand why the authors believe that a correct imbalance price (or real-time price or cash out price) will not result in proper scarcity signals e.g. in day-ahead prices.*".

An interpretation of this proposal it to replace the alpha component in the imbalance price with the ORDC adder in the settlement of equation (4):

$$\lambda^I = \lambda^B + \lambda^R \tag{6}$$

The effect of this proposal is notably different from the alpha component. In this case, and in contrast to the case of the Belgian alpha component, BSPs are rewarded for providing upward activation. Therefore, the tradeoff of agents to use their balancing capacity in order to self-balance their portfolio now creates conditions for certain agents (those with the lowest marginal costs in the merit order stack for upward activation) to take their chances by activating their reserve upwards without being asked to do so by the system operator. With the presence of an ORDC adder, the payoff of the imbalance price can be adequate to cover the losses of sometimes taking an imbalance with an imbalance price which is lower than the marginal cost of the resource which is used for self-balancing. Since the balancing market does not apply the same adder, this will make it advantageous for cheaper resources in the upward balancing stack to migrate towards self-balancing through active imbalances of their reserve resources, and keeping these resources out of the balancing market.

This effect is interesting, because it creates an opportunity cost for bidding these resources into the balancing market, and therefore produces a back-propagation of a day-ahead reserve price. However, the opportunity cost is lower than that of the average ORDC adder (Papavasiliou, 2020), and only applies to a subset of the balancing resources. Therefore, something is being back-propagated, but it is not the average ORDC scarcity adder. Moreover, it is worth noting that self-balancing results in economic inefficiencies, since the imbalances that are being resolved within a balancing portfolio could have been balanced by potentially cheaper resources in the balancing market.

The most interesting effect of this design is the fact that the resources which are self-balancing are replacing resources that would have otherwise been activated in the balancing market. This depresses the balancing energy price, and counteracts the effect that the scarcity adder has on the imbalance price. Although it is true that day-ahead trades contribute towards the imbalance that a BRP would need to buy back at imbalance prices, the system lambda λ^B is depressed by the migration of upward reserves to self-balancing, and this counteracts the effect of the ORDC adder λ^R. Alas, the fundamental problem remains (Papavasiliou, 2020): in the absence of a real-time market for reserve capacity, back-propagation is undermined.

4.5 Our Proposal

The appeal of implementing a real-time market for reserve capacity is that it achieves the back-propagation of scarcity adders based on ORDC, without removing balancing resources from the balancing market (Papavasiliou, 2020). Applying an identical argument to the one of section 4.2, one concludes that agents have an interest in bidding their *entire* reserve capacity at its marginal cost to

[4] Paul Giesbertz, "The power market design column – The scarcity of scarcity pricing", available online: https://www.linkedin.com/pulse/power-market-design-column-scarcity-pricing-paul-giesbertz/.

the balancing auction. But there is also an opportunity cost in selling this capacity away in the day-ahead market, since the capacity needs to be bought back in real time against a scarcity adder which is anchored against the level of stress in the system, as expressed in formula (1).

Concretely, the proposal for implementing scarcity pricing with a scarcity adder based on ORDC relies on the following three changes to the existing design.

- **Design proposal 1**: Replace the alpha components with the scarcity adders of formula (1) in the imbalance price.
- **Design proposal 2**: Align the balancing and imbalance prices by augmenting the clearing price of the balancing auction by the ORDC adder.
- **Design proposal 3**: Introduce a real-time market for reserve capacity by introducing imbalance settlement.

5. Progress on the Implementation of Scarcity Pricing in Belgium

Scarcity pricing is unfolding in a number of markets. This evolution is likely driven to some extent by the integration of renewable resources which is shifting value from energy to reserves. Lessons from the implementation of the mechanism can prove valuable for the rollout of the design in the EU.

5.1 Scarcity Pricing Outside the EU

The scarcity pricing design was launched originally in ERCOT (ERCOT, 2014). ERCOT does not co-optimize energy and reserves in real time, instead the ORDC adder is computed ex post. Two adders are computed, with one applying to reserves that can respond within 30 minutes and one to reserves that can respond within one hour. Operating reserve demand curves are also being introduced in the real-time and day-ahead market of PJM (Hogan & Pope, 2019). Similarly to the ERCOT market, PJM plans to apply adders for reserves of varying quality, referred to as tier 1 and tier 2 reserves respectively. The Federal Energy Regulatory Commission has recently approved the filing of PJM for implementing ORDC (FERC, 2020).

The UK system operator balances the system using a mix of balancing market bids (the analog of free bids, in Belgian market terminology) and the so-called Short-Term Operating Reserve (STOR), which is the analog of frequency responsive reserve (aFRR and mFRR). STOR receives so-called availability payments which can be interpreted as activation costs for energy. The value of these payments is not closely linked to the real-time stress of the system but is rather based on an ex-post calculation (see article 3.46 and figure 3.48 of (Flamm & Scott, 2014)). The UK regulator (OFGEM) recently proposed the introduction of a real-time operating reserve demand function that would set the real-time energy price and more accurately reflect scarcity in the system. The UK ORDCs were introduced in early winter 2015/16 (article 3.51, (Flamm & Scott, 2014)). The ORDC is constructed by using the product of VOLL with loss of load probability as a function of available reserve capacity.

5.2 Scarcity Pricing in EU Legislation

Capacity remuneration mechanisms have traditionally been treated with concern by the European Commission (Papavasiliou & Smeers, 2017). Diverse approaches have been implemented for remunerating capacity in Europe, including capacity obligations, capacity payments, and strategic reserves. This has led to concerns about the potential balkanization of the European electricity market. Moreover, some of these measures have drawn scrutiny as possibly constituting State Aid, and therefore contravening EU competition law.

Scarcity pricing has been considered as an option for alleviating the need to lean on capacity remuneration mechanisms in order to resolve missing money problems. This favorable view of the European Commission towards scarcity pricing is reflected in article 44(3) of the EBGL (European Commission, 2017):

"*Each TSO may develop a proposal for an additional settlement mechanism separate from the imbalance settlement, to settle the procurement costs of balancing capacity pursuant to Chapter 5 of this Title, administrative costs and other costs related to balancing. The additional settlement mechanism shall apply to balance responsible parties. This should be preferably achieved with the introduction of a shortage pricing function. If TSOs choose another mechanism, they should justify this in the proposal. Such a proposal shall be subject to approval by the relevant regulatory authority.*"

The article refers to scarcity pricing as a "shortage pricing function", which should apply to BRPs. We argue in the next section about how our proposal relies on this article in order to align the requirements of our proposed design with EU legislation. It is also interesting to note that the article sets a certain precedence: "*If TSOs choose another mechanism, they should justify this in the proposal*".

The favorable view of the European Commission towards scarcity pricing is similarly reflected in article 20(3) of the Clean Energy Package (European Parliament, 2019):

"*Member States with identified resource adequacy concerns shall develop and publish an implementation plan with a timeline for adopting measures to eliminate any identified regulatory distortions or market failures as a part of the State aid process. When addressing resource adequacy concerns, the Member States shall in particular take into account the principles set out in Article 3 and shall consider:*

…

(c) introducing a shortage pricing function for balancing energy as referred to in Article 44(3) of Regulation 2017/2195;"

The Clean Energy Package stipulates that Member States with adequacy concerns should consider scarcity pricing as an option for relieving the missing money problem. Belgium is one such country with adequacy concerns, as indicated by the procurement of strategic reserve capacity in Belgium in recent years that were considered tight for the system. Strategic reserve comprises plants that do not participate in the market, but can be activated in case of scarcity, as determined by the system operator (Hoeschle & De-Vos, 2016). Strategic reserve is essentially a temporary capacity market that is evaluated on a periodic basis and that needs to be approved by European competition authorities.

5.3 The Belgian Scarcity Pricing Studies

During the fall (September - October) of 2014, four nuclear units in the Belgian system were placed out of order simultaneously. This resulted in an unplanned outage of approximately 4 GW. Given that the capacity of the Belgian system at the time amounted to approximately 15 GW, this corresponded to a severe shock to the system. The winter of 2014-2015 was fortunately relatively mild, however the system experienced significant stress.

The tight fall of 2014 triggered an inquiry by the Belgian regulator about whether adequate measures are in place in the Belgian market for attracting investment. The specific question that was posed by the regulator was how electricity prices would change if we would introduce scarcity pricing through ORDC adders (Hogan, 2005) in the Belgian market. The detailed analysis that is presented in

(Papavasiliou & Smeers, 2017) attempts to address this question by developing a simulation model that reproduced the market conditions from January 2013 until September 2014 in Belgium. The analysis arrived to two main conclusions: (i) under existing market conditions, CCGT units in Belgium are able to cover their short-term variable costs, however they do not earn sufficient profits in order to recover long-run investment costs; (ii) introducing a scarcity adder based on ORDC can restore financial viability for CCGT units, provided ORDC adders are back-propagated to the forward market.

The analysis of (Papavasiliou & Smeers, 2017) was based on 21 months of simulated market conditions, and assumed a back-propagation of the scarcity prices to day-ahead energy prices, with all the limitations that these assumptions imply. Nevertheless, the first conclusion of the finding was consistent with the fact that certain CCGT units in Belgium have been placed in strategic reserve since 2014, and the second conclusion inspired a follow-up inquiry by the CREG about the sensitivity of the results to various assumptions. The follow-up analysis (Papavasiliou, et al., 2018) investigates the sensitivity of the scarcity adder to the restoration of nuclear capacity in Belgium, the assumed value of lost load, and the interaction of scarcity pricing with strategic reserve.

Subsequently, the regulator commissioned a market design proposal for implementing scarcity pricing in Belgium. The proposal (Papavasiliou, et al., 2019) aimed at analyzing the impact of various market design options on the back-propagation of scarcity adders. These design options include the effects of virtual trading, the co-optimization of energy and reserves in the day-ahead market, and the implementation of a real-time market for reserve capacity on back-propagation. The major conclusion from this market design proposal is that the least disruptive and most important measures that should be adopted for a successful application of scarcity pricing in the Belgian market are the alignment of balancing and imbalance prices, and the implementation of a real-time market for reserve capacity.

5.4 Current Status of Scarcity Pricing in Belgium

In 2018, at the request of the Belgian regulatory authority, the Belgian transmission system operator undertook a counterfactual analysis that aimed at computing the ORDC scarcity prices that would have occurred in Belgium in 2017 based on historically telemetered data (ELIA, 2018). For this purpose, ELIA used the so-called Available Reserve Capacity (ARC), which is the amount of reserve capacity that is recorded by ELIA telemetry. The ARC can be used as an input to the adder formula (1), thereby allowing scarcity adders to be computed ex post. The study found few occurrences of a non-zero scarcity price in the system, which is consistent with the fact that 2017 was a comfortable year for the Belgian system. The study further identified a significant sensitivity of the adder to the set of resources that are considered to contribute to ARC. Since October 2019, the Belgian system operator publishes[5] scarcity prices one day after operations based on the ARC that has transpired during the previous day.

The concerns of the Belgian regulator and system operator about the adequacy of Belgian capacity remain, and are exacerbated by the political uncertainty that surrounds the future of nuclear capacity in Belgium. This has led the Belgian government to request the implementation of a capacity remuneration mechanism. This, in turn, has resulted in the activation of article 20(3) of the Clean Energy Package, which calls for the implementation of a "*shortage pricing function*" for Member States

[5] https://www.elia.be/en/electricity-market-and-system/studies/scarcity-pricing-simulation

with adequacy concerns. In response, in the 2019 national implementation plan (Federal Public Service Economy, 2019) the Belgian system operator has argued that

"*the existing alpha component in the imbalance price mechanism … already exhibits quite some characteristics of a scarcity pricing mechanism*".

In this respect, it is interesting to note the response of the European Commission to the Belgian national implementation plan (European Commission, 2020):

"*The Commission also considers that the scarcity pricing function should be triggered by the scarcity of reserves in the system and it should be calibrated to increase balancing energy prices to the Value of Lost Load when the system runs out of reserves. The Commission invites Belgium to consider amending its scarcity pricing scheme accordingly by no later than 1 January 2022.*".

The analysis of paragraph 4.3, which draws a distinction between an imbalance penalty and a scarcity adder, is in line with the position of the European Commission. The rationale is that a large system imbalance per se is not an indication of scarcity. Crucially, there is an alignment between design proposal 1 in section 4.5 and the position of the European Commission.

Another important element in the response of the European Commission to the Belgian national implementation plan (European Commission, 2020) is the following:

"*The Commission, however, invites Belgium to consider whether the scarcity pricing function should apply not only to BRPs but also to balancing service providers (BSPs). This may support security of supply by ensuring that BRPs and BSPs face the same price for the energy produced / consumed, as price differentiation may result in inefficient arbitrage from market players.*"

Design proposal 2 in section 4.5 is thus perfectly aligned with the position expressed by the European Commission.

6. Ongoing Implementation Questions

6.1 Compatibility with EU Legislation

The legal implementation of our proposal can be aligned with European legislation, and specifically the provisions of the EBGL. Two articles of the EBGL constitute the cornerstone of our proposal: (i) the attribution of each BSP to a unique BRP, as per article 18.4(d) of the EBGL, and (ii) the possibility of introducing an additional settlement mechanism separate from imbalance settlement, as per article 44.3 of the EBGL.

Concretely, we break down the settlement terms of formula (3) into individual components (where we set $ai = 0$, since we show (Papavasiliou, 2020) that active imbalances are suboptimal under our proposed market design), and comment on their compatibility with the EBGL:

(A): $\lambda^B \cdot qa$: This is the standard balancing payment that will be produced by the integrated European platforms (e.g. MARI and PICASSO).

(B): $-(\lambda^B + \lambda^R) \cdot Imb$: This is an imbalance energy settlement, where the alpha component of paragraph 4.2 is replaced by an ORDC scarcity adder. This is design proposal 1 in section 4.5, and it is perfectly aligned with the recent position of the European Commission (European Commission, 2020), as we discuss in paragraph 5.4.

(C): $\lambda^R \cdot qa$: This is a scarcity adder to the balancing price which aligns the BRP and BSP settlement for real-time energy. This is design proposal 2 in section 4.5, and it is perfectly aligned with the recent position of the European Commission (European Commission, 2020), as we discuss in paragraph 5.4.

(D): $\lambda^R \cdot (P^+ - qa - qa^R)$: This is a reserve imbalance settlement mechanism, and essentially implements a real-time market for reserve capacity. This is design proposal 3 in section 4.5.

The proposed settlement reproduces exactly the incentives of a real-time reserve capacity market. Note that terms (C) and (D) are essentially related to BSP quantities, but can be attributed to the relevant BRP. The terms (C) and (D) are compatible with EBGL, since (i) every BSP is associated to a BRP according to article 18.4(d) of the EBGL, and (ii) the Belgian TSO is allowed to introduce an additional settlement mechanism separate from imbalance settlement, which shall apply to domestic BRPs, according to article 44.3 of the EBGL.

Thus, the EBGL sanctions the implementation of design proposals 1-3. The implementation of scarcity pricing and a real-time market for reserve capacity in the European electricity market is perfectly compatible with the provisions of the EBGL.

6.2 Multi-Area Considerations

Since the early discussions of scarcity pricing in Belgium, the question has emerged about whether Belgium is in a position to implement scarcity pricing unilaterally, even if neighboring markets would not follow suit. Insofar as cross-border effects are concerned, there is no contradiction between the proposal of section 4.5 and EU law. The crucial observation is that only Belgian BSPs are affected by the ORDC scarcity adder, therefore the adder does not apply on foreign balancing resources which may be activated in the integrated European balancing platforms (e.g. MARI and PICASSO).

The possible economic implementations of this change in market design on neighboring markets are a different matter. Preliminary analysis indicates that the unilateral implementation of the mechanism in Belgium may leave day-ahead reserve prices of neighboring markets unaffected, but may have an effect on the equilibrium of neighboring real-time energy markets. The latter effect would be driven by the fact that Belgium would be more likely to lean on cross-border balancing resources if Belgian BSPs would face an increased opportunity cost due to scarcity pricing.

6.3 Adder Computation

The collaboration with ELIA (ELIA, 2018) for the publication of D+1 scarcity prices brought forth important practical considerations that should be incorporated in the computation of the adder (Papavasiliou, et al., 2019). The set of resources that should be counted towards the Available Reserve Capacity (which constitutes the reserve capacity argument R in the scarcity price formula (1)) can have an important effect on the value of the adder, as the analysis of the Belgian system operator points out. The assumptions about the statistical behavior of imbalances (i.e. whether imbalance increments within a given balancing interval should be considered as being perfectly correlated or perfectly independent (Hogan, 2013)) also have an important effect on the adder. The longer duration of the European balancing market time interval (15 minutes, compared to the 5 minutes in certain US real-time markets) motivates an additional dilemma in the design of the mechanism, namely whether the capacity R in formula (1) should correspond to the leftover capacity before or after the activation of balancing capacity (or some intermediate solution). In systems with shorter balancing market intervals, this distinction is less relevant.

7. Conclusions

The European electricity market, like any electricity market that aims at relying increasingly on renewable resources, will need to adapt to the value shift from energy to reserve capacity that is induced by renewable resources. The absence of a real-time market for reserve capacity, i.e. a market for settling reserve imbalances in real time, is a serious handicap of the European electricity market towards achieving this transition. Fortunately, European legislation allows for this transition to take place, and it is up to the competent Member State regulatory authorities and transmission system operators to exploit the degrees of freedom that are afforded by the European Commission to the fullest for implementing a future-proof balancing market.

Scarcity pricing emerges as a no-regret measure in this respect. The appeal of scarcity pricing is that, if a system is not under stress, the scarcity pricing mechanism dissipates, and the market reverts back to its default state. The mechanism only becomes active when the system is under stress, and works towards relieving this stress. The dilemma between capacity markets and scarcity pricing is false: scarcity pricing does not preclude capacity remuneration mechanisms. It is perfectly compatible with capacity remuneration mechanisms. Precedence, however, matters: before we proclaim the 'energy-only' market dead, let us give it an opportunity to function properly.

Since 2014, the Belgian regulator with the help of the system operator has pioneered the implementation of scarcity pricing in Central Europe. The Belgian regulator and system operator have at their disposal the necessary legal institutions and technological infrastructure to implement scarcity pricing, and ORDC scarcity prices are already being computed in Belgium since October 2019. We argue in this paper that the next step in the process is to introduce scarcity adders to balancing prices and imbalance prices, and to introduce a real-time market for reserve capacity.

We summarize our proposal in three items: (i) replace alpha components with scarcity prices based on LOLP considerations; (ii) align BSP and BRP settlement; and (iii) put in place a real-time market for reserve capacity that settles reserve imbalances. The first two items are perfectly aligned with the recommendations of the European Commission in response to the Belgian national implementation plan. The third step is essential, in order to ensure that scarcity prices can back-propagate to forward reserve and energy markets, and European legislation allows for its implementation.

Acknowledgements

The author would like to thank William Hogan, Alain Marien, Marijn Maenhoudt, and Gilles Bertrand for valuable discussions. This research has been supported financially by the Belgian Regualtory Commission for Electricity and Gas, the Francqui Foundation Research Professorship 2018-2021, and the EPOC 2030-2050 project (Belgian energy transition funds).

References

De-Vos, K. et al., 2019. Dynamic Dimensioning Approach for Operating Reserves: Proof of Concept in Belgium. *Energy Policy,* Volume 124, pp. 272-285.

ELIA, 2015. *Tariffs for maintaining and restoring the residual balance of the individual access responsible parties 2016-2019,* Brussels: Belgian Transmission System Operator.

ELIA, 2018. *Study report on scarcity pricing in the context of the 2018 discretionary incentives,* Brussels: Belgian transmission system operator.

ELIA, 2019. *Tariffs for maintaining and restoring the residual balance of individual access responsible parties 2020-2023,* Brussels: Belgian Transmission System Operator.

ENTSO-E, 2017. *Explanatory document to all TSOs' proposal for the implementation framework for a European platform for the exchange of balancing energy from frequency restoration reserves with manual activation,* Brussels: ENTSO-E.

ERCOT, 2014. *ERCOT Market Training: Purpose of ORDC, Methodology for Implementing ORDC, Settlement Impacts for ORDC,* TX: ERCOT.

European Commission, 2017. *Commission regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing,* Brussels: European Commission.

European Commission, 2020. *Commission opinion of 30/04/2020 pursuant to article 20(5) of Regulation (EC) No 2019/943 on the implementation plan of Belgium,* Brussels: European Commission.

European Parliament, 2019. Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast). *Official Journal of the European Union,* pp. 1-71.

Federal Public Service Economy, 2019. *Belgian electricity market: implementation plan,* Brussels: Federal Public Service Economy.

FERC, 2020. *Docket Nos. EL19-58-000 and ER19-1486-000,* Washington, DC: United States Federal Energy Regulatory Commission.

Flamm, A. & Scott, D., 2014. *Electricity Balancing Significant Code Review - Final Policy Decision,* London: Office of Gas and Electricity Markets.

Hoeschle, H. & De-Vos, K., 2016. *Implementation of a strategic reserve in Belgium: Product design and market results.* Paris, Proceedings of the International Council on Large Electric Systems (CIGRE).

Hogan, W., 2005. *On an 'energy only' electricity market design for resource adequacy,* Cambridge, MA: Center for Business and Government, JFK School of Government, Harvard University.

Hogan, W., 2013. Electricity scarcity pricing through operating reserves. *Economics of Energy and Environmental Policy,* 2(2), p. 65–86.

Hogan, W. & Pope, S., 2019. *PJM Reserve Markets: Operating Reserve Demand Curve Enhancements,* Cambridge, MA: Center for Business and Government, JFK School of Government, Harvard University.

Papavasiliou, A., 2020. *Analytical Derivation of Optimal BSP / BRP Balancing Market Strategies,* Louvain la Neuve: Center for Operations Research and Econometrics, UCLouvain.

Papavasiliou, A. & Smeers, Y., 2017. Remuneration of Flexibility under Conditions of Scarcity: A Case Study of Belgium. *The Energy Journal,* 38(6), pp. 105-135.

Papavasiliou, A., Smeers, Y. & Bertrand, G., 2018. An Extended Analysis on the Remuneration of Capacity under Scarcity Conditions. 7(2).

Papavasiliou, A., Smeers, Y. & De-Maere-D'Aertrycke, G., 2019. *Study on the general design of a mechanism for the remuneration of reserves in scarcity situations,* Louvain la Neuve: Center for Operations Research and Econometrics, UCLouvain.