Study on the general design of a mechanism for the remuneration of reserves in scarcity situations

Anthony Papavasiliou, UCLouvain

Yves Smeers, UCLouvain

Gauthier de Maere d'Aertrycke[1], Tractebel ENGIE

[1] The information and views set out in this report are those of the author(s) solely and do not necessarily reflect the official opinion of ENGIE and any of its subsidiaries

Table of Contents

Executive Summary

Scarcity pricing is the principle of pricing electricity at a value above the marginal cost of the marginal unit during conditions of high system stress, according to the incremental value that flexible capacity offers to the system in terms of keeping loss of load probability in check. Concretely, scarcity pricing is implemented by including an adder to the imbalance price on top of the marginal cost of the marginal unit, and by rewarding that same adder to standby reserve capacity. The effect of this mechanism is that (i) it rewards flexible resources for being available, even if not activated, and (ii) it rewards flexible resources for reacting to system imbalances when the system is short on flexible capacity. Through economic arbitrage, scarcity pricing creates the potential of giving rise to a long-term investment signal for building flexible capacity or mobilizing demand response that can deliver security to the system.

Detailed numerical analyses of the Belgian market have demonstrated the potential of scarcity pricing to overturn the financial viability of flexible technologies in Belgium, and also to create a strong investment signal for mobilizing demand response. In response to these encouraging indicators about the potential of scarcity pricing to attract flexibility in the Belgian market, the present report discusses concrete market design measures that would enable scarcity pricing to function effectively in the context of the Belgian market design.

We propose a range of increasingly disruptive measures for the evolution of Belgian market design that would enable scarcity pricing to deliver its intended benefits to the Belgian market. The sequence of market design evolutions that we consider are presented in the following figure.

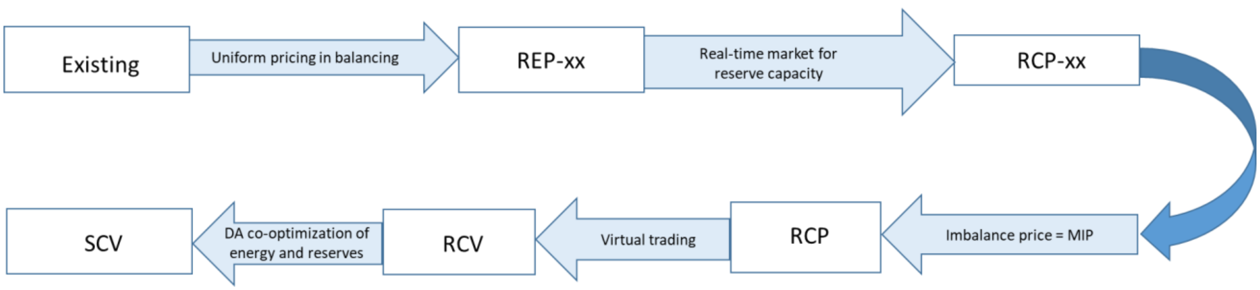


Figure 1: The evolution of market designs that have been considered in our study.

The concrete measures that would be required for this sequence of evolutions can be described in the following table.

From: Existing	To: REP-xx
Resources (free bids / reserves) that are activated in the balancing market are paid as bid. Resources that cause imbalances pay an imbalance price.	Resources (free bids / reserves) that are activated in the balancing market are paid a uniform price, which is the price paid by resources that are causing the imbalances.
From: REP-xx	**To: RCP-xx**
Single-product auction: only energy is traded in real time. No real-time reserve capacity price exists.	Multi-product auction: activated energy *and* reserve capacity are settled in real time. A real-time reserve capacity price is computed in real time.

Free bids are only paid a real-time energy price if they are activated to clear imbalances. BSPs are only paid a real-time energy price if they are activated to clear imbalances.	Free bids that are standing by (but not activated) are paid a real-time reserve capacity price. BSPs that are activated are paid a real-time energy price for activated energy, but pay a real-time reserve price for using up activated capacity.
From: RCP-xx	**To: RCP**
Resources that are short in real time pay a penalty α in addition to MIP whenever the system is (very) short. Resources that are long in real time are paid MDP minus a penalty α whenever the system is (very) long.	Resources that are short pay the system marginal cost for oversupply. Resources that are long are paid the system marginal cost for undersupply.
From: RCP	**To: RCV**
Only entities with physical assets are allowed to participate in the day-ahead market. Portfolio bids correspond to physical assets.	Entities without physical assets can trade in the day-ahead *energy* market. Virtual bids are separately identified from portfolio bids of physical assets.
From: RCV	**To: SCV**
Reserve capacity is auctioned before the clearing of the day-ahead energy market. The day-ahead exchange only trades energy products.	There is no separate reserve capacity auction. Option 1 (exchange approach): The day-ahead exchange introduces reserve products. Option 2 (pool approach): Bids in the day-ahead market correspond to individual resources, and the allocation of energy and reserves is co-optimized.

Table 1: Implementation measures that would be required for the transitions that are considered in our study.

These market designs have been simulated against a realistic model of the Belgian electricity market. The energy prices resulting from these different designs are indicated in the following table.

Month	SCV	RCV	RCP	RCP-0.1	REP-0.1	REP-0.1 inelastic	Hist. DA	Hist. RT
1	40.09	41.00	41.00	41.12	35.61	56.39	52.50	39.51
2	31.12	31.17	31.17	31.31	30.68	32.63	55.41	61.04
3	45.94	46.88	46.88	47.05	30.60	66.23	43.12	36.57
4	35.76	37.38	37.36	37.49	28.77	51.50	35.94	33.31
5	38.40	41.25	41.25	41.32	27.17	63.36	32.61	29.48
6	20.91	21.71	21.73	21.87	19.61	26.42	25.39	21.80
7	20.82	21.12	21.13	21.23	20.74	21.44	27.13	25.11
Average	33.29	34.36	34.36	34.48	27.60	45.42	38.87	35.26

Table 2: Energy price (€/MWh) for the models considered in the study.

The effect of the proposed designs on the profitability of flexible resources and demand response can be used as a basis for assessing their expected benefits, and weighing them against the level of disruption that they would introduce to existing Belgian market operations. On the basis of this analysis, we arrive to the following conclusions:

- An introduction of scarcity pricing as an adder to the real-time energy price alone is not expected to have any material impact on the price of reserves or the profitability of flexible resources.
- The introduction of a real-time market for reserve capacity is the lowest-hanging fruit in the Belgian market design: it is the easiest measure to implement, and it is expected to have a great effect on the long-run incentive to invest in flexible resources.
- The introduction of virtual trading and the co-optimization of energy and reserves in the day-ahead market are more disruptive measures, relative to the introduction of a real-time market for reserve capacity. In an environment of risk-neutral agents, they are also expected to have a minor impact relative to the introduction of a real-time market for reserve capacity in terms of back-propagating scarcity prices.

On the basis of these observations, the concrete recommendation to the Belgian regulator is to proceed with the introduction of a real-time market for reserve capacity. This market can be put in place with the introduction of an energy adder, and a price for real-time secondary reserve capacity and real-time tertiary reserve capacity.

The report is structured as follows:

In section 1 we compare the Belgian market design to the ERCOT market design. We summarize various definitions that are used in these markets in the appendix, and note similarities or equivalences whenever they exist.

In section 2 we outline the main features of the two ends that we consider in the market design spectrum: the current Belgian design, and the US two-settlement system. We propose a stochastic equilibrium model for quantifying the effects of these designs on the back-propagation of scarcity prices, and we define the three main market design questions that were raised by the CREG in the terms of reference:

(i) What is the effect of a real-time market for reserve capacity on the back-propagation of scarcity prices?

(ii) What is the effect of virtual trading on the back-propagation of scarcity prices?

(iii) What is the effect of the co-optimization of energy and reserves on the back-propagation of scarcity prices?

In section 3 we present the measures that would be required for introducing a real-time market for reserve capacity in the Belgian market design. We also discuss the interaction of such a market with existing initiatives for balancing coordination. In section 4 we introduce virtual trading and discuss the measures that would be required for implementing virtual trading in the Belgian market. In section 5 we discuss options for the co-optimization of energy and reserves in the Belgian market.

Section 6 presents a numerical case study of how the aforementioned market design measures (real-time market for reserve capacity, virtual trading, and day-ahead co-optimization of energy and reserves) would affect the profitability of CCGT units and flexible loads that can offer demand response to the system in the form of a reserve service. We conclude that the first measure, i.e. the introduction of a real-time market for reserve capacity, is the simplest measure to implement, and the one with the greatest effects on the back-propagation of scarcity adders in a market with risk-neutral agents.

In section 7 we discuss various practical considerations that relate to the computation of scarcity adders and settlement. This work has been performed in conjunction with ELIA in the context of the 2018 and 2019 scarcity pricing incentive, and a parallel report has been published by ELIA [1] where the topics of section 7 are developed in further detail.

Section 8 concludes with the main findings of our analysis.

1. General Design of a Scarcity Pricing Mechanism

In this section we describe the general principles of scarcity pricing. We then describe the Texas market, and compare it to the Belgian market, highlighting certain differences that relate to scarcity pricing explicitly. We close this section with a short description of scarcity pricing in the Texas market. The transposition of this mechanism to the Belgian market is the topic of the following sections of the report.

1.1. Principles of ORDC

The proliferation of renewable resources in electric power systems, combined with the planned retirement of a significant amount of nuclear capacity in Belgium, have recently raised concerns about whether adequate capacity, especially flexible capacity, is ensured for the future needs of the Belgian system [1]. This increased need for flexible capacity is compounded by an uncertain investment environment, which has discouraged investors from maintaining or expanding the commitment of capital to flexible resources such as combined cycle gas turbines.

The design of the Belgian electricity market, which is an energy-only market[2], is to some extent relevant for this slowdown in investment. Any energy-only market is prone to a missing money problem, whereby caps on energy prices result in capital that cannot be recovered [2]. This issue is especially relevant for peaking generators with high marginal costs, such as combined cycle gas turbines. These peaking technologies are the ones that are the first to be pushed out of the merit order as a result of the introduction of low marginal cost resources such as renewables in the system. But these are also the technologies that are best suited for balancing the system in the presence of renewable resources. This paradoxically results in these technologies, which are needed most for supporting system flexibility, to being the least attractive from the point of view of risk-averse investors. Ultimately, this paradox is a reflection of a problematic valuation of reserve services in energy-only markets.

Scarcity pricing has been proposed as a partial correction to the imprecise valuation of reserve services [2]. The principle of scarcity pricing is to add a correction to the real-time price which rewards generators that can respond rapidly so as to balance the system. The theoretical justification of the approach is that it adjusts the real-time price of energy and reserve capacity such that the resulting dispatch of profit-maximizing generators would reproduce the optimal dispatch that would be obtained if the contribution of reserve capacity towards reducing the loss of load probability would be accounted for [3]. This principle was already retained in the E&W pool (the first restructuring in the UK). At that time the adder was explicitly introduced in the day-ahead market. It led to the exercise of market power but this was mainly due to the implementation (not the principle).

The mechanism has an equivalent, intuitive interpretation in terms of a demand function for operating reserve. This has resulted in touting scarcity pricing equivalently as "Operating Reserve Demand Curve" (ORDC). The rationale for an ORDC interpretation can be developed as follows. Consider the marginal value of reserve capacity when there is very little capacity left. In Texas, when the system reserve capacity drops below 2000 MW the system operator is willing to involuntarily curtail demand in order to prevent cascading outages [4]. Effectively, the marginal value of reserve capacity under these conditions is equal to the value of lost load, which in Texas is set administratively to 9000 $/MWh [4]. When abundant reserve capacity is available in the system (e.g. above 5000 MW in Texas [4]), the marginal value of capacity is equal to zero. For intermediate values, the marginal value of reserve depends on the loss of load probability. This corresponds to the introduction of a demand function for operating reserve capacity that the system operator submits to a multi-product auction that simultaneously clears energy and reserve in

[2] The presence of strategic reserve implies hidden subsidies [18] and potentially distorts the signals that would be generated by an energy-only design.

the market. The effect of this demand function is that, under conditions of scarcity in reserve capacity, it lifts the energy price by a scarcity adder, which also applies to reserve capacity. A simplified formula for this adder (the formula is described in detail in section 7) when there exists a single type of reserve is given by the following expression:

$$(VOLL - \lambda) \cdot LOLP(R).$$

The notation here is as follows: $VOLL$ corresponds to the value of lost load, λ is a proxy of the marginal cost of the marginal unit, R is the amount of remaining reserve capacity, and $LOLP$ is the loss of load probability. Note that as the system becomes tight (R decreases), adding this term to the energy price tends to push the energy price to $VOLL$. The distinction with a pure energy-only market is that this occurs in a smooth and more predictable fashion. When abundant capacity is available (R is very large), the adder dissipates and has no effect on the energy price.

In a two-settlement system, the scarcity adder directly impacts in real time the resources that can rapidly be dispatched upward: they receive the scarcity adder in addition to the marginal cost of the marginal unit (as discussed later, the adder is actually embedded both in the real-time energy price as well as in the real-time price of reserve capacity) . But this scarcity signal is not meant to only apply to real-time operations. Financial arbitrage between day-ahead and real-time markets back-propagates the scarcity signal to the day-ahead market and hence creates a favorable environment for all resources that can offer reserve capacity. Such resources are inherently required in systems with significant shares of renewable power supply. With that being said, a notable difference between scarcity pricing and capacity mechanisms is the built-in 'pay for performance' attribute of the scarcity pricing mechanism. Indeed, under scarcity pricing, the stress of the system is signaled by the real-time price which is enhanced by a scarcity adder, therefore it is in the best interest of resources to perform exactly when the system is most stressed (otherwise they pay for their shortfall in real time, or forgo profit opportunities). In a capacity mechanism, this performance attribute needs to be closely specified in the mechanism (by defining ad-hoc de-rating of capacities depending on their characteristics or penalties for unavailability during stress events) and requires ex-post monitoring of those performances.

Scarcity pricing is not a panacea. The mechanism is designed to reward short-run flexible capacity. Systematic adequacy problems that result in scarcity will of course be reflected in the mechanism, but that would occur anyways under an energy-only market design. Therefore, the mechanism will not result in price adjustments that would be entirely different from those of an energy-only market design in a system that faces the risk of multiple days or weeks of low renewable output. Such systems require long-range storage, a coupling with the heating sector, or other alternatives that fall out of the scope of scarcity pricing. At the same time, however, an appealing aspect of the mechanism is the fact that it can co-exist with capacity markets [2]. Capacity markets are especially well suited for hedging investors in an environment with significant regulatory uncertainty, such as the European market. It is therefore especially appealing that the two can co-exist. Regardless of the implementation of capacity mechanisms, however, it should be noted that real-time markets need to be designed correctly. This is especially true for future systems where real-time conditions are expected to vary in extremely unpredictable patterns due to renewable energy integration. For this reason, some form of scarcity pricing merits careful consideration in systems that lean substantially on operating reserves.

One important aspect of the present report which is not in the scope of the analysis is how scarcity co-exists with neighboring markets that do not implement scarcity pricing. This issue is especially relevant for Belgium, which is a relatively small market compared to its German and French neighbors. This issue, which is especially complicated by the treatment of transmission capacity in the European market coupling design, is left for future investigation.

1.2. The Texas Market Compared to the Belgian Market

We next compare certain elements of the Texas markets to elements of the Belgian market that relate to scarcity pricing. A glossary is added in the appendix, which summarizes certain terms that are used in the Texas and Belgium markets, and notes correspondences wherever relevant. This material is drawn from Texas operations manuals [4], [5], [6], as well as the online resources that are made available by ELIA [7], [8], [9], [10], [11], [12]. Since transmission constraints are intentionally left out of the scope of the present analysis, the following presentation of operations and terminology is limited to energy and reserves.

1.2.1. The ISO Model

Texas

The Texas Independent System Operator integrates the operation of the power system and the operation of the electricity market. This integrated operation takes place both in the day ahead as well as in real time.

Belgium

In Belgium, the Transmission System Operator operates the system in real time. In the day ahead the system operator is responsible for procuring reserve capacity, whereas the Power Exchange clears the energy market.

1.2.2. Types of Reserve

Texas

The ERCOT market offers the following types of reserves: regulation, responsive reserve services, and non-spinning reserve. Each of these services is described below.

Regulation service is an ancillary service that responds within three to five seconds, in response to changes in system frequency. This service is deployed by the load frequency control function, which is described below. There are two types of regulation service, upward and downward. Both directions of regulation can be offered by both loads as well as generators.

Responsive reserve is deployed when upward regulation is used up. It is also used when the security constrained economic dispatch function does not have enough capacity to dispatch. Responsive reserve needs to make its full capacity available within 10 minutes. Responsive reserve capacity is activated proportionally by all resources that provide responsive reserve. Every time responsive reserve capacity is activated, this triggers a run of the SCED.

Non-spinning reserve is used in order to free up responsive reserve. It is expected to respond within 30 minutes, although the actual response time may vary depending on the type of resource. Responsive reserve can be made available by off-line generators, on-line generators, or controllable load. Once non-spinning reserve is activated, its set point is determined by a SCED run.

ORDC adders effectively influence all of the above reserves, because ORDC adders are applied to the energy which is activated by the load frequency control (see description of the LFC function below).

Belgium

The Belgian market has three main types of reserve: FCR, aFRR and mFRR. Each of these services is described below.

Frequency control reserve (FCR) / primary reserve is used for frequency control, and needs to respond within 0 to 3 seconds. Primary reserve is driven by automatic controllers, and depends on European-wide frequency measurements.

Automatic frequency restoration reserve (aFRR) / secondary reserve is reserve that needs to be available within 7.5 minutes.

Manual frequency restoration reserve (mFRR) / tertiary reserve is a reserve product that needs to respond within 15 minutes. Tertiary reserve for generators is classified between R3 standard and R3 flexible. The distinction between the two types of tertiary reserve relates to the fact that R3Flex has a limited number of activations. This distinction is not relevant for the application of scarcity pricing, since the amount of system capacity that is available at a given imbalance interval is updated every 15 minutes (this is the case with the computation of Available Regulation Capacity (ARC), which is updated every 15 minutes). The demand side can offer two types of tertiary reserve products, R3 flexible and ICH (although the latter has been discontinued since 2018). Demand-side R3 flexible products are portfolio based and the system operator does not have an exact view on the amount of demand response inside the portfolio, especially not differentiated in a 15-minute granularity. R3 flexible is traded with a monthly time step, ICH used to be traded with a yearly time step.

1.2.3. Real-Time Markets

Texas

Texas operates a real-time energy market. Texas also operates a real-time reserve market, even if there is no co-optimization in real time. The real-time process is a co-optimization of energy and transmission. In detail, the real-time energy dispatch process consists of the following processes: (i) real-time network security analysis, (ii) security constrained economic dispatch, and (iii) load frequency control.

Real-time network security analysis collects data from QSEs regarding generator and load resources, and from TSEs regarding network elements, in order to confirm that the system is secure. The result of this process is a set of *constraints and shift factors* that need to be input into SCED when deciding the dispatch of units, in order to respect binding constraints either in the base case or in contingencies. Shift factors determine the contribution of shadow prices for transmission constraints to the locational marginal price of a certain bus.

Security constrained economic dispatch (SCED) uses the economic offers of market participants as well as the results of the real-time network security analysis in order to determine both the real-time dispatch of units, as well as the real-time prices for energy and reserves. SCED is executed every five minutes, but may also be executed more often, as needed by ERCOT, for example when responsive or non-spinning reserve is deployed. Instructions that are issued by ERCOT need to be followed by QSEs, and non-compliance may imply charges. SCED produces *locational marginal prices*, which are combined with *reserve price adders* (scarcity adders) in order to determine *real-time settlement point prices*. When SCED runs out of capacity, it generates scarcity prices even in the absence of ORDC.

Load frequency control (LFC) is performed every 4 seconds. LFC is the function which is responsible for deploying upward regulation, downward regulation and responsive reserve. The deployment of these services is averaged over 15-minute intervals and the resulting average is paid the real-time settlement point price. LFC is essentially a proportional controller which deploys regulation in proportion to the deviation between measured and target frequency.

Belgium

Belgium operates a real-time balancing market. This market is used for activating reserve in order to eliminate the imbalances of *balancing responsible parties*. Activated resources are paid as bid for their activated energy, although the introduction of pay-as-cleared has recently been advocated by the

European Commission[3]. BRPs are charged for their imbalances at the *balancing price*. A different balancing price may apply for upward and for downward imbalances, depending on whether the system is longer or shorter than a certain threshold (typically 200 MW). The difference between the prices in the two directions is equal to a parameter referred to as *alpha*[4]. These prices apply to the quantity of upward regulation and the quantity of downward regulation that is activated within a given imbalance interval. If the two prices are equal, then they coincide with the marginal cost of the marginal unit that has been activated for balancing the system.

1.2.4. Imbalance Settlements

Texas

Imbalances in Texas are settled at the real-time settlement price, which is the LMP plus the ORDC adder. This price applies to the difference between the financial position of a resource and its physical position, as metered by ERCOT. The settlements are broken into 15-minute intervals. Imbalance charges are also applied to delivered reserve capacity. This is in stark difference to the Belgian market.

In addition to imbalance charges, resources are required to pay *base point deviation charges* if they do not follow dispatch instructions and ancillary service deployments within defined tolerances. These tolerances are defined as being equal to the maximum of +/-5% or +/5 MW. These charges do not apply if a resource is contributing to the correction of frequency when the frequency deviation is greater than 0.05 Hz, or for intervals during which responsive reserve is deployed. Recall that responsive reserve is deployed when regulation capacity is used up.

Belgium

Imbalances in Belgium are cleared by *balancing service providers*, who either make pre-contracted reserve capacity available in real time, or bid excess capacity to the balancing market in the form of *free bids*[5]. As mentioned previously, BSPs are paid as bid, whereas the BRPs who are creating the need for balancing are charged a uniform imbalance price, which may be different depending on whether the system was short or long. BRPs are actively encouraged to stay balanced in real time, although there is no formal way of imposing this requirement[6].

The notion of BRPs which are responsible for balancing their financial positions with their physical net injection or offtake differs from US operations. US real-time market design is based on the notion that individual resources should be adjusted, even if very close to real time, to a level that helps the system, even if that implies significant changes with respect to previously defined set-points. The philosophy of operating a system with BRPs is that, as long as BRPs take all the necessary actions to balance within their own perimeter of resources, there is a small residual uncertainty left over to the system operator. This view ignores transmission constraints and the benefits of diversifying resources close to real time, and is linked with the view of treating the day-ahead market as the spot market for trading electricity, with real-

---

[3] See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R2195&from=EN, article 30.

[4] The alpha parameter is a stress signal resulting from a high Net Regulation Value. This parameter is not a scarcity price in itself, in the sense that it may be non-zero even if the system has abundant reserve capacity.

[5] The marginal cost that is bid by asset owners for free bids is only restricted by a cap that is defined in order to avoid market distortions due to interactions with the regulated imbalance price in case of activation of strategic reserve (see balancing rules [10], sections 8.5.1 and 8.6.1).

[6] The requirement of BRPs remaining in balance is only a 'soft' requirement in European market operations. For example, article 10.2 of [7] stipulates that BRPs can be rewarded for supporting the system in real time by deviating from their balance, if certain conditions are met. It should be emphasized that there is a distinction between doing so by responding to a dispatch/activation signal, versus doing so by speculating on the direction of the system imbalance. The latter is clearly more risky.

time balancing being treated as a set of services that are put in place in order to support the spot market trading positions, which in the view of European market design are the day-ahead positions[7]. With the increasing integration of renewable resources, and the consequent need for market players to adjust their positions to the rapidly evolving real-time conditions of the system, it is becoming increasingly difficult for the day-ahead market outcomes to reflect real-time conditions, and it worth considering an approach whereby the day-ahead market is treated as a forward financial market, and real time acts as the spot market against which forward positions are settled financially.

Given this decentralized view of balancing the system, there are two philosophies of balancing, which are referred to as *reactive* and *proactive balancing*. **Reactive balancing** places increased responsibilities for balancing on market parties. The idea is to provide advance information about whether the market will be long or short, so that BRPs can take all necessary actions leading up to real time so as to prevent imbalances within their perimeter. This is the approach that is adopted in the Belgium-Netherlands-Luxembourg area. **Proactive balancing** is applied in France. Three hours before real time, BRPs are required to freeze their positions, and if there are imbalances the TSO is responsible for managing them. Germany applies an intermediate solution, in the sense that a reactive balancing approach is adopted in practice, however the information which is used for indicating the direction and magnitude of imbalances may be up to one week old.

An important feature of scarcity pricing is the notion of rewarding resources for responding in a direction which benefits the balancing of the system. *By contrast, BRPs in Belgium are encouraged to maintain their day-ahead and intraday positions through imbalance tariffs* (for a discussion about this tension, see [13]). Table 3 describes the Belgian imbalance tariff. In this table, MIP refers to the marginal increment price, MDP refers to the marginal decrement price, and a1 and a2 are penalties[8].

		Net regulation volume	
		Negative (downward regulation)	Positive (positive regulation)
BRP imbalance	Positive	**MDP-a1**	**MIP**
	Negative	**MDP**	**MIP+a2**

Table 3: The Belgian imbalance tariffs (source: ELIA[9]).

The way to read the table is the following:

- Upper right bold entry: If the system is short (i.e. in conditions of undersupply), and a resource is in positive imbalance (i.e. helping the system), then the resource is paid the system incremental price. This is identical to real-time pricing.
- Lower left bold entry: If the system is long (i.e. in conditions of oversupply) and a resource is in negative imbalance (i.e. helping the system), then the resource pays back the decrement price. This is identical to real-time pricing.

[7] An artefact of this viewpoint is that if the day-ahead market clears at its cap, 3000 €/MWh, then the imbalance price is set by default at 4500 €/MWh (even before real-time conditions are revealed) so that BRPs cannot lean on the balancing market for avoiding expensive procurements in the day-ahead market.
[8] Note that it is possible that ELIA will reinforce the alpha penalties in the immediate future in order to strengthen the incentive for BRPs to retain a balanced perimeter in real time.
[9] http://www.elia.be/~/media/files/Elia/Products-and-services/Balancing/Tarifaire-fiches_Na-beslissing-CREG_FR.pdf

- Upper left bold entry: If the system is short (undersupply) and a resource is in negative imbalance (hurting the system), then the resource pays the incremental price *plus* a penalty.
- Lower right bold entry: If the system is long (oversupply) and a resource is in positive imbalance (hurting the system), then the resource is paid the decrement price minus a penalty.

1.2.5. Day-Ahead or Earlier Markets

Texas

The day-ahead operations of Texas ensure transparent price signals and operational reliability by separating day-ahead operations into three steps: the **day-ahead market** (10am – 1:30pm), which is followed by the **adjustment period** (1:30pm – 2:30pm), which is followed by **day-ahead reliability unit commitment** (2:30pm – 6pm) the day before operations. The idea is to allow the generation of price signals in the day-ahead market, and incorporate self-commitments in the adjustment period before the system operator steps in and commits additional resources as needed for reliability reasons in the reliability unit commitment stage.

Day-ahead market. The day-ahead market model is a co-optimization of energy and ancillary services. Participation in the day-ahead market is voluntary, however ancillary services carry physical commitment to deliver in real time. Texas operates a day-ahead multi-product auction, using co-optimization. Entities bid unit technical and economic characteristics. Texas uses the following terminology for day-ahead bids: (i) *energy bids* are demand-side bids that can be submitted by QSEs in order to procure energy; (ii) *energy-only offers* are supply-side bids that can be submitted by QSEs in order to sell energy; (iii) *three-part supply offers* are supply-side bids that are linked to specific generators and can be submitted by QSEs in order to sell energy. To be clear, energy bids and energy-only offers can be virtual bids, as they are not specific to a resource. Therefore, virtual bidding is allowed in the Texas day-ahead market. The three-part bid consists of a startup minimum energy, and an energy offer. Energy offer curves internalize startup costs, whereas three-part offers declare these startup costs and allow the market clearing algorithm to decide whether the resource should be committed or not.

Every QSE in the day-ahead market has an ancillary service obligation. Ancillary service obligations are allocated as a function of the load that each QSE is serving. These ancillary services are either self-provided, or purchased from ERCOT. Self-provided ancillary services are either provided by private backup generation, or guaranteed through bilateral agreements.

Adjustment period. The adjustment period allows QSEs to make changes to the current operating plans (COPs) of reliability must-run (RMR) units and other resources that are committed in the day-ahead market. The COPs consist of the on/off status of units, their technical min and max, and their ancillary service responsibilities for the following day. COPs are individually defined for every different generator. After the clearing of the day-ahead market, these COPs can be updated from 1:30pm until 2:30pm the day before operations in the adjustment stage. These updates may occur due to outages, de-ratings, RUC, or self-commitments following the clearing of the day-ahead market. The adjustment period allows bilateral trades of energy, capacity and ancillary services between QSEs. The adjustment period is also used for executing trades between QSEs and neighboring systems through the DC ties.

Day-ahead reliability unit commitment. DRUC consists of *transmission security analysis* (TSA) and *reliability unit commitment* (RUC). The TSA process collects unplanned transmission outage data from TSPs, unplanned resource outages from QSEs, and performs a network security analysis which produces a list of contingencies which are input into the RUC. The process resembles the day-ahead market model, in the sense that it uses the techno-economic data of generators and loads as input, but also accounts for the screened contingency constraints that are generated by the transmission security analysis. The RUC process may commit or de-commit resources.

The dispatch or commitment of *any* resource may be changed in RUC. In RUC the marginal cost of the generators is ignored, and ERCOT only accounts for startup costs and minimum energy costs. Ramp constraints are ignored in RUC. Thus, even if it is very costly to resolve a congestion using an online resource, ERCOT will use this resource and will generally avoid committing new units in RUC unless it is needed for reliability reasons.

The two-settlement system. The ERCOT market, like most US markets, operates under the principle of the two-settlement system, whereby the day-ahead market is treated as a forward financial market and the real-time market is treated as the spot market and sets the price for the settlement of forward trades. The principle of the two-settlement system is that the real-time price is used for settling deviations between forward positions and actual production or consumption in real time. This is equivalent to buying out an agent's financial position at the real-time price and being paid for physical transactions at the real-time price. The attraction of the two-settlement system is that it hedges risk while ensuring that agents are exposed to an efficient real-time price signal. Price convergence between day-ahead and real-time prices in two-settlement systems is discussed in further detail throughout the report.

Pool. The ERCOT market may be considered as a pool, in the sense that generators bid the fixed and variable costs of individual generators into the day-ahead auction. The auction then conducts a co-optimization of energy, transmission and reserves, and determines the welfare maximizing allocation of these resources simultaneously. A market clearing price is also generated as a by-product of the auction, and side payments (uplift) are used in order to compensate generators if there exists a deviation between their profit maximizing position and the dispatch instructions of the ISO.

By contrast, the European day-ahead market is operated by several *power exchanges,* clearing the market by running a centralized algorithm (EUPHEMIA). This algorithm seeks to maximize the surplus of the bids while respecting a set of additional constraints. Those constraints are linked to the bid characteristics, but also with respect to the price (usually referred to as strict linear pricing). Indeed, some bids involve lumpiness decisions (fill or kill decisions for block bids and complex bids), and it is known that the pure welfare maximizing solution usually does not have linear prices to support it [14]. In Europe, the choice was made to impose the constraint that the clearing prices should necessarily support the accepted bids (i.e. all accepted bids are making a surplus), whereas some bids that could make a surplus may not be activated (the so-called paradoxically rejected bids).

<u>Belgium</u>

In contrast to ERCOT, the procurement of energy and reserve is separated in the Belgian market. The Belgian market has three notable stages of operation: reservation auctions, the day-ahead power exchange, and the nomination of reserve capacity.

The **reservation** auctions are the markets in which the TSO procures reserve capacity from BSPs on a portfolio basis. Primary and secondary reserve capacity is procured in weekly auctions. Tertiary reserve is procured in monthly auctions. ICH was procured on a yearly basis. The required total volumes are fixed for each category of reserve. These reserve capacities can also be procured in year-ahead auctions. These reserves are secured on a portfolio basis, meaning that no specific units are attached to the provision of these reserves at the reservation stage.

Following reservation, market participants can enter the day-ahead **power exchange**. Only BRPs can participate in the power exchange. The European day-ahead power exchange trades energy and implicitly trades transmission capacity, in the sense that transmission constraints are approximated by a zonal network model. The treatment of transmission is out of the scope of the present analysis, therefore this report will not expand on the treatment of transmission constraints in the day-ahead market clearing model[10]. The day-ahead power exchange uses a pricing rule the principle of which is based on *paradoxically rejected bids* (PRBs) that may be in the money. This is in stark contrast to the ERCOT approach, where the welfare maximizing commitment and dispatch is preserved, and where side payments are used in order to settle discrepancies between market clearing prices and fixed costs (e.g. minimum load or startup costs). By contrast, the European power exchange may discard a solution, even if it is profit maximizing, in order to ensure that another solution, with typically lower welfare, can be reached for which a uniform price ensures that the dispatch instructions are consistent with the market clearing price. This, in itself, is not necessarily relevant to scarcity pricing, because previous work has shown [1] that there exist indications that the welfare maximizing solution of the European power exchange may not be too different from the solution produced by paradoxically rejected bids. What is relevant to scarcity pricing is the fact that this pricing rule creates already today an extremely challenging problem for the market clearing algorithm (EUPHEMIA) which needs to produce a solution within 10 minutes. This issue is discussed further in section 5.2.1.

While the current available bids allow specifying some generation constraints for the resources (e.g. with block orders, complex orders, exclusive bids,…), they are still too simple to accurately represent the production set of most generation assets (e.g. no bid types for thermal generators except in Spain, no bid types for storage assets). The introduction of reserve products (as potentially foreseen by the integration of a day-ahead ORDC) and the co-optimization of energy and reserve would necessarily require to define new bid products where participants can offer a bulk of energy and reserve linked to their resources (i.e. in a similar way as the *three-part-offers* in ERCOT). A possible modification of EUPHEMIA bids is the redefinition of the complex block orders that could be allocated optimally between energy and reserve (probably keeping a minimum acceptance ratio for energy). The introduction of more complex new bids, as well as the joint clearing of energy and reserve, might impact EUPHEMIA performance: on a negative side, the size of the optimization problem increases as it includes more complex products and an additional reserve demand to be cleared; on a positive side, pricing reserve might actually ease the search of strict linear prices[11].

The European day-ahead exchange obeys *portfolio bidding*, meaning that resources are bid in the day-ahead market as portfolios. The disaggregation of the resulting market participant positions to individual resources takes place in the nomination stage, and is entirely decided by the owners of the resources.

The day-ahead energy market closes at 2pm the day before operations. This is followed by the so-called **nomination** stage. The nomination stage refers to the nomination of production, and the nomination of reserve. The nomination of production takes place at 3pm the day before, at this stage the technical maximum and set-point of *individual units* is announced by the owners of the resources to the system operator. The nomination of secondary and tertiary reserve capacity occurs after the nomination of production. The selection of the resources that will provide secondary and tertiary reserve is based on techno-economic criteria, which are described in detail in the CIPU documentation. ELIA specifically checks that the marginal cost declared by generators is consistent with the market values of the required fuels, with the details of bidding restrictions being stipulated in the CIPU contract [12]. The amount of capacity

[10] We note that transmission constraints enter in a natural way in the US market model that we propose later in this report. By contrast, the model is not tailored to capture gaming opportunities, such as the DEC game, in zonal network models.

[11] The difficulties related to the search of strict linear prices arise from a "missing market" issue: there exist products – the plant/order indivisibilities - for which there exist no prices (cf. [28]). Introducing reserve pricing in the search of price might make the missing market issues less severe and hence simplify the search of strict linear prices. As this additional clearing does not fully solve the issue, only numerical experiments can conclude on its effect on algorithm performance.

that BSPs offer for reserve nomination needs to be at least as much as the capacity that has been promised in the reservation auctions, since the TSO expects to be able to count on that capacity. The availability of tertiary reserve is verified through random checks and activation tests by the TSO, once specific units have been nominated. No payment is made in the nomination stage. The system operator checks that the nominated units can indeed deliver the promised reserves, according to the schedule that they have nominated.

The nomination stage of the Belgian market carries a legacy terminology for conventional resources (as opposed to renewable or demand-side resources), referred to as CIPU. CIPU units should be understood as conventional units, that form a subset (and the majority) of resources in the Belgian electricity system. The CIPU classification will be dropped in the coming years, but it is necessary to introduce it here, because this terminology is used repeatedly in the report.

1.2.6. Intraday Markets

Texas

ERCOT implements an hour-ahead reliability unit commitment for every operating period. There are no intraday adjustments after the RUC and before the hour-ahead RUC.

Belgium

The intraday price coupling of regions (IDPCR) is the intraday market, which takes place at 6pm the day before operations, and runs up to a few minutes before real time. The intraday market consists of a *continuous intraday market*. Bids are matched on a first-come-first-serve basis bilaterally. Products in the intraday market are differentiated according to whether they are divisible (limit orders) or not[12] (block orders).

1.2.7. Other Relevant Markets

Texas

Texas does not operate a capacity market, although ORDC is compatible with the existence of a capacity market.

Belgium

Strategic reserve. Belgium operates strategic reserve, which is an emergency measure that is intended to keep units that are planned for mothballing as standby capacity in order to serve peak load during winter months.

Secondary reserve market. There exists a secondary market for R1, R2 and R3 where market players can trade their obligations [9]. As described in the new balancing rules which entered into force in January 1, 2018, section 6.7, the secondary markets for R1, R2 and R3 cover both day-ahead and intraday timescales for CIPU as well as non-CIPU resources.

1.3. Implementation of ORDC in Texas

In this section we provide an overview of the design of the Texas electricity market, as it relates to scarcity pricing. This section is based on [6]. We ignore issues that relate to transmission, in order to focus the discussion on scarcity pricing.

1.3.1. Performance of ORDC Adder in the Texas Market

According to the Potomac 2017 state of the market report [15], the ORDC adder contributed to 0.24 $/MWh to the real-time price, which corresponds to less than 1% of the annual average real-time price in

[12] See https://www.belpex.be/trading/product-specification/.

Texas. This is due to the fact that the system was rarely short of reserves in 2017. The most notable impact of ORDC on real-time energy prices was in July and August of 2017. Currently, there is too much capacity in Texas, although the forecast is that this problem will correct itself.

1.3.2. Loss of Load Probability

In our proposed approach, we use historical imbalance data in order to estimate a distribution for computing the loss of load probability which is used for calculating the scarcity pricing adder. ERCOT, on the other hand, uses the so-called historical reserve error, which is the difference between the hour-ahead available reserves and the real-time available reserves. The amount of hour-ahead reserves measured in ERCOT is the difference between COP capacities and load forecasts. In this sense, 'free bids' (i.e. resources that are available in real time, even if they have not been cleared for reserve capacity) are counted towards real-time available capacity. The real-time reserve is computed as the difference between the measured generator capacities and their SCED set-points.

ERCOT assumes a normal distribution for the reserve error, and estimates a mean and standard deviation for characterizing this distribution. The parameters of the normal distribution are changed in 4-hour blocks for every season. The ERCOT ORDC uses a value of lost load of 9000 $/MWh, and imposes a minimum operating reserve requirement of 2000 MW. This is referred to as the *minimum contingency level* (MCL).

1.3.3. ORDC Adder

ERCOT computes one adder for resources that can respond within 30 minutes (which is called the *real-time online reserve price adder*), and one adder for resources that can response within 60 minutes (which is called the *real-time offline reserve price adder*). Both of these adders are computed every five minutes, i.e. every time that SCED runs. However, because the Texas real-time market settles transactions every 15 minutes (as opposed to every five minutes), the real-time online and offline reserve price adders are averaged over three time intervals into the *real-time reserve price for online* and *for offline reserve* respectively.

ERCOT currently does not perform a real-time co-optimization of energy and reserves. Nevertheless, the Texas market computes real-time reserve prices by virtue of the ORDC adder. Resources that can respond within 30 minutes are paid the online reserve price, while resources that can respond within 60 minutes are paid the offline reserve price.

ERCOT uses hourly estimates of the reserve error for the offline reserve adder, and uses an assumption of independent increments for estimating the online reserve price. Concretely, if μ and σ are the hourly mean and standard deviation for the offline adder, the online adder is computed by assuming a mean of 0.5μ and a standard deviation of 0.707σ. This is also explained in [3].

As mentioned previously, the real-time settlement point price is the sum of the locational marginal price and the reserve price adder for every 15-minute interval. The adder is computed as the average adder over the 3 SCED runs that take place within a 15-minute interval. The offline reserve adder is actually not used for correcting the settlement point price (i.e. the energy price). We return to this issue in section 7.6.

In order for ERCOT to balance its budget when introducing the ORDC adder, the reserve price adder is effectively passed on to loads. Recall in the description of the ERCOT day-ahead market that reserve responsibilities are shared among loads on a proportional basis. Therefore, even though ERCOT procures the reserve, since QSEs are responsible for self-supplying this reserve, effectively loads pay for this reserve. Thus, an adder that increases the reserve prices is effectively paid for by loads. Of course, loads can also provide reserve to the market, so they also stand to benefit from the mechanism. This is explained in the numerical examples of section 7.5.4.

1.3.4. How Many Adders?

Texas has the following reserve products: regulation up (responds in 3-5 seconds), regulation down (responds in 3-5 seconds), restoration reserve services / RRS (responds within 10 minutes), and non-spin (responds within half an hour). Day-ahead prices for these reserves are posted with hourly resolution.

The ORDC adder is broken into two components ([6], slide 29), one component that corresponds to capacity that can be made available immediately (the online adder), and one component that corresponds to capacity that can respond in 30 minutes (the offline adder). We therefore interpret the online adder as being applicable to reserve capacity that can be offered by regulation and restoration reserve, and the offline adder as being applicable to reserve capacity that can be offered by all reserves (including non-spinning reserve). The change in the energy price is driven by the online adder, as explained in the previous paragraph and further justified in section 7.6.. We therefore effectively have three adders.

1.4. ORDC in Other Systems

1.4.1. PJM

PJM is currently moving forward with the implementation of ORDC [16] in the day-ahead[13] and real-time market. An important ongoing debate in PJM is about whether the ORDC should be present in the day-ahead market clearing model[14]. The argument in favor of introducing it to the day-ahead market clearing is that it would better align real-time and day-ahead pricing.

The construction of an ORDC in PJM would involve two steps. The first step is the computation of a minimum reserve requirement (Texas has a corresponding quantity in the real-time ORDC, which amounts to 2000 MW). The idea is to compute this quantity by using forced outage rate data, load forecast error data, and adapting this quantity by season and also by the occurrence of extreme weather events. The second step is the construction of the part of the curve that would relate to loss of load probability, and would rely on the same data.

PJM currently operates two tiers of reserve products. Tier 1 has a 10-minute response time, an obligation to respond, is subject to a non-compliance penalty, and is paid for response to an event. It is available generation capacity that is synchronized and can be loaded within 10 minutes. Tier 2 has a 10-minute response time, an obligation to respond, is subject to a non-compliance penalty, and is paid the market clearing price regardless of deployment. It consists of resources that are committed or dispatched out of merit in order to provide reserves. In reforming its reserve market, PJM is interested in consolidating tier 1 and tier 2 reserves.

PJM plans to propose ORDC for all reserve products (tier 1 and tier 2). PJM uplift is currently 300 thousand dollars per day on average. The introduction of ORDC is expected to reduce this uplift. The distribution of uncertainty that will be used for the computation of the ORDC will be based on real-time load forecast errors, renewable (solar and wind) forecast errors, and conventional generator failures. This effort is driven to a significant extent by the projected increase of wind nameplate capacity installation to 36,159 MW by 2029 (it is currently at approximately 10 GW).

PJM currently performs real-time dispatch with 5-minute frequency. The penalty factor for falling below a capacity of 1500 MW is currently set at 850 $/MWh, and to 300 $/MWh for falling below a capacity of 1.7 GW. The goal is to introduce a downward sloping ORDC curve in the real-time dispatch. The curve jumps to 2000 $/MWh at the 1.5 GW level.

---

[13] https://www.rtoinsider.com/pjm-ferc-energy-price-111313/

[14] See https://www.pjm.com/-/media/committees-groups/task-forces/epfstf/20180608/20180608-item-03c-day-ahead-scheduling-reserve-operating-reserve-demand-curve.ashx.

In addition to consolidating its tier 1 and tier 2 products, PJM is moving to three types of reserve requirements. These are synchronized, primary, and 30-minute reserve. In implementing ORDC, each reserve product will be associated with a penalty factor, a minimum requirement, and an associated probability distribution. The idea is that: (i) spinning reserve will contribute to all requirements; (ii) non-spinning reserve will contribute to primary reserve and 30-minute reserves, and secondary reserve will contribute to the 30-minute requirement. PJM proposes to base its reforms on a 30-minute look ahead for uncertainty for the synchronized and primary reserve requirements and a 60-minute look ahead for the 30-minute reserve requirement. For the first half of the 60-minute period, the 30-minute uncertainty will apply only to the valuation of two types of 10-minute reserves. For the second half of the period, the full one-hour uncertainty will apply to the combined levels of 10-minute and 30-minute reserves.

1.4.2. UK

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) in Belgian market terminology. It appears that the UK market does not involve real-time reserve capacity payments, but only real-time energy payments.

Balancing market bids can adjust their activation cost in real time. By contrast, STOR receives so-called availability payments which can be interpreted as activation costs for energy, but the value of which 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 [17]). In this sense, STOR unit owners do not have the freedom of adapting their real-time bids for energy. This creates a challenge in creating a real-time energy signal which accurately reflects scarcity.

For this reason, 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, [17]). The ORDC is constructed by using the product of VOLL with loss of load probability as a function of available reserve capacity. The original estimate for VOLL was equal to 3000 British pounds per MWh, and was planned to be raised to 6000 British pounds per MWh by early winter 2018/2019 (article 3.55, [17]).

The intent of OFGEM is to use a dynamic LOLP for computing ORDC, as is the case in Texas, and as also recommended in this report for the Belgian market. The LOLP uses both STOR capacity, as well as free balancing bids, as we also recommend in the present report. The LOLP is computed using data of the current balancing interval, and publishes the LOLP information shortly thereafter to balancing market participants. Indicative LOLPs are recommended to be published shortly in advance of real time (e.g. 4, 3 or 2 hours ahead of real time) in order to reduce the risk of balancing market participants. Advance information is not required in central dispatch systems such as Texas or PJM, since it is the system operator that dispatches resources in a way that is automatically consistent with real-time prices.

2. Design Options

The goal of a scarcity pricing mechanism is to create a signal for quantifying the value of reserve in keeping the loss of load probability in check. This additional value is reflected in the real-time price of energy and the real-time price of reserve capacity. Under conditions of perfect arbitrage, it would be expected that this adjustment to the real-time value of energy and reserve would be back-propagated to the forward price of energy and reserve.

Section 2.1 discusses the exact mechanism by which this back-propagation occurs under different market designs. We then delineate two market design prototypes that envelope the full range of market design options that we consider. In section 2.2 we present the first market design, which we refer to the as the US model, and which is a *conceptual approximation* of the existing market design in ERCOT. In section 2.3 we present the second market design, which we refer to as the Belgian model, and which is a *conceptual approximation* of the current Belgian market. These two envelope models differ along three principal features:

- whether or not reserve capacity is traded in real time,
- whether or not virtual trading is allowed, and
- whether or not energy is cleared simultaneously with reserve in the day-ahead market.

The introduction of each of these features creates a family of market design options that we present in sections 3, 4, and 5.

2.1. Back-Propagation / Convergence of Energy Prices

For power suppliers, the marginal values of generation assets are ultimately their value at the final settlement period (the closest to generation time). All the other market activities happening before this (intraday, day-ahead, futures trading) are necessary to provide hedges, support provisional planning and help in key decisions/commitments of the resources (e.g. fuel contracting, generation scheduling, maintenance). The services contracted by the network operator ensure stable final delivery and prevent costly rolling blackouts. In this sequence, the day-ahead market has today a specific role: while still being a forward market for real-time activities, it also corresponds to a milestone where key decisions on the commitment of resources need to be taken (as generators are not fully flexible and their production must accordingly be planned in advance), providing an initial production/consumption plan. Real-time activities should then correct for all the unforeseen deviations that ensue. Due to the inflexibility of certain resources, the space of actions in real time is more limited (plants are already committed and start-up/loading can take some time). Hence it is important to ensure an efficient provisional planning in the day-ahead time frame, recognizing the possible deviations happening afterwards in real time. This ideal sequence of decisions can only efficiently work if there are no structural deviations between the economic incentives in day ahead and real time, i.e. if there is convergence[15] of prices between day-ahead and real-time operations. Regulators in the US recognize this price convergence as a sign of efficient day-ahead/real time market operations, and continuously monitor this statistic in their state of the market reports[16]. This price convergence is also crucial in the context of scarcity pricing. The scarcity signal should indeed be back-propagated to earlier markets in order to incentivize long-term decisions on resources. In terms of

[15] This convergence should not occur at every time delivery, as unforeseen deviations can indeed lead to differences in price, but rather in expectation, a condition for a well-functioning forward market where risk premia should be low.

[16] See, for example, https://www.potomaceconomics.com/wp-content/uploads/2018/05/2017-State-of-the-Market-Report.pdf, http://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2018/2018q2-som-pjm-sec3.pdf.

market design, there exist several ways to ensure/promote this price convergence. If none of the following options are present in the design, obliging resources to bid their marginal cost in the day-ahead would certainly suppress back-propagation.

(i) Allow opportunity cost bidding in the day-ahead market. For efficient operations, generating resources should bid their opportunity cost in the day-ahead market, i.e. anticipating their value in real time. This value is a function of the expected price distribution in real time and of the flexibility of the resource to adapt to this price signal. Therefore, efficient day-ahead bids need not be linked to the physical characteristics of a resource, i.e. to its marginal cost, but rather to its opportunity cost. The fact that European day-ahead markets permit portfolio bidding provides market actors with significant flexibility to internalize such factors in their bids.

(ii) Introduce virtual trades. Allow market participants to choose the amount of power contracted in day-ahead versus real-time and arbitrage between the two markets.

(iii) Align as much as possible the market organization/pricing mechanism in day ahead and real time. A significant difference in terms of market organization can introduce structural differences in terms of prices. In terms of scarcity pricing, this is why there has been a debate on the need to implement ORDC both in the day-ahead as well as the real-time market.

2.1.1. Price Convergence in Texas

According to the state of the market report [15], the average real-time and day-ahead prices in Texas in 2017 were both at 26 $/MWh.

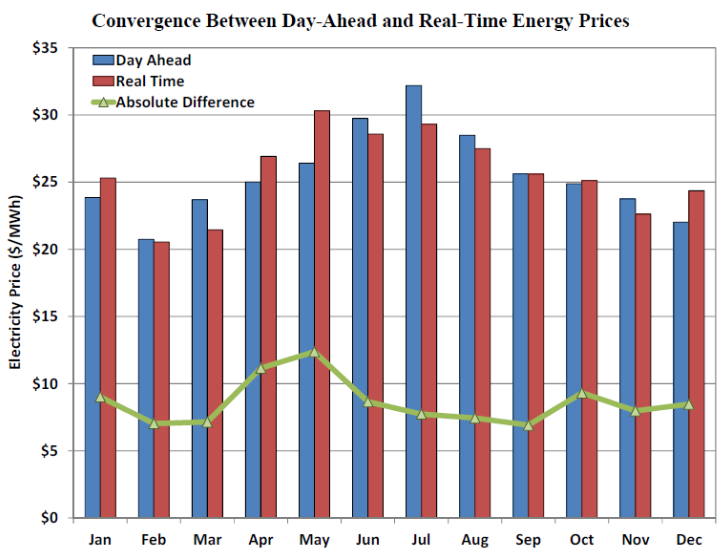


Figure 2: Price convergence in the ERCOT market for 2017. Source: [15].

Figure 2 presents the monthly average day-ahead and real-time prices for the Texas market in 2017. The green figure is the absolute difference. The month of July presents a notable risk premium, in the sense that the average day-ahead price is notably larger than the average real-time price. According to Potomac: "*Risk is lower for loads purchasing in the day-ahead market and higher for generators selling day-ahead. The higher risk for generators is associated with the potential of incurring a forced outage and having to buy back energy at real-time prices. This explains why the highest premiums occurred during the summer months in 2017 with the highest relative demand and highest prices.*".

2.1.2. Price Convergence in Belgium

We compare the Belgian market data to the data presented in the Potomac report. The price data that we use was provided to us by the CREG for the first and second scarcity pricing study [1], [18]. The average day-ahead and real-time prices are reported in Table 4. We note that the average real-time price is quite close to the average day-ahead price for the first study (which spans 26 months), with the average day-ahead prices being slightly lower. We note that the average real-time price is notably lower in the second study (which spans 7 months).

	Data start	Data end	Average DA [€/MWh]	Ave imb up [€/MWh]	Ave imb down [€/MWh]
First study	01/2013	02/2015	44.3	45.5	44.4
Second Study	09/2015	03/2016	38.9	36.2	35.4
Aggregate			42.3	43.5	42.5

Table 4: Day-ahead and average imbalance prices in Belgium for the time horizons spanned by the two previous studies on scarcity pricing.

Figure 3 reproduces the analogs of the Potomac report for Belgium. For the aggregate data, the average day-ahead price is very close to the negative imbalance price. Interestingly, the average positive imbalance price is on average 1 €/MWh higher than the negative imbalance price, which implies that the dual pricing mechanism of balancing in Belgium has a non-negligible effect. For the aggregate data set, the month with the highest average monthly deviation was April 2013, at 15.5 €/MWh. The month with the lowest average monthly deviation was June 2013, at -13.6 €/MWh. In contrast to the Texas market, the Belgian market is most heavily loaded in the winter. This does not necessarily coincide with the most positive deviations between day-ahead prices and real-time prices (this should be contrasted with the Texas market, where the greatest positive deviation coincided with the heaviest loading of the system). The absolute difference, as a percentage of average prices, is comparable to that of the Texas market for the data of the first study, i.e. the absolute difference corresponds to approximately half of the average market price. By contrast, the absolute differences are more pronounced for the data of the second study compared to the Texas data.

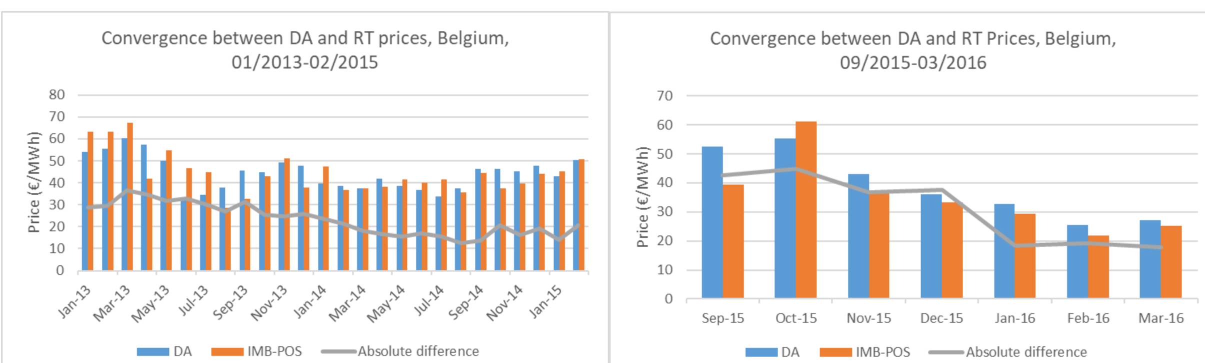


Figure 3: Price convergence in the Belgian market. Left: January 2013 – February 2015. Right: September 2015 – March 2016.

2.2. The US Model (SCV)

The blueprint of a market design for which scarcity pricing is conceived, which is depicted in Figure 4, comprises of the following set of agents:

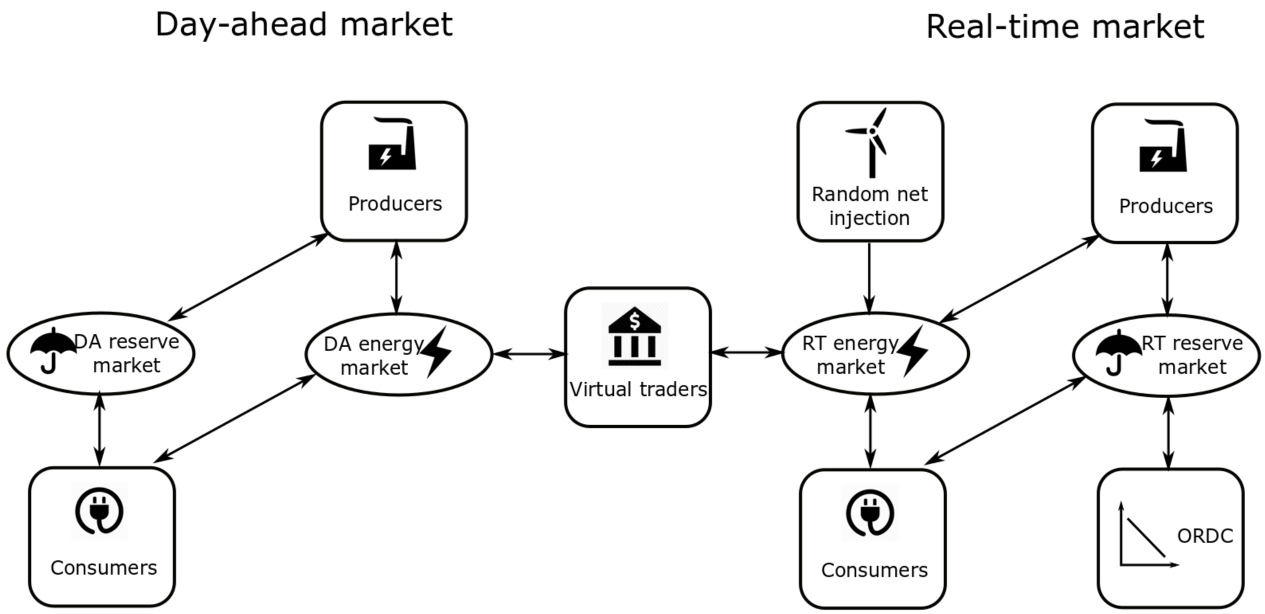


Figure 4: A blueprint of the US two-settlement system (SCV).

- generators that own plants, some of which may be technically capable of offering reserves,
- loads,
- a system operator, and
- virtual traders that do not necessarily own physical capacity, but are rather arbitraging price differences between real-time and earlier markets.

Two products are traded in both the day-ahead and real-time market:

- energy, and
- reserve, which may be further differentiated depending on its quality (response time).

2.2.1. Real-Time Market Equilibrium

We describe the economic equilibrium in the real-time market, where we assume that agents are price takers.

<u>Generators</u>

Generators participate in the real-time energy and reserve market. They seek to maximize profits, which consist of revenues earned for selling power and reserve capacity:

$$\max_{p\geq 0, r\geq 0} \Pi_{g,\omega}^{RT} = \lambda_{\omega}^{RT} \cdot p_{g,\omega}^{RT} + \lambda_{\omega}^{R,RT} \cdot r_{\omega}^{RT} - C_g \cdot p_{g,\omega}^{RT},$$

where λ_{ω}^{RT} is the real-time price of power, $\lambda_{\omega}^{R,RT}$ is the real-time price of reserve capacity, $p_{g,\omega}^{RT}$ is the real-time amount of power production, and $r_{g,\omega}^{RT}$ is the amount of real-time reserve capacity *after*[17] reserves have been activated in response to an imbalance. The parameter C_g corresponds to the marginal cost of a generator. The subscript ω indicates that the real-time conditions are uncertain (due to forecast errors,

[17] The qualification 'after' is important – this issue is revisited later in the report in the description of the Belgian market model.

outages of system components, etc.), and therefore decisions of power production and reserve capacity allocation are in response to a realization ω of the state of the world.

Generators allocate their limited capacity between power and reserves:

$$\left(\alpha_{g,\omega}^{G,RT}\right): p_{g,\omega}^{RT} + r_{g,\omega}^{RT} \leq P_{g,\omega}^{RT,+} \cdot y_g,$$

where $P_{g,\omega}^{RT,+}$ is the amount of power generation capacity that is available in real time and y_g is the decision of committing a unit. We assume that this decision is determined in the day-ahead time frame, and is therefore a fixed parameter when entering real time. Dual multipliers, such as $\alpha_{g,\omega}^{G,RT}$, are indicated to the left of the corresponding constraints.

The amount of reserve capacity that can be made available is limited by the ramp rate of a unit and the response time of the relevant reserve product:

$$\left(\beta_{g,\omega}^{G,RT}\right): r_{g,\omega}^{RT} \leq R_g,$$

where the total reserve capacity that a generator can make available is indicated by R_g.

<u>Loads</u>

Loads can generally participate in both the energy and reserve market (and we account for this in our quantitative analysis), although we do not represent the contribution of loads to reserve in this section in order to keep the exposition as clear as possible without obscuring the key ideas. Loads that participate in the real-time market are represented by the following profit maximization:

$$\max_{d \geq 0} \Pi_{l,\omega}^{RT} = V_l \cdot d_{l,\omega}^{RT} - \lambda_{\omega}^{RT} \cdot d_{l,\omega}^{RT},$$

where V_l is the valuation of a consumer, and $d_{l,\omega}^{RT}$ is its power consumption.

The power consumption of consumers is limited by their total real-time demand:

$$\left(\alpha_{l,\omega}^{L,RT}\right): d_{l,\omega}^{RT} \leq D_{l,\omega}^{RT,+},$$

where $D_{l,\omega}^{RT,+}$ is the real-time demand.

<u>System operator</u>

The system operator participates in the real-time reserve market. It procures reserve capacity according to an Operating Reserve Demand Curve (ORDC):

$$\max_{d^R \geq 0} \sum_{l \in RL} \left(V_l^R - \lambda_{\omega}^{R,RT}\right) \cdot d_{l,\omega}^{R,RT}$$

$$\left(\alpha_{l,\omega}^{R,RT}\right): d_{l,\omega}^{R,RT} \leq D_l^{R,RT}$$

where (V_l^R, D_l^R) are the price-quantity pairs of the system operator ORDC.

<u>Market clearing</u>

The market clearing constraint for the real-time energy market can be written out as follows:

$$\sum_{g\in G} p_{g,\omega}^{RT} = \sum_{l\in L} d_{l,\omega}^{RT},$$

where G is the set of generators, and L is the set of loads.

The market clearing constraint for the real-time reserve market is expressed as follows:

$$\sum_{g\in G} r_{g,\omega}^{RT} = \sum_{l\in RL} d_{l,\omega}^{R,RT}$$

where RL is the set of segments in the system operator demand function for operating reserve.

This model can be enriched with various additional features, such as (i) technical minimum constraints, (ii) multiple reserve products with different response times, and (iii) detailed unit commitment constraints. We account for these features in our numerical analysis, but simplify the exposition here in order to keep the minimal number of model features that convey the key messages of the analysis.

2.2.2. Day-Ahead Market Equilibrium

The day-ahead market in a US-style two-settlement system is an auction that simultaneously clears energy and reserve. In developing an economic equilibrium model for the two-settlement system, we consider agents that may be averse to risk. We represent their aversion to risk by a risk function $\mathcal{R}$, which maps uncertain real-time profits Π_ω^{RT} to agent utility. For example, in the case of risk-neutral agents, this risk function corresponds to the expectation over real-time profits:

$$\mathcal{R}(\Pi_\omega^{RT}) = \mathbb{E}[\Pi_\omega^{RT}].$$

Our model is sufficiently general to account for risk aversion, but the numerical treatment of the resulting equilibrium model is challenging and therefore the case study of the Belgian market will be conducted under the assumption of risk-neutral agents.

<u>Generators</u>

As in the case of the real-time market, generators trade power and reserves in the day-ahead market. The goal of generators in the day-ahead market is to take positions that maximize their profits, where profits include the immediate rewards that the generators earn in the day-ahead market, as well as the risk-adjusted profits that the agents will actually face in the real-time market:

$$\max_{y\geq 0,p,r\geq 0} \lambda^{DA}\cdot p_g^{DA} + \lambda^{R,DA}\cdot r_g^{DA} - K_g\cdot y_g + \mathcal{R}_g\left(\Pi_{g,\omega}^{RT} - \lambda_\omega^{RT}\cdot p_g^{DA}\right),$$

where the real-time profits $\Pi_{g,\omega}^{RT}$ have been defined above in the exposition of the real-time equilibrium. The energy and reserve prices in the day-ahead market are indicated as λ^{DA} and $\lambda^{R,DA}$ respectively.

Note that the on-off status of a unit is decided in this day-ahead time frame, and indicated by the variable y_g. We use a linear relaxation of the unit commitment variables, in order to focus the analysis on reserve pricing and disentangle it from the pricing of non-convexities:

$$(\delta_g): y_g \leq 1$$

US day-ahead energy markets permit virtual trading in power, but not in reserve. This implies that agents can take financial positions in the *energy* market which are not restricted by the physical capabilities of their assets, however the positions that they take in *reserve* markets are limited by their response capability. We model this requirement by the following constraint:

$$(\beta_g^{G,DA}): r_g^{DA} \leq R_g.$$

Note that, in order to allow for virtual trading, no explicit constraint is imposed on the amount of power p_g^{DA} that is traded.

Loads

Loads participate in the energy market, whereas we assume that they do not participate in the reserve market in order not to clutter the notation. Note, however, that the model can easily be extended to allow for the participation of loads in the reserve market. Since loads are also risk averse in general, their profit maximization problem in the day-ahead market is described as follows:

$$\max_{d \geq 0} V_l \cdot d_l^{DA} - \lambda^{DA} \cdot d_l^{DA} + \mathcal{R}_l\left(\Pi_{l,\omega}^{RT} + \lambda_\omega^{RT} \cdot d_l^{DA}\right),$$

where the real-time profits, $\Pi_{l,\omega}^{RT}$, correspond to the objective function of the loads in real time.

System operator

The system operator may participate in the day-ahead market through an operating reserve demand curve. This day-ahead ORDC needs to be designed in a way that reflects the fact that certain resources that can offer reserve need to be committed in the day-ahead time frame.

$$\max_{d^R \geq 0} \sum_{l \in RL} (V_l^R - \lambda^{R,DA}) \cdot d_l^{R,DA}$$

$$(\alpha_l^{R,DA}): d_l^{R,DA} \leq D_l^{R,DA}$$

In this formulation, $\left(V_l^R, D_l^{R,DA}\right)$ represent the price-quantity pairs of the day-ahead ORDC.

Market clearing

The market equilibrium conditions for the energy and reserve market respectively are given as follows:

$$\sum_{g \in G} p_g^{DA} = \sum_{l \in L} d_l^{DA}$$

$$\sum_{g \in G} r_g^{DA} = \sum_{l \in RL} d_l^{R,DA}$$

2.2.3. Arbitrage and Back-Propagation

Arbitrage between energy and reserve capacity

For an interior solution ($p_{g,\omega}^{RT} > 0$ and $r_{g,\omega}^{RT} > 0$), the generator profit is given by:

$$\Pi_{g,\omega}^{RT} = \left(\lambda_\omega^{RT} - C_g\right) \cdot p_{g,\omega}^{RT} \cdot P_{g,\omega}^{RT,+} = \lambda_\omega^{R,RT} \cdot P_{g,\omega}^{RT,+},$$

with the profit margin being determined as the scarcity value of generation capacity:

$$\lambda_\omega^{RT} - C_g = \alpha_{g,\omega}^{G,RT}.$$

The conclusion is that the price of reserve drives the profit of the generator, and is connected to scarce generator capacity, as reflected by the multiplier $\alpha_{g,\omega}^{G,RT}$. By consequence, energy prices will follow in lock step with reserve prices (with a constant equal to the marginal cost of the marginal unit separating them) and determine the profit margin of those generators that offer reserve.

Arbitrage between Day Ahead and Real Time

The optimality conditions of the producer problem lead to the following no-arbitrage condition for energy and reserve prices, in the case of risk-neutral agents[18] and non-binding ramp constraints ($\beta_g^{G,DA} = 0$):

$$\lambda^{DA} = \mathbb{E}[\lambda_\omega^{RT}]$$

$$\lambda^{R,DA} = \mathbb{E}\left[\lambda_\omega^{R,RT}\right]$$

These relations can be generalized for risk-averse agents, by replacing the expectation operator with an expectation over the risk-neutral measure of agents[19]. These no-arbitrage conditions express the intuitive fact that any deviation between average real-time prices and day-ahead prices (in either energy or reserve) can lead to profit opportunities, and will therefore induce agents to exploit these opportunities until equilibrium is restored. This is the essence of back-propagation:

- The fact that agents exploit arbitrage opportunities between energy and reserve capacity in real time results in a tight coupling between real-time energy and reserve prices.
- Reserve prices can be driven by ORDC, so as to reflect scarcity in the system, as indicated by loss of load probability. Since real-time energy prices follow in lock step, they too are lifted in response to scarcity.
- The fact that agents can exploit arbitrage opportunities between day-ahead and real-time markets implies that the impacts of ORDC on the real-time market will not only apply to incremental changes in real-time power and reserve capacity, but will eventually see their way through to day-ahead prices, which impact the *entire* quantity of power and reserve capacity. This will create an investment signal for building out resources that can offer reserve capacity.

2.3. The Belgian Model (REP)

Our current proxy of the Belgian market is presented in Figure 5. There are three principal differences between the existing Belgian design and the US design:

- The Belgian electricity market does not operate a real-time market for reserve capacity.
- No virtual trading is permitted in the Belgian market.
- The trading of reserve capacity takes place before the trading of energy in the Belgian electricity market.

[18] These equalities rely on virtual trading. The introduction of constraints on trading will affect the back-propagation, as we describe in more detail later in the report.

[19] The risk-neutral measure is an agent-specific way of weighting the probability of uncertain outcomes, which tends to place higher weight on adverse outcomes and give rise to a risk premium.

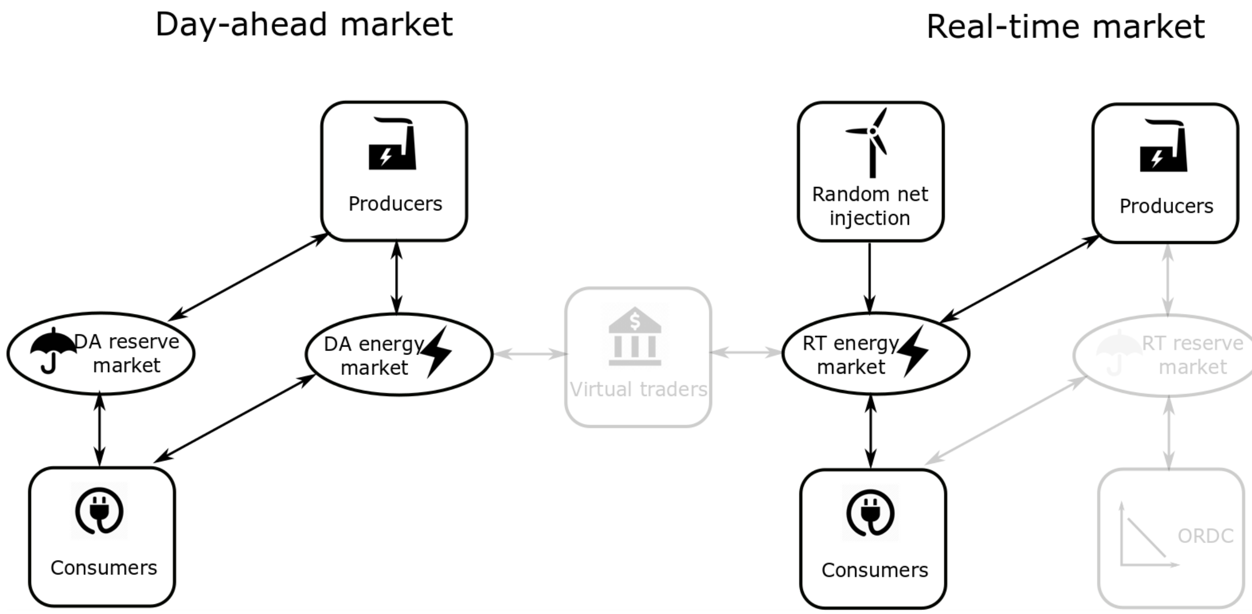


Figure 5: A blueprint of the existing Belgian market design (REP).

2.3.1. Real-Time Market Equilibrium

In the absence of a real-time market for reserve capacity, the generator profit maximization problems are written out as follows:

$$\max_{p,s\geq 0} \Pi^{RT}_{g,\omega'} = \lambda^{RT}_{\omega'} \cdot p^{RT}_{g,\omega'} - C_g \cdot p^{RT}_{g,\omega'} - \epsilon^{+}_{g} \cdot s^{RT,+}_{g,\omega'} - \epsilon^{-}_{g} \cdot s^{RT,-}_{g,\omega'}$$

$$\left(\alpha^{G,RT}_{g,\omega'}\right): p^{RT}_{g,\omega'} \leq P^{RT,+}_{g,\omega'} \cdot y_{g,\omega}$$

$$\left(\gamma^{G,RT,+}_{g,\omega'}\right): p^{RT}_{g,\omega'} - p^{DA}_{g,\omega} - s^{RT,+}_{g,\omega'} \leq 0$$

$$\left(\gamma^{G,RT,-}_{g,\omega}\right): p^{DA}_{g,\omega} - p^{RT}_{g,\omega'} - s^{RT,-}_{g,\omega'} \leq 0$$

The slack variables $s^{RT,+}_{g,\omega'}, s^{RT,-}_{g,\omega'}$ have been introduced in order to model the upward and downward deviation of generators from their day-ahead positions[20]. These deviations may be explicitly or implicitly penalized by ϵ^{+}_{g} and ϵ^{-}_{g} respectively. An implicit penalization corresponds to a “gentlemen’s agreement” between agents and system operators that the former shall “do their best” to maintain balance in real time. One form of explicit penalization are imbalance charges, as we describe them in section 1.2.4.

Concretely, consider a generator which is marginal ($p^{RT}_{g,\omega'} < P^{RT,+}_{g,\omega'}$), and is in positive imbalance ($s^{RT,+}_{g,\omega'} > 0$). Then it follows that:

$$\lambda^{RT}_{\omega'} = C_g + \epsilon^{+}_{g}.$$

[20] In practice, deviations between day-ahead and real-time positions are penalized on the level of the portfolio of a BRP. Since we do not have access to precise ownership data in our study, we use resource-specific penalties as a proxy of the penalty on portfolio level.

The interpretation is that a generator will only be marginal if it can break even both on its fuel cost as well as its imbalance charges. We can adapt the parameters $\epsilon_g^{+/-}$ to model various degrees to which the TSO penalizes real-time imbalances.

The load profit maximization in real time can be modeled similarly:

$$\max_{d \geq 0} \Pi_{l,\omega'}^{RT} = V_l \cdot d_{l,\omega'}^{RT} - \lambda_{\omega'}^{RT} \cdot d_{l,\omega'}^{RT}$$

$$(\alpha_{l,\omega'}^{L,RT}): d_{l,\omega'}^{RT} \leq D_{l,\omega'}^{RT,+}$$

The equilibrium of the real-time energy market is enforced by the following constraint:

$$\sum_{g \in G} p_{g,\omega'}^{RT} = \sum_{l \in L} d_{l,\omega'}^{RT}.$$

Note that, in this model, reserve capacity is not traded in real time. Therefore, the system operator is absent from the real-time market in terms of procuring reserve capacity. Also note that, in this model, real-time available reserve capacity *after* activation for balancing is not bound in any way to reserve capacity that is committed in the day-ahead market. We revisit this issue when we discuss the results of our models in section 6.

2.3.2. Day-Ahead Market Equilibrium

The first notable difference of the Belgian *day-ahead* market from the US design is the separation of energy and reserve capacity. We represent this separation using the following scenario tree, which describes the gradual revelation of information between reserve and energy day-ahead auctions.

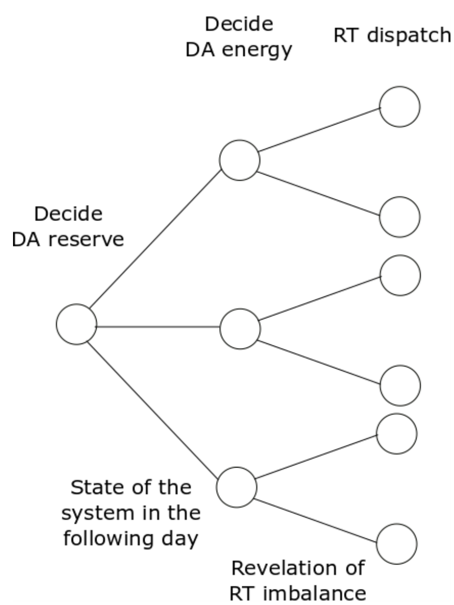


Figure 6: The sequence of reserve and energy day-ahead auctions in the Belgian design.

The sequence of events in this clearing process is as follows:

- What is observed in the third stage is the realization of real-time renewable supply and real-time demand, which from the point of view of the agents implies a real-time price for energy. Decisions and parameters that are revealed in this stage are indexed by ω' and belong to the set of outcomes indicated in the third stage of the scenario tree of Figure 6 (i.e. the six nodes presented in the third stage of the tree).
- The uncertainty which is observed in stage 2 is the state of the world which will influence the real-time imbalances of stage 3, and which from the point of view of the agents implies an observable price for day-ahead energy. Decisions and parameters that are revealed in this stage are indexed by ω and belong to the set of outcomes indicated in the second stage of the scenario tree of Figure 6 (i.e. the three nodes presented in the second stage of the tree).
- The first stage reveals to the agents the price of day-ahead reserve capacity, as a result of their competition.

<u>Day-ahead energy market</u>

Generators that trade power in the day-ahead energy market decide on the commitment of their generators, as well as the power that they wish to trade in the day ahead. Concretely, generators solve the following profit maximization:

$$\max_{y\geq 0, p\geq 0} \Pi_{g,\omega}^{DA} = \lambda^{DA} \cdot p_{g,\omega}^{DA} - K_g \cdot y_{g,\omega} + \mathcal{R}2_{g,\omega}\left(\Pi_{g,\omega'}^{RT} - \lambda_{\omega'}^{RT} \cdot p_{g,\omega}^{DA}\right).$$

$$(\delta_{g,\omega}): y_{g,\omega} \leq 1$$

$$\left(a_{g,\omega}^{G,DA}\right): p_{g,\omega}^{DA} + r_g^{DA} \leq P_g^{DA,+} \cdot y_{g,\omega}$$

The risk function $\mathcal{R}2$ maps the risky real-time profit of generators to a risk-adjusted day-ahead payoff. Note that this model suppresses *virtual trading*, in the sense that generators may only trade in the day-ahead energy market up to the level of their generation capacity. At this stage of the market, generators must decide whether or not to activate their units, and to what extent they wish to hedge in the day-ahead energy market. Note that their decision about whether or not to commit a unit is influenced by prior decisions that they have made in the day-ahead reserve market. Concretely, if the generators have been cleared for offering reserve capacity in the day-ahead market, they are required to turn a unit on in order to deliver that reserve.

Similarly, loads maximize their profits in the day-ahead energy market by solving the following profit maximization:

$$\max_{d\geq 0} \Pi_{l,\omega}^{DA} = -\lambda^{DA} \cdot d_{l,\omega}^{DA} + \mathcal{R}2_{l,\omega}\left(\Pi_{l,\omega'}^{RT} + \lambda_{\omega'}^{RT} \cdot d_{l,\omega}^{DA}\right).$$

$$\left(a_{l,\omega}^{L,DA}\right): d_{l,\omega}^{DA} \leq D_l^{DA,+}$$

The market-clearing constraint of the energy market is expressed as follows:

$$\sum_{g\in G} p_{g,\omega}^{DA} = \sum_{l\in L} d_{l,\omega}^{DA}.$$

<u>Day-ahead reserve market</u>

The reserve market involves generators and the system operator, and takes place in the first stage of the scenario tree of Figure 6. The profit maximization problem of generators is expressed as follows:

$$\max_{r\geq 0} \lambda^{R,DA} \cdot r_g^{DA} + \mathcal{R}1_g\left(\Pi_{g,\omega}^{DA}\right).$$

$$\left(\beta_g^{G,DA}\right): r_g^{DA} \leq R_g$$

In this model, $\mathcal{R}1_g$ corresponds to the risk-adjusted day-ahead profits of generators. These profits correspond to the objective function of the second-stage generator problem, which is described in the previous paragraph.

The system operator procures reserve capacity in the day-ahead market through a demand curve (which may be inelastic, i.e. a fixed reserve requirement, as is currently the case in Belgium).

$$\max_{d^R\geq 0} \sum_{l\in RL} (V_l^R - \lambda^{R,DA}) \cdot d_l^{R,DA}$$

$$\left(\alpha_l^{R,DA}\right): d_l^{R,DA} \leq D_l^{R,DA}$$

We complete the model by including the market clearing constraint for the day-ahead reserve market:

$$\sum_{g\in G} r_g^{DA} = \sum_{l\in RL} d_l^{R,DA}.$$

2.3.3. Formation of Reserve Prices

A major difficulty with the absence of a real-time reserve market is that it becomes difficult to value reserve precisely. To illustrate this, consider a generator that is offering reserve in the day-ahead market: $r_g^{DA} > 0$. Then, the reserve price is given by the following expression:

$$\lambda^{R,DA} = \beta_g^{G,DA} + \mathbb{E}\left[a_{g,\omega}^{G,DA}\right].$$

Recall that $\beta_g^{G,DA}$ is related to limited *capacity*, while $a_{g,\omega}^{G,DA}$ is related to limited *ramp rate*. This implies that the formation of the reserve price relies on day-ahead scarcity conditions. However, it is only in real time when extreme outcomes of forecasting errors or outages are revealed, therefore it is much less likely that scarcity conditions emerge in day-ahead market clearing. This tends to depress reserve prices.

An alternative representation of the Belgian market is to impose the requirement that the amount of real-time reserve capacity after the activation of reserves within a given imbalance interval should be at least as much as the amount of reserves that generators have committed to offer in the day-ahead market. This requirement is expressed by the following constraint in the real-time market equilibrium:

$$\left(\gamma_{g,\omega'}^{G,RT}\right): r_g^{DA} - r_{g,\omega'}^{RT} \leq 0.$$

This constraint introduces a scarcity signal in the day-ahead market that is related to the amount of reserve capacity that is available in real time. This scarcity can back-propagate to the day-ahead market, and gives rise to the following day-ahead reserve price signal:

$$\lambda^{R,DA} = \beta_g^{G,DA} + \mathbb{E}\left[a_{g,\omega}^{G,DA}\right] + \mathbb{E}\left[\gamma_{g,\omega'}^{G,RT}\right].$$

The problem with this requirement is that the scarcity signal that it produces is *too strong*. According to this constraint, generators should carry the *entire* reserve requirement of the system in real time, even after reserves have been activated. This is clearly too demanding: reserves are there in order to protect against a shortage before the unexpected happens. The system operator should have the freedom to deplete these reserves once a shortage occurs. This means that the introduction of this constraint is likely to be binding even if the system capacity is adequate to deal with imbalances.

The conclusions of this discussion are the following: (i) The precise requirement for available reserve capacity after activation within an imbalance interval is crucial in terms of reflecting the value of reserve. (ii) The approach of not imposing any requirement on leftover capacity after activation leads to reserve prices that are too weak, whereas the approach of imposing that the full quantity of reserve capacity be available after activation produces reserve prices that are too strong. (iii) The scarcity pricing mechanism remedies this challenge by simply measuring how much leftover capacity is available after activation, and pricing it according to how tight the system is, following a loss of load probability calculation. This can be interpreted as introducing a real-time market for reserve capacity and a demand function from the system operator for real-time reserve capacity.

2.4. The CREG Market Design Questions

The preceding discussion highlights the fact that market design is crucial in producing a forward signal for reserve capacity, because the market design influences two mechanisms by which long-run reserve prices are formed: (i) the arbitrage between power and reserve capacity, and (ii) the arbitrage between day-ahead and real time.

The specific design options that we are interested in analyzing have been distilled from the CREG terms of reference for this study:

- <u>Question 1</u>: Do we need a market for reserve capacity in real time, or can we just rely on the clearing of energy?
- <u>Question 2</u>: Do we require virtual trading?
- <u>Question 3</u>: Should energy and reserve be cleared simultaneously in the day-ahead market, or should reserve be cleared first?

Along these three axes, we can map the ERCOT and Belgian markets as follows:

- Regarding *question 1*: In the Texas real-time market, there is no explicit co-optimization of energy and reserves, however reserve capacity is traded in real time, in the sense that market participants receive a real-time price for the real-time quantity of reserve capacity that they make available to the system. Belgium does not trade reserve capacity in real time, meaning that there is no real-time price signal for reserve capacity (even if there is a price for activated reserve energy). Nevertheless, entities are required to hold enough reserve in real time *before* the activation of reserve, so as to honor their day-ahead reserve commitments, even if this capacity is not priced in real time.
- Regarding *question 2*: Virtual trading of energy is allowed in Texas. Virtual trading is not allowed in the Belgian market.
- Regarding *question 3*: The Texas day-ahead market trades reserves and energy simultaneously in a multi-product auction. Reserve is cleared before energy in the Belgian market.

From the comparison of the previous paragraph, we note that the Texas and Belgian market differ along all three design options that are of interested to the CREG. In the remainder of the report we will investigate a possible evolution from one design to the other. We present this evolution from the least to the most disruptive steps:

- In section 3 we consider the introduction of a real-time market for reserve capacity, and present the resulting model (RCP).
- In section 4 we consider the introduction of virtual trading, and present the resulting model (RCV).

- In section 5 we consider the introduction of co-optimization in day-ahead energy and reserves, which leads to the model SCV that has been presented in section 2.2.
- In section 6 we consider the implication of these changes on the profitability of CCGT resources and loads that can offer reserve, and examine the tradeoff between the costs and benefits of introducing these market design changes to the Belgian market.

Throughout sections 3-5, we also discuss certain implementation challenges, as they relate to the features of the Belgian market. The discussion of the design options is structured according to Figure 7, which presents the entire set of options that have been simulated in section 6. Throughout the document, we describe the differences between each of the designs in this chain. In section 6 we provide a recommendation about how far to go along this chain, taking into consideration the level of disruption that is implied by each of the transitions.

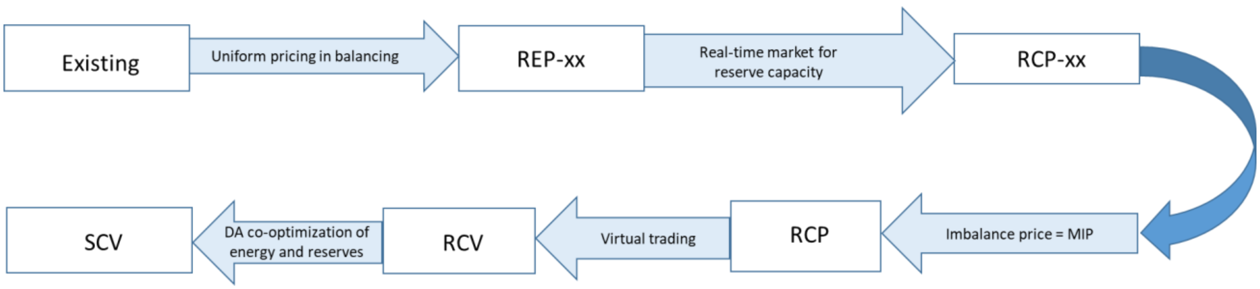


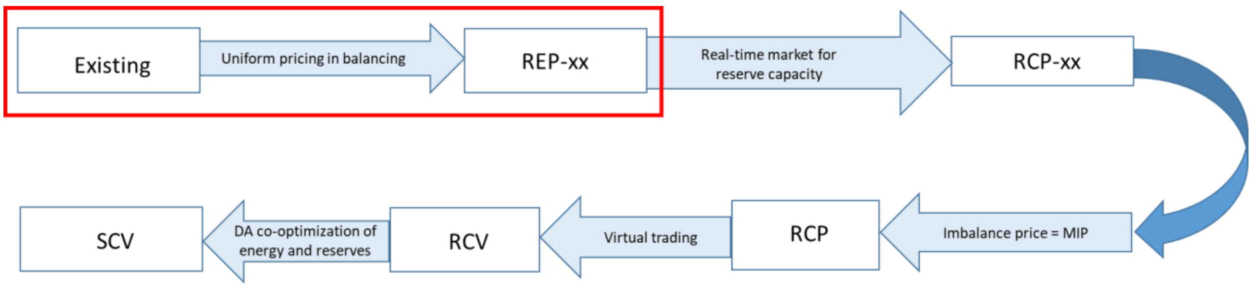
Figure 7: The chain of designs that have been considered in this study.

3. The Real-Time Market

In this section we focus on the real-time market, and discuss the pricing of real-time energy and real-time reserve capacity.

3.1. Pricing Real-Time Energy

The first difference between existing practice in Belgian market operations and the REP model that is described in section 2.3 is the fact that the real-time price of energy is identical for both the entities that cause imbalances and the entities that resolve them. This difference is illustrated in Figure 8.



Existing	**REP-xx**
Resources (free bids / reserves) that are activated in the balancing market are paid as bid. Resources that cause imbalances pay an imbalance price.	Resources (free bids / reserves) that are activated in the balancing market are paid a uniform price, which is the price paid by resources that are causing the imbalances.

Figure 8: Difference between the existing Belgian market design and the REP model that is presented in section 2.3.

The rationale for the uniform pricing of activated real-time energy is that buyers and sellers of the same product should face the same price for the product that is being exchanged. In the case of real-time operations, the product in question is energy. The rationale of the REP model is that generators produce energy, loads consume energy, and energy is exchanged at a uniform price $\lambda_{\omega'}^{RT}$. The total settlement is then:

$$\lambda_{\omega}^{DA} \cdot p_{g,\omega}^{DA} + \lambda_{\omega'}^{RT} \cdot p_{g,\omega'}^{RT} - \lambda_{\omega'}^{RT} \cdot p_{g,\omega}^{DA}$$

There are two equivalent ways of viewing this settlement expression:

(i) In terms of total quantities: producers receive a payment in the day-ahead market for their forward position (first term). In real time, they buy back their forward quantity at the real-time price (third term), and are paid the real-time price for the *entire* real-time quantity (second term).

(ii) In terms of deltas: the settlement can be rewritten as follows:

$$\lambda_{\omega}^{DA} \cdot p_{g,\omega}^{DA} + \lambda_{\omega'}^{RT} \cdot (p_{g,\omega'}^{RT} - p_{g,\omega}^{DA})$$

According to this equivalent point of view, producers receive a payment in the day-ahead market for their forward position (first term), and in real time they are paid the real-time price for their deviation (second term).

The second point of view is more in line with the notion of balancing, but the underlying principle is that there exists a unique real-time energy product with a unique settlement price. According to the REP model, there is no reason why the real-time market should differentiate the payment to the entities that cause imbalances from the entities that relieve imbalances.

According to the REP market design, in a simple dispatch model the real-time energy price $\lambda_{\omega'}^{RT}$ is equal to the marginal cost of the marginal unit. The study that was conducted recently by ELIA [11] considers the possibility of adding an *energy* scarcity adder to the marginal cost of the marginal unit. Assuming that agents can anticipate this adder price, this is effectively an administrative intervention. We argue that this measure, in isolation, will have no effect on dispatch or settlement. To see this, note that in a simple dispatch model (without reserve constraints or other complicating features) resources are dispatched in order of increasing marginal cost. Adding an administratively determined energy adder to this real-time signal will not affect the order of dispatch. Said otherwise, if an adder is included in the MIP, generators will simply adapt their bid marginal costs so as to internalize this adder, assuming that they can anticipate the adder. If the adder is known or can be predicted in advance, the end effect will be the exact same dispatch and settlement as if the adder had not been introduced in the first place: (i) the same set of resources will be dispatched, and (ii) the inclusion of an adder to a MIP which has been adjusted downwards by the anticipated adder will have the net effect of cancelling the effect of the adder. Thus, introducing an adder in the energy price alone has no effect on the market equilibrium or long-run profits of generators.

3.2. Pricing Real-Time Capacity (RCP)

The next design that we consider replaces the pure real-time energy market of the REP model with a simultaneous real-time auction of energy and reserve capacity. The resulting market blueprint is illustrated in Figure 9. Note that, compared to Figure 5, there is an additional product being traded in real time: reserve capacity. This reserve capacity is procured by the system operator as a public good, and is being offered to the system operator by generators and loads.

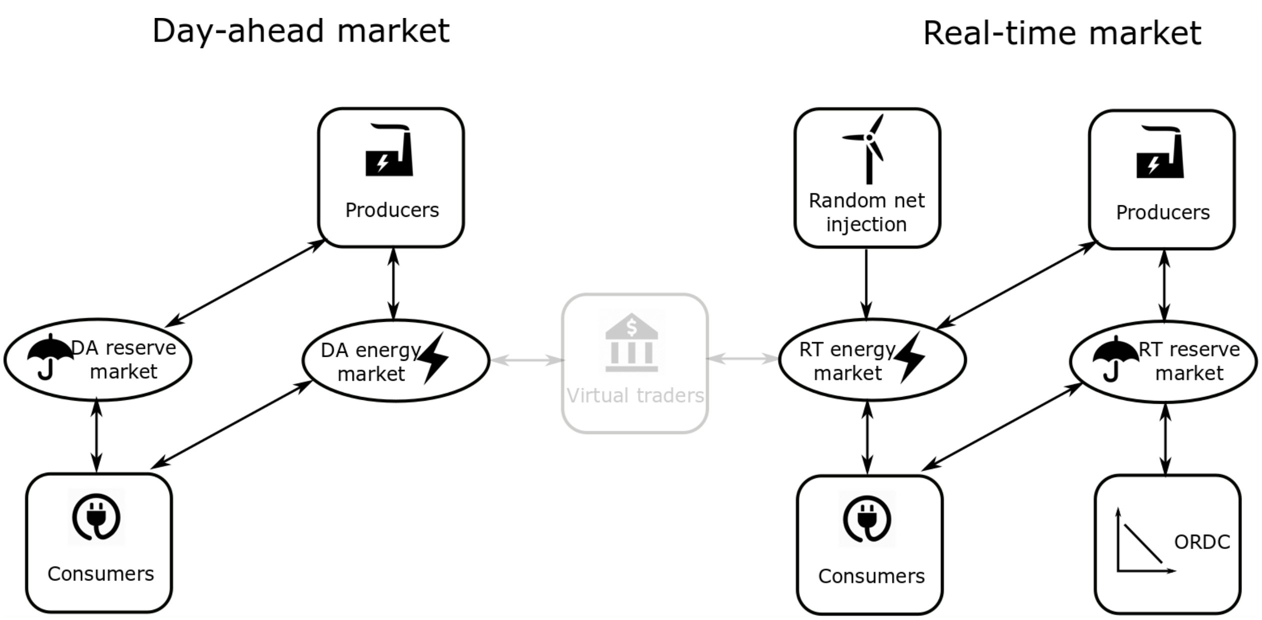
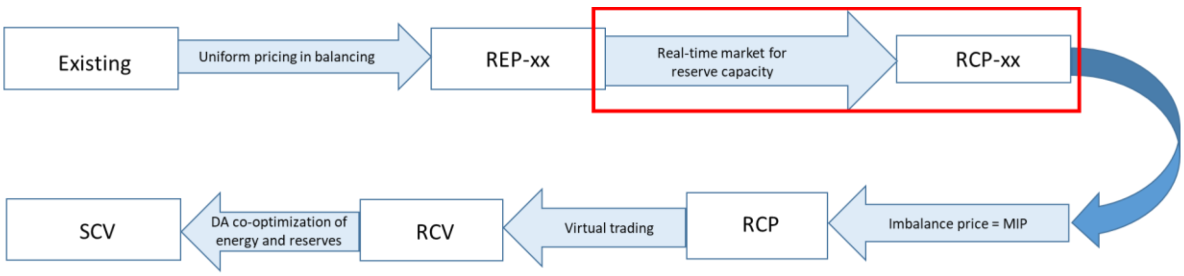


Figure 9: A blueprint of the RCP market design.

The differences between the REP design and the RCP design are indicated in Figure 10. The principal difference between the REP and RCP design is that the real-time market in the RCP design is a multi-product auction that trades energy and reserve capacity simultaneously. In practice, this requires the system operator to determine a real-time price for reserve capacity. There are two ways in which this can be achieved in practice:

(i) An ex ante calculation solves a real-time dispatch problem subject to reserve requirement constraints. The dual multipliers on the reserve balance constraints set the price for real-time reserve capacity. This model explicitly co-optimizes real-time reserve and energy capacity.

(ii) The system is dispatched so as to minimize energy production costs. An ex post calculation determines the reserve capacity price as the marginal value, to the system operator, for an increment of reserve capacity, as determined by the ORDC formula. This is the approach that is adopted in Texas.



REP-xx	RCP-xx
Single-product auction: only energy is traded in real time.	Multi-product auction: activated energy *and* reserve capacity are settled in real time.
No real-time reserve capacity price exists.	A reserve capacity price is computed in real time.
Free bids are only paid a real-time energy price if they are activated to clear imbalances.	Free bids that are standing by (but not activated) are paid a real-time reserve capacity price.
BSPs are only paid a real-time energy price if they are activated to clear imbalances.	BSPs that are activated are paid a real-time energy price for activated energy, but pay a real-time reserve price for buying back activated capacity.

Figure 10: Difference between the REP model (section 2.3) and the RCP model (section 3.2).

Note that most existing European markets do *not* account for the real-time reserve capacity market. Resources that are activated in real time are paid a price for the incremental fuel cost that they incur, but no payment is foreseen for the amount of reserve capacity that they hold in reserve. According to the RCV model, free bids are paid for standby reserve capacity. Even if these bids are not incurring additional costs for standing by (in contrast, for example, to activated free bids, which are consuming fuel in order to provide upward regulation), they are offering value to the system by keeping the loss of load probability under check. Moreover, the existence of a real-time market for reserve capacity increases the attraction for such free capacity to be in place in real time.

The fact that activated energy bids *pay* the system operator for reserve capacity when being activated upwards is consistent with their profit maximizing incentives. If generators are asked to be activated upwards, it is because they are getting a better payment from the energy market (even if they are depleting their reserve capacity and paying back for the capacity that they deplete) than they would receive from standing idle and not being activated upwards. To put it more simply: the clearing prices of the real-time market in the RCP design are consistent with the activation instructions that generators are asked to execute.

In recent discussions, ELIA raised concerns about the exposure of real-time reserve capacity markets to market power. Given that real-time dispatch in the Belgian market is closely monitored by the TSO and is akin to central dispatch, market power should be straightforward to mitigate. Individual generators are nominated in the day-ahead, and real-time generator capacity is monitored by the TSO through telemetry, therefore it is unlikely that units should be in a position to conceal information about the real-time availability of their units, or to self-dispatch in real time by withholding capacity without TSO permission. Once units have been nominated in the day ahead, the commitment and dispatch of units is *not* voluntary.

3.2.1. Real-Time Market Equilibrium

The real-time market equilibrium of the RCP model is expressed as follows. Generators solve the following profit maximization problem:

$$\max_{p,s,r\geq 0} \Pi_{g,\omega'}^{RT} = \lambda_{\omega'}^{RT} \cdot p_{g,\omega'}^{RT} + \lambda_{\omega'}^{R,RT} \cdot r_{g,\omega'}^{RT} - C_g \cdot p_{g,\omega'}^{RT} - \epsilon_g^{+} \cdot s_{g,\omega'}^{RT,+} - \epsilon_g^{-} \cdot s_{g,\omega'}^{RT,-}$$

$$(\alpha_{g,\omega'}^{G,RT}): p_{g,\omega'}^{RT} + r_{g,\omega'}^{RT} \leq P_{g,\omega'}^{RT,+} \cdot y_{g,\omega}$$

$$(\beta_{g,\omega'}^{G,RT}): r_{g,\omega'}^{RT} \leq R_g$$

$$(\gamma_{g,\omega'}^{G,RT,+}): p_{g,\omega'}^{RT} - p_{g,\omega}^{DA} - s_{g,\omega'}^{RT,+} \leq 0$$

$$(\gamma_{g,\omega}^{G,RT,-}): p_{g,\omega}^{DA} - p_{g,\omega'}^{RT} - s_{g,\omega'}^{RT,-} \leq 0$$

Loads solve the following profit maximization problem:

$$\max_{d\geq 0} \Pi_{l,\omega'}^{RT} = V_l \cdot d_{l,\omega'}^{RT} - \lambda_{\omega'}^{RT} \cdot d_{l,\omega'}^{RT}$$

$$(\alpha_{l,\omega'}^{L,RT}): d_{l,\omega'}^{RT} \leq D_{l,\omega'}^{RT,+}$$

The system operator solves the following profit maximization problem:

$$\max_{d^R\geq 0} \sum_{l\in RL} (V_l^R - \lambda_{\omega'}^{R,RT}) \cdot d_{l,\omega'}^{R,RT}$$

$$(\alpha_{l,\omega'}^{R,RT}): d_{l,\omega'}^{R,RT} \leq D_l^{R,RT}$$

The price quantity pairs $(V_l^R, D_l^{R,RT})$ determine the real-time ORDC of the system operator. This demand function can either be inelastic (in which case it results in relatively infrequent price spikes) or it may reflect the decreasing valuation that the system operator places on additional increments of reserve capacity, in which case reserve prices and energy prices exhibit more frequent, but less pronounced, price spikes.

The market-clearing constraints are expressed as follows:

$$\sum_{g\in G} p_{g,\omega'}^{RT} = \sum_{l\in L} d_{l,\omega'}^{RT}$$

$$\sum_{g \in G} r_{g,\omega'}^{RT} = \sum_{l \in RL} d_{l,\omega'}^{R,RT}$$

3.2.2. Day-Ahead Market Equilibrium

The day-ahead market model is identical to that of section 2.3.2. A relevant market design decision that is required in this stage is the choice of day-ahead operating reserve demand curve, and how it relates to the real-time ORDC.

Both the day-ahead and real-time ORDC parameters can be determined on the basis of parameters that are pre-computed, and adapted to the season and hour of the day. Texas employs 24 different ORDC functions, depending on the season and 4-hour interval of the day. In the case study of section 6, the same approach is adopted for the Belgian system, using the parameters that are estimated in **Error! Reference source not found.**.

According to the simulation results of section 6, the RCP design is the least challenging and most effective step towards improving the long-run valuation of reserve capacity. As generators arbitrage day-ahead and real-time market profit opportunities, the following back-propagation formula emerges for day-ahead reserve capacity prices:

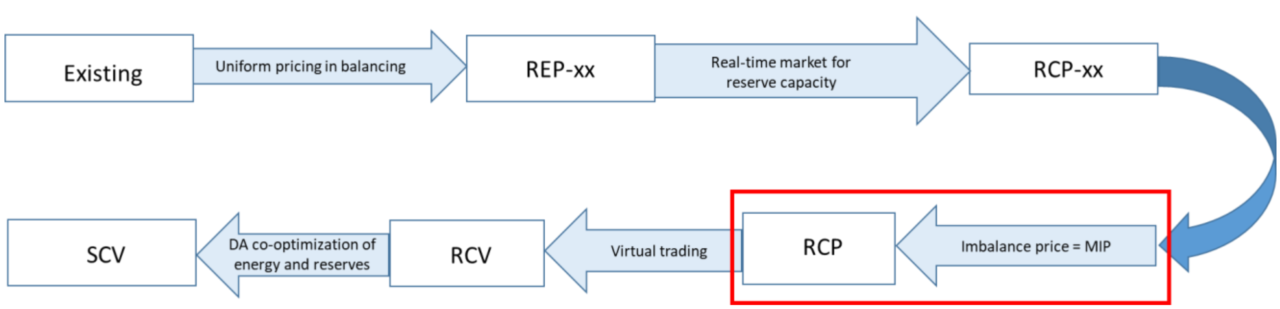
$$\lambda^{R,DA} = \beta_g^{G,DA} + \mathbb{E}\left[a_{g,\omega}^{G,DA}\right] + \mathbb{E}\left[\lambda_{\omega'}^{R,RT}\right].$$

Compared to the REP design, the day-ahead reserve price formation is driven by the expected real-time price of reserve capacity, $\mathbb{E}\left[\lambda_{\omega'}^{R,RT}\right]$. The introduction of a real-time ORDC allows this real-time price to automatically adjust according to the real-time scarcity in the system, which is the result of a loss-of-load probability calculation. The first and second term in the back-propagation formula are related to scarce capacity and scarce ramping in the day-ahead auction.

We argue in section 2.2.3 that, depending on the requirements that the TSO places on leftover reserve capacity *after* activation, the valuation of reserve may become too low (if the TSO places no requirement) or too high (if the TSO requires the full amount of reserve to be available after activation). The real-time ORDC automates this calculation in a self-correcting fashion, and arbitrage propagates this price to the day-ahead market, thereby signaling the need for investment in reserve capacity in case of tight system conditions.

3.3. Imbalance Penalties

The third aspect of the real-time market that we concentrate on is the presence of administrative penalties on imbalances. These administrative penalties are illustrated in Table 3 for the Belgian real-time market. These administrative penalties reflect the notion that BRPs should strive to balance their own perimeter, and the TSO should be responsible for any residual imbalance by activating reserve capacity or free bids.



RCP-xx	RCP
Resources that are short in real time pay a penalty $\alpha 1$ in addition to MIP whenever the system is (very) short. Resources that are long in real time are paid MDP minus a penalty $\alpha 2$ whenever the system is (very) long.	Resources that are short pay the system marginal cost for oversupply. Resources that are long are paid the system marginal cost for undersupply.

Figure 11: Differences between the RCP-xx model (section 3.2) and the RCP model (section 3.3).

The rationale of removing the administrative penalties is that a single product is traded in the real-time market (namely, real-time energy) and should be priced consistently between those who produce it (e.g. the entities that over-produce relative to their forward positions) and those who consume it (e.g. the entities that over-consume relative to their forward positions). Nevertheless, as we demonstrate in the numerical simulations of section 6, the presence of these administrative penalties has a minor effect on the forward prices and profit margins of reserve suppliers. These penalties are akin to uninstructed deviation charges that are set in place in other systems.

The equilibrium formulation of the RCP model without administrative penalties is almost identical to that of section 3.2, with the exception that the generator profit maximization is modified in order to reflect the removal of the administrative penalties:

$$\max_{p,r\geq 0} \Pi_{g,\omega'}^{RT} = \lambda_{\omega'}^{RT} \cdot p_{g,\omega'}^{RT} + \lambda_{\omega'}^{R,RT} \cdot r_{g,\omega'}^{RT} - C_g \cdot p_{g,\omega'}^{RT}$$

$$(\alpha_{g,\omega'}^{G,RT}): p_{g,\omega'}^{RT} + r_{g,\omega'}^{RT} \leq P_{g,\omega'}^{RT,+} \cdot y_{g,\omega}$$

$$(\beta_{g,\omega'}^{G,RT}): r_{g,\omega'}^{RT} \leq R_g$$

Effectively, generators are no longer penalized for deviations relative to their day-ahead positions.

To a certain extent, the central dispatch organization of US markets is conducive towards eliminating imbalance penalties because it ensures that generators follow the system operator set-points by effectively handing over control to the system operator in real time. This renders administrative penalties redundant to some extent, and the focus instead is on creating price signals that are consistent with the central dispatch from a decentralized, profit-maximizing point of view. The philosophy of US market design is that operational efficiency is ensured by co-optimized central dispatch, and a separate pricing exercise takes place which produces price signals that strive to be consistent with this central dispatch[21]. Nevertheless, the setup in European *real-time* operations is not entirely different. Also in European real-time operations, TSOs carry significant authority and can largely control the set-point of individual units that are nominated in the day-ahead and that are expected to closely track dispatch instructions in real time. One major difference between European and US markets is the absence of sophisticated optimization in European real-time operations, which is replaced by simplified heuristics (based on merit order dispatch), operator judgement, and experience. Notwithstanding, the argument can still be made that European real-time operations are, to some extent, controlled centrally by the system operator, at least insofar as balancing for small bidding zones (such as Belgium) is concerned.

[21] There are challenges in the US design that relate to the settlement period. This is why the FERC has been pushing for 5-minute settlement.

3.4. Idiosyncratic Features of the Belgian Market

In this section, we provide a qualitative discussion of some other notable features of the Belgian real-time market. We have not attempted to model these features.

3.4.1. Coordinated Balancing

The European balancing market is moving towards increasing integration. Towards this end, two initiatives that stand out are the "Platform for the International Coordination of Automated Frequency Restoration and Stable System Operation" (PICASSO[22]) and the "Manual Activated Reserves Initiative" (MARI[23]). The goal of these initiatives is to enable the sharing of aFRR and mFRR capacity in real-time operations.

PICASSO and MARI aim at activating secondary and tertiary reserve bids by accounting for international transmission capacity constraints. The output of these platforms would be dispatch instructions to individual BRPs, as well as market clearing prices for activated energy.

It is unclear whether the separation of reserve activation into secondary and tertiary reserves promotes operational efficiency. Empirical evidence suggests that modern optimization algorithms can handle the computational complexity of multi-product auctions that co-optimize energy, transmission, and multiple substitutable reserve products, and can generate price signals as a result of this optimization that are consistent with the arbitrage between energy, transmission capacity, and reserve capacity. The separation of these processes in real time may create inconsistencies in the pricing of these products that could be avoided in a coordinated optimization.

Notwithstanding, as we have argued throughout the present section, the most crucial aspect of real-time market design is to generate a signal that values excess reserve capacity *after* resources have been activated for balancing. This is a calculation that can be performed ex post, as is currently the case in Texas. In other words, the absence of co-optimization of reserve and energy does not preclude the computation of scarcity prices, and prices for reserve and energy can be constructed that come close to the result that would be produced from co-optimization.

What is unclear, however, is how these scarcity prices will interact with the prices generated by PICASSO and MARI. The separation of the PICASSO and MARI balancing processes suggests that energy activated as part of aFRR is a different product than energy activated as part of mFRR, and should be priced differently[24]. By contrast, in the present section we argue that energy in real time is one product, independently of whether it is upward or downward, if the system is long or short, and if it originates from mFRR, aFRR, free bids, or self-dispatch. A simple application of the adder to the prices resulting from PICASSO and MARI is possible, however it is understood that the resulting settlement may be inconsistent with the dispatch instructions issued by the PICASSO and MARI platform. Concretely, BSPs that were dispatched by PICASSO/MARI at a certain price generated by the PICASSO/MARI platform (which is consistent with the PICASSO/MARI dispatch instructions) would face scarcity adders for energy and reserves which were not accounted for when the PICASSO/MARI algorithm produced its market clearing results.

---

[22] https://electricity.network-codes.eu/network_codes/eb/picasso/

[23] https://www.entsoe.eu/network_codes/eb/mari/

[24] The possibility of defining the price of energy as the marginal of the marginal price of these platforms has been suggested.

3.4.2. Strategic Reserve

Since 2014, Belgium has put a strategic reserve in place as a means of mitigating the risk of capacity shortages during winter months when load in the system peaks. The strategic reserve consists of units that are planned to be mothballed[25], but are paid a capacity price for remaining at the disposal of the system operator during periods of capacity shortage. Strategic reserve that is activated in the day-ahead market triggers a very high day-ahead market clearing price, as well as a default very high imbalance price, so as to prevent entities from leaning on the system in real time.

Strategic reserve should not be counted towards the computation of scarcity adders. The inclusion of strategic reserve in the adder calculations would effectively eliminate the adder signal, independently of system conditions, and would defeat the purpose of scarcity pricing.

[25] For example, the units that were available as strategic reserve for 2015 included Seraing (485 MW), Vilvoorde (265 MW), Angleur (50 MW), Izegem (20 MW) and Esche-sur-Alzette (357.1 MW), as well as 358.4 MW of demand response, thereby totaling a strategic reserve capacity of 1535.5 MW. Source: http://www.creg.be/fr/producte9.html.

4. Virtual Trading

Virtual trading is the practice of allowing agents to trade energy in the day-ahead market, even if they do not own physical assets. The intended benefit of virtual trading is to exploit the "wisdom of the crowds" so as to permit day-ahead prices to converge to the expected real-time prices. For example, if there is a sense by traders that day-ahead prices are over-valued compared to the expected value of electricity in real time, virtual trading would allow traders to sell energy in the day-ahead market (even if they do not own generating assets) and pay back that for that energy in the real-time prices (assuming traders do not own physical assets, they cannot actually produce the power that they sold in the day-ahead market, so the only way for them to honor their day-ahead trade is by buying back their position at the real-time price). The end result of this increased sale of electricity in the day-ahead market is to exert an upward pressure on day-ahead prices, and bring them closer to the average real-time prices. In this way, the expert knowledge of virtual traders about the expected real-time price contributes to the formation of day-ahead prices that are consistent with the average real-time prices. This contributes towards back-propagating efficient investment and operational planning signals to the day-ahead and earlier forward markets, thereby promoting short and long-term operational efficiency and effective risk management.

Note that the introduction of virtual trading, and more broadly the treatment of the real-time market as the spot market, is not meant to compromise security. Even under virtual trading, resources that are nominated will need to be present in real time. US markets that permit virtual trading do, at the same time, require resources that have been committed and that receive real-time dispatch signals to be physically present in real time. The fact that virtual trading is allowed does not imply that physical units can deviate arbitrarily from central dispatch and buy back their energy without suffering consequences for not being available when the system operator was counting on them to be available. Virtual trading is intended to affect day-ahead prices, and to some extent day-ahead unit commitment (ideally by steering this commitment towards economic efficiency). It is not meant to undermine real-time operations and security.

This section considers the second major market design question raised by CREG, namely whether virtual trading should be allowed in the Belgian market. The motivating argument behind this question is that virtual trading should permit a more precise back-propagation of expected real-time prices for reserve and energy, thereby refining the precision of the scarcity signal. In section 6 we demonstrate that this effect is limited in the case of risk-neutral agents. Given the level of disruption that the institution of virtual bidding would imply in European market operations, the resulting benefits may not justify the implementation effort.

4.1. How Virtual Trading Works

We provide a simple numerical example of how virtual bidding works. Before discussing the example in detail, we summarize the conclusion of the example here:

- The absence of virtual trading may lead to day-ahead prices that are not equal to average real-time prices.
- Virtual trading tends to eliminate these differences. Risk-neutral agents will tend to push the day-ahead prices to their average real-time value. If virtual trading is limited, this may not occur. If virtual trading is allowed at greater quantities, this is more likely to occur.

We consider a system with the composition presented in Table 5. We introduce uncertainty to the system by including a renewable generator that may produce 56 MW, or 156 MW, with equal probability.

	Capacity (MW)	Marginal cost (€/MWh)	Ramp capacity (MW)
Blast furnace	323	38.1	323
Renewable	56 or 156	35.7	0
Gas-oil	5	85.0	2.5
LVN	212	315	106

Table 5: Virtual trading example. A system with three reliable technologies and an unreliable renewable generator.

The supply function of the real-time market under conditions of low and high supply are shown in Figure 12. We observe that, depending on the available supply, the real-time price may be 35.7 €/MWh or 38.1 €/MWh.

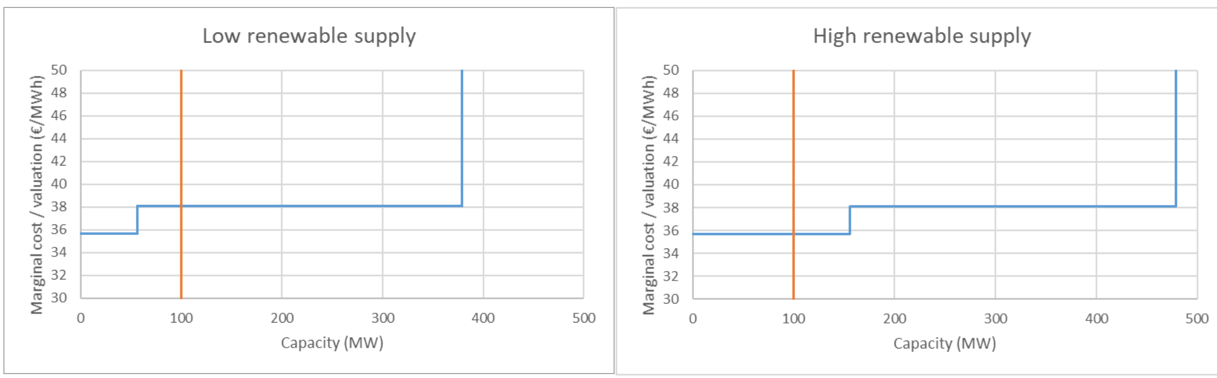


Figure 12: Virtual trading example. The real-time supply function of the system under low (left) and high (right) supply.

Consider, now, a day-ahead market in which virtual trading is not allowed, in the sense that conventional producers are only allowed to bid the cost characteristics of their physical assets, renewable producers are only allowed to bid their forecast production, and consumers are only allowed to bid their demand forecast. The outcome of the market is shown in the following figure, and results in a market clearing price of 35.7 €/MWh. Here, the renewable supplier bids its day-ahead forecast of 106 MW. Notice that the average real-time price and the day-ahead price are actually not equal.

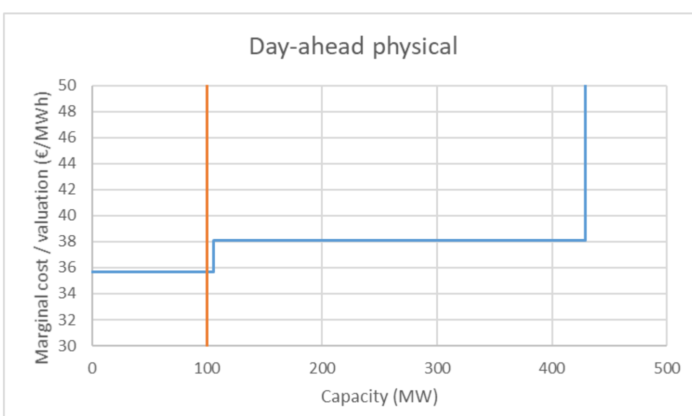


Figure 13: Virtual trading example. The outcome of a day-ahead market where virtual bidding is not allowed.

Consider now a market in which we introduce a single risk-neutral agent for which virtual trading is allowed. The market allows virtual trading in the sense that the agent is allowed to place a bid in the day-ahead market for which there is no corresponding physical capacity to back it up. For example, the virtual bidder may be a financial institution which specializes on forecasts and controls no physical assets. Since this agent is risk-neutral, and assuming that it can anticipate the real-time price, the agent will place a supply bid in the day-ahead market for buying energy at the price that it expects to sell it back in real time, which is the average real-time price, i.e. 36.9 €/MWh. This results in the following supply and demand functions in the day ahead market.

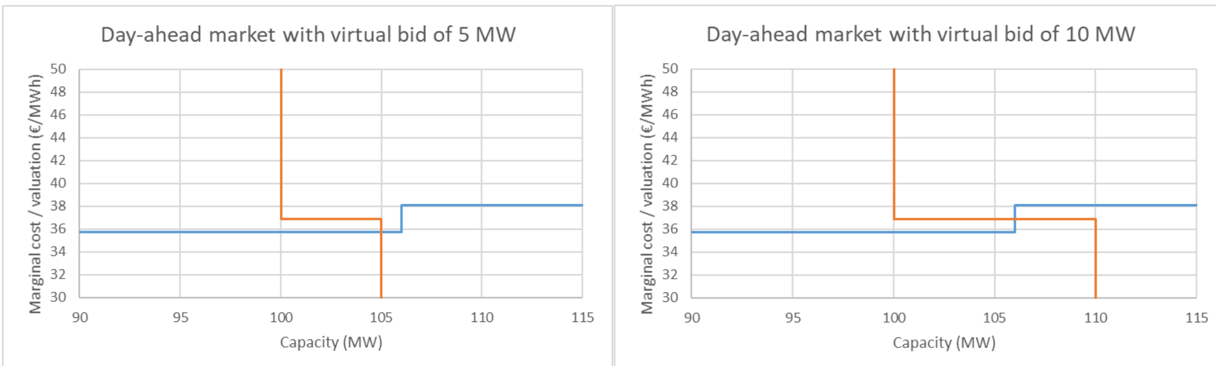


Figure 14: Virtual trading example. Left: the day-ahead market clearing with a virtual bid of 5 MW (limited virtual trading) at 36.9 €/MWh, which is the average real-time price. Right: the day-ahead market clearing with a virtual bid of 10 MW (twice as much virtual trading) at 36.9 €/MWh.

Note the effect of the virtual bid. The virtual bid has effectively introduced a block in the demand function at the average real-time price. Depending on how exposed this risk-neutral agent wishes to be, this will effectively tend to align the day-ahead price to the average real-time price.

Note the distinction between the day-ahead market and the real-time market. The special attribute of the real-time market which distinguishes it from an intra-day, day-ahead or other forward market is that it is the only moment during which the true physical production capabilities of the system meet the true physical demand of the system, and therefore it is the only moment in time where the actual physical scarcity of the system is revealed. Anything that happens in earlier markets is just based on forecasts.

As demonstrated in the example above, the functioning of virtual bidding effectively hinges on allowing agents to deviate from the true techno-economic characteristics of their assets, and incorporating their perceived opportunity costs into the bidding process. Therefore, any system that allows agents to bid, in a day-ahead market, in a way that is not perfectly consistent with physical assets effectively enables virtual bidding to a certain extent. In fact, even in the way that the existing Belgian system is set up with agents bidding portfolios as opposed to true physical assets, it can be argued that virtual trading is effectively allowed to a certain extent[26].

[26] In practice, BRPs have a certain degree of flexibility in the day-ahead time frame in terms of bidding. For example, BRPs can submit their own renewable supply forecasts. Since forecasts are private information, there is no way to strictly enforce physical constraints on the energy position of BRPs. On the other hand, financial institutions without any ownership of physical assets are not allowed to participate in the day-ahead energy market, which is in stark contrast to certain US market designs.

4.2. The Virtual Trading Model (RCV)

The economic equilibrium model with virtual trading is presented in Figure 15. Virtual traders are included in the model, which take positions in the day-ahead energy market and close their positions in real time. Note that the virtual traders can only engage in the energy market. Trading in the reserve market requires physical assets.

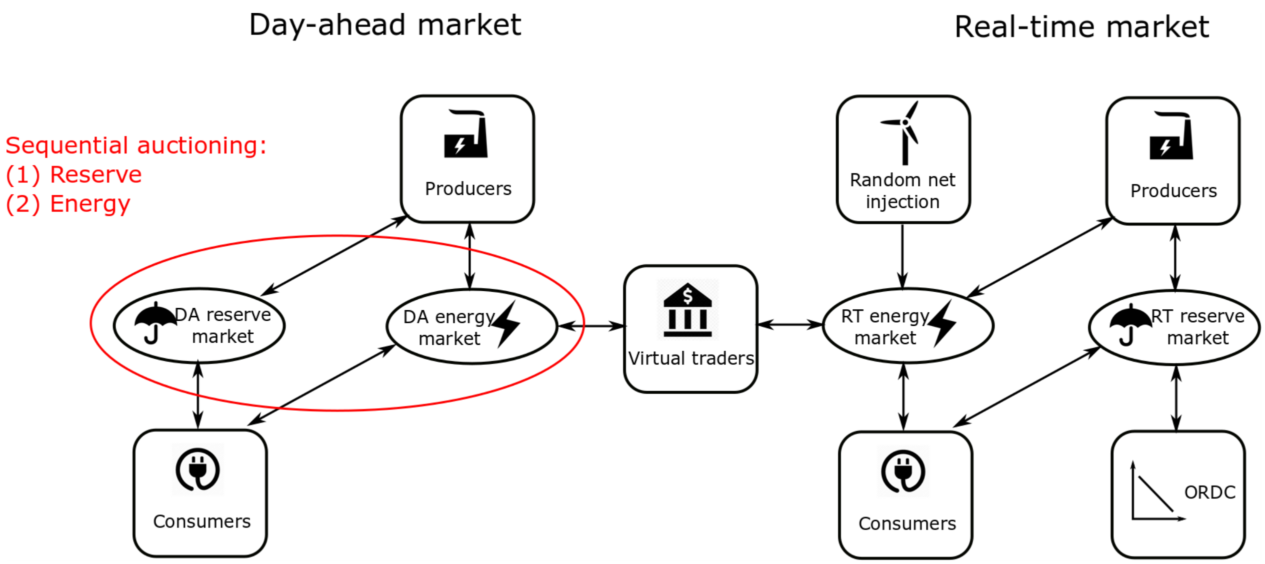
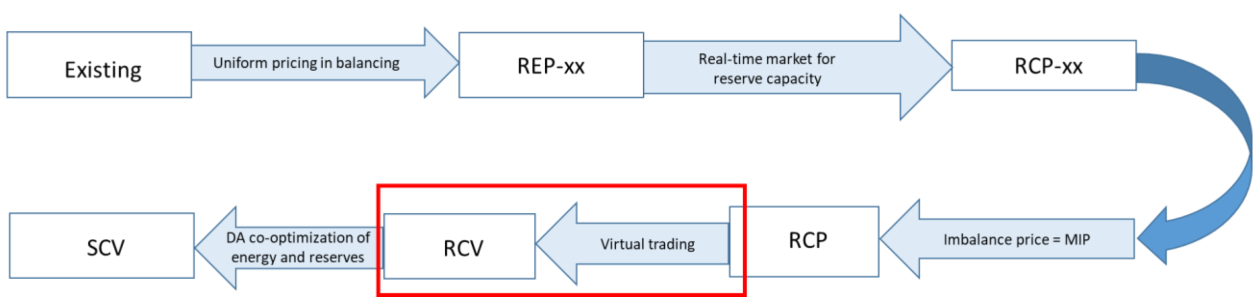


Figure 15: A blueprint of the RCV market design.

In US-style pools, individual generators are bid into the day-ahead market through multi-part bids that describe their physical characteristics and the costs of various operations (starting up the units, minimum load cost and their marginal cost function). For these individual assets, their PMin, PMax, ramp rates, and associated costs are represented accurately in the day-ahead unit commitment model[27]. Virtual bids may then be placed separately from the bids of these individual physical resources. For example, the PJM day-ahead market accepts an average of 1210 three-part offers for generators, 10000 demand bids, and 50000 virtual bids[28]. In a European setting where resources are bid into the market as portfolios, the implementation of virtual bidding could follow analogously, in the sense that the bidding of physical assets as portfolios would remain intact, but additionally virtual bids could be introduced which would arbitrage away deviations between the real-time and day-ahead market.



[27] Attempts to deviate from this in order to game make-whole payments have been detected and settled by the Federal Energy Regulatory Commission (https://www.ferc.gov/CalendarFiles/20130730080931-IN11-8-000.pdf).
[28] https://www.ferc.gov/CalendarFiles/20100601131610-Ott,%20PJM.pdf.

RCP	RCV
Only entities with physical assets are allowed to participate in the day-ahead market. Portfolio bids correspond to physical assets.	Entities without physical assets can trade in the day-ahead *energy* market. Virtual bids are separately identified from portfolio bids of physical assets.

Figure 16: Differences between the RCP model (section 3.3) and the RCV model (section 4.2).

The real-time equilibrium of the RCV model is identical to that of RCP (section 3.3). The day-ahead energy market model reads very similarly to that of the REP model (section 2.3.2). The only difference to the REP model is the generator profit maximization problem:

$$\max_{y\geq 0,p} \Pi_{g,\omega}^{DA} = \lambda^{DA} \cdot p_{g,\omega}^{DA} - K_g \cdot y_{g,\omega} + \mathcal{R}2_{g,\omega}\left(\Pi_{g,\omega'}^{RT} - \lambda_{\omega'}^{RT} \cdot p_{g,\omega}^{DA}\right).$$

$$\left(\delta_{g,\omega}\right): y_{g,\omega} \leq 1$$

$$\left(a_{g,\omega}^{G,DA}\right): r_g^{DA} \leq R_g \cdot y_{g,\omega}$$

Compared to the corresponding day-ahead energy market equilibrium of REP, the second constraint no longer implicates the day-ahead energy trading position, and the energy trading position is also unconstrained in sign. In the day-ahead energy market, generators still need to decide which generators they will activate in order to ensure that they can deliver the reserve that they have committed in the day-ahead reserve market, however their position in the energy market is in no way limited by the physical capacity of the generating unit. This models the fact that an agent which owns physical assets can also submit separate virtual bids.

4.3. Virtual Trading in Practice

Virtual trading has recently been criticized by Parsons [19]. PJM has also challenged the idea. Hogan [20] does not negate these dysfunctions but argues that dropping virtual trading would do more harm than good because of the essential role of virtual trading for allocating risk.

Even if virtual trading is not allowed explicitly in the Belgian market, it can be achieved indirectly due to the fact that the forecasts of utilities regarding demand and renewable supply are private information. More broadly, there are many examples of places where generators (typically renewables) do not participate in the day-ahead market. Virtual trades effectively take their place.

5. Day-Ahead Clearing of Energy and Reserves

In this section we discuss the timing of reserve capacity auctions.

5.1. Clearing Reserve before Energy

The current practice in Belgian markets is to conduct a reserve capacity auction before energy. Depending on the specific reserve product that is being considered, the tendering may be conducted on a monthly (R3) or weekly (R2) basis. The trend in Belgium is towards daily tendering [21], in order to exploit the day-ahead information that can support an adaptive sizing of reserve capacity. This move is expected to reduce the cost of reserving capacity while achieving the same target level of 99.x% reliability. Day-ahead information that can support precise sizing includes weather conditions that affect forecast errors (e.g. temperature, insolation, wind conditions), forecast day-ahead schedules of generators which affect capacity outage distributions, and schedules on large interconnectors, such as the Belgium-UK NEMO link. Once reserve capacity is cleared, agents need to nominate individual units in the day-ahead to the system operator, in order to ensure that this reserve can indeed be delivered in real time.

The separation of reserve auctions from energy auctions has a precedent in US market design. A challenge that emerged in this separation relates to gaming. This is an aspect that is not captured in our model, but we discuss it briefly here[29].

The gaming of reserve markets in auctions that separate the clearing of energy and reserves is facilitated by the fact that bidding in such reserve auctions is based on opportunity costs. The quantification of opportunity costs is subjective to some extent, since it depends both on anticipated real-time energy prices, as well as the risk attitude of agents, which are both subjective.

The gaming that occurred in the early designs of the California market related to the substitutability of reserve products [22] and so-called **price reversals** in the California reserve market. Price reversals refer to the fact that higher quality products are cleared at lower prices. Reserve products are naturally substitutable, because generator capacities with response speeds that are adequate for covering the needs of R2 are clearly also fast enough to cover the needs of R3. In the California market, various auction designs were considered in which fast reserves were cleared before slower reserves. These designs suffered from the fact that lower quality products (R3) were priced higher than higher quality products (R2). This is clearly problematic, since it induces generators to migrate their capacity to the lower-quality reserve auction (where they are paid better for offering a product that is easier for them to deliver), thereby causing liquidity problems in the R2 market.

It is worth noting that such gaming issues related to price reversals have not been documented in the Belgian market. However, it is also worth noting that a price of 25 €/MW-h (the 2014 R2 price [23]) is difficult to justify purely on the basis of CCGT operating costs (even when startup costs are accounted for), and that since the R2 market has opened up to storage resources, R2 prices have been observed to decline.

[29] The resolution proposed in [22] is a co-optimization of reserve products, which produces prices that are guaranteed to reward high-quality reserve products at least at the level of low-quality reserve products. The drawback with this approach is that it presents the highest bill to the system operator, since agents are effectively collecting information rents from the system operator.

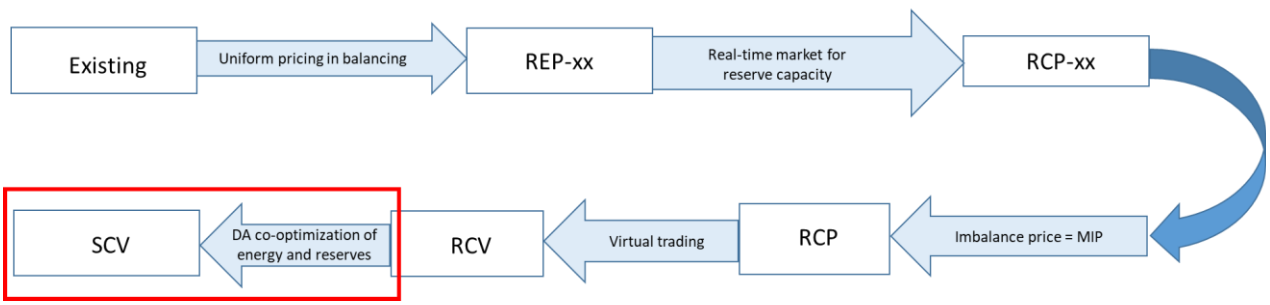
5.2. Simultaneous Clearing of Reserve and Energy

The simultaneous clearing of reserve and energy mitigates gaming opportunities, since the opportunity cost valuation of reserve capacity becomes endogenous to a co-optimization model that receives verifiable information, such as the technical characteristics and costs of physical units. Another appealing effect of co-optimization is that it uses the latest possible information for scheduling units in the day-ahead time frame.

The access to more precise information, and how it compares to the sequential clearing of energy in the RCV model and reserves (section 4.2, see also Figure 6), is the aspect that we quantify with the SCV equilibrium model of section 2.2. This follows in spirit the work of Oggioni and Smeers [24].

It turns out from the simulation results of section 6 that, in a risk-neutral setting, this effect is limited. The implementation overhead that is involved with the introduction of co-optimization in day-ahead market clearing may therefore not be justified at this stage by a back-propagation argument, especially in a setting with risk-neutral agents.

As indicated in Figure 17, there are two principal means of introducing a simultaneous clearing of energy and reserves in the day-ahead market. According to the first approach, which we refer to as the exchange approach (section 5.2.1), reserve products are introduced to the existing day-ahead energy exchange. According to the second approach, which we refer to as the pool approach (section 5.2.2), individual resources are bid into the market, and the allocation of the capacity of these resources between energy and reserves is co-optimized.



RCV	SCV
Reserve capacity is auctioned before the clearing of the day-ahead energy market. The day-ahead exchange only trades energy products.	There is no separate reserve capacity auction. Option 1 (exchange approach, section 5.2.1): The day-ahead exchange introduces reserve products. Option 2 (pool approach, section 5.2.2): Bids in the day-ahead market correspond to individual resources, and the allocation of energy and reserves is co-optimized.

Figure 17: Differences between the RCV model (section 4.2) and the SCV model (section 2.2).

The implications of co-optimization in back-propagating the value of reserve are discussed in section 2.1. One major appeal of co-optimization is that it results in an automatic arbitrage between energy and reserve capacity, meaning that the output of a co-optimization model is such that agents are indifferent between allocating their capacity between reserve and energy. Thus, if reserve prices are lifted as a consequence of a scarcity pricing mechanism, energy prices follow suit. In the absence of co-optimization, this arbitrage step needs to be introduced 'manually' by traders who are required to estimate the opportunity cost of reserve capacity in day-ahead reserve auctions. This estimation step is subjective to some extent, and depends on beliefs about real-time prices and the attitude of market participants towards risk.

5.2.1. Enhancing EUPHEMIA with Reserve Products[30]

The presence of binary commitment variables in day-ahead electricity market models renders the derivation of prices for these markets a non-trivial task. Equilibrium prices in the presence of integer decision variables may not exist, depending on the parameters of the market clearing problem at hand. This can lead to (i) paradoxically accepted bids, i.e. bids that are accepted at a loss for the bidding participant, and (ii) paradoxically rejected bids, i.e. rejected bids that if accepted would generate a profit for the bidding participant. Paradoxically rejected block bids are allowed and not compensated in most electricity markets [25], [26]. The treatment of paradoxically accepted bids, on the other hand, differs among markets: US markets allow paradoxically accepted bids and cover the losses of bidding participants through uplift payments [26], while European markets do not allow paradoxically accepted bids [25].

The inclusion of reserve products in day-ahead market clearing has been demonstrated in a simplified version of the EUPHEMIA model by Aravena [27]. Aravena derives a model for the day-ahead market clearing problem that respects the European pricing restrictions, following the MILP framework of Madani and Van Vyve [14].

The idea presented by Aravena [27] is to formulate the day-ahead market clearing problem in terms of continuous bids, block bids and linked families. Reserve products are represented by quantity bids, i.e. no opportunity cost is associated to the submission of reserve offers. The day-ahead market clearing model also includes a day-ahead demand function for operating reserve, which is bid in by the TSO of each zone. This is in contrast to the current approach for bidding reserve capacity, where bids are submitted in price-quantity pairs, with bid prices corresponding to an estimate of the opportunity cost of reserve capacity.

The model of Aravena generates market clearing prices for reserve as an output of the day-ahead market clearing model, while respecting the general principle of European market clearing whereby resources may be paradoxically rejected, but may not be paradoxically accepted. The interesting implication of Aravena's model insofar as reserves are concerned is that reserves pass surplus to their parent. This implies that, even if a block bid cannot generate enough surplus by being accepted for a given market price, it may end up being accepted if the price of reserve creates sufficient surplus for the resource to be committed.

From a computational standpoint, the model of Aravena as it is presented in [27] introduces an additional product in EUPHEMIA that resembles continuous orders, and interacts with block bids. In a static setting, the problem is tractable as the computational experiments of Aravena demonstrate. The implementation of the approach on a realistic instance of the problem, with the inclusion of additional products that are known to cause serious convergence challenges for EUPHEMIA (such as Minimum Income Condition orders and PUN orders), goes beyond the scope of the present study. Here we limit ourselves to describing how the EUPHEMIA model could be extended in order to accommodate the introduction of reserve products.

[30] The following section is based on Aravena [26].

5.2.2. Pools

The alternative to an exchange which is broadly adopted in the US for the simultaneous clearing of energy and reserves are power pools. As we explain in the previous section, the presence of non-convexities that complicate pricing is present both in modern renderings of power exchanges as well as pools. Instead, the two major distinctions are the definition of products and the bidding of portfolios instead of individual resources.

Generators are bid into a power pool through multi-part bids that include all the data that is required for populating a day-ahead unit commitment model that co-optimizes energy and reserve capacity. Such information includes startup costs, min load costs, an increasing marginal cost curve, and technical information that includes PMin, PMax, generator ramp rates, minimum up and down times, and even more complex data for resources such as multi-stage generators. In fact, the introduction of increasingly complex products in exchanges attempts to proxy the numerous combinations in which resources can be committed in a unit commitment model. Thus, the complexity of the unit commitment problem in a pool is transferred, in exchanges, to the definition of increasingly sophisticated products.

The requirement to bid generators individually in a power pool is a consequence of how a unit commitment problem is defined. It has been argued that the ability of market participants to bid portfolios instead of resources is an appealing aspect of the current European exchange, since it offers market participants increased flexibility. On the other hand, it is clear that aggregations in resource allocation problems deteriorate the quality of the resulting solution. So long as commercial optimization solvers can handle the scale of the problems at hand[31], there is an argument to be made about tackling these problems without resorting to aggregations.

5.3. Clearing Reserve after Energy

Italy and Spain both clear reserve after energy. They are based on similar philosophies except for the zonal granularity.

Like all other zonal markets of the EU, the Italian market can be described in terms of the sequence of a day-ahead and intraday market followed by re-dispatching and balancing. But all these phases offer significant differences with what is found in the rest of the EU. In particular, the sophistication of the re-dispatching and its relation with balancing are of particular relevance for scarcity pricing.

The day-ahead Italian market is sub-zonal in the sense that it comprises several zones that account for transmission constraints in approximate form. It is also cleared by its own algorithm. The intraday market departs from the target model in the sense that it consists of a sequence of 5 auctions (it is not a continuous order book market) constructed along the same spatial structure and market clearing logic as the day-ahead market. There is no rupture of the trading pattern between day-ahead and intraday.

These markets are marginal price based. Producers are remunerated at the regional zonal price, consumers pay a unique average price (the PUN) the calculation of which is part of the equilibrium algorithm.

The TSO takes over after the closing of the day-ahead and intraday market. The TSO works with a nodal representation of the grid and procures ancillary services for reserve, re-dispatching and balancing. This is

[31] The scale of unit commitment problems that can be solved has improved significantly in recent years. See, for example, https://arpa-e.energy.gov/sites/default/files/13_PNNL_GD_OP-15_HIPPO2_public.pdf.

achieved through co-optimization, which means that re-dispatching, seen as an energy activity, is co-optimized with reserve and balancing. This ancillary service activity can thus be considered as a version of a gross pool where most agents must participate. This action implies a correction of the clearing in quantity of the day-ahead and intraday markets to the extent that it accounts for the technical constraints that were only approximated (in the zonal description) in the day-ahead and intraday market. But because this takes place in a pay-as-bid regime, there is no change of price, which thus remains compatible with the EU day-ahead market. Because of the different granularity there is rupture of structure between the energy and ancillary service market that would imply, in case ancillary services were cleared as single-price auction and following Parsons' argument (and basic economic logic) that a virtual trading system between the ancillary market seen as a co-optimization of energy and reserve and the day-ahead/intraday market is unlikely to be efficient. Virtual trading thus seems excluded by construction.

Notwithstanding the above, the (intuitive) implication is that an "ancillary service market" provides a natural basis for implementing the computation of a scarcity signal of the ORDC type. The signal would not add to the energy price in re-dispatching which is pay as bid, but this is not necessary to get the signal: it would even permit keeping differentiated signals for reserves of different flexibility. In contrast with the usual European view, re-dispatching takes balancing needs forecasts into account: besides the pay-as-bid remuneration and the pricing of imbalances that is not single price, the ancillary services market thus seems to capture (through co-optimization) most of what we need for scarcity pricing. It indeed looks like a full real-time market, which corrects day-ahead volumes (but not prices) by taking network constraints on board.

There are other issues to consider, some positive, some negative (for scarcity pricing) On the positive side this activity of the ancillary services market takes place through a sequence of sessions each coming after the energy intraday session and before entering in the balancing session before real time. As mentioned above, plants are remunerated as pay-as-bid, which means that the ancillary services do not produce another set of prices (different from those from the energy market clearing, which would have caused problems). On the negative side, the pricing of imbalances follows the usual dual pricing system for qualified units (those that participate to balancing), which blurs the scarcity signal in real time.

Summing up, the argument is that assigning reserves after the clearing of the energy market in the day ahead offers the opportunity to insert a scarcity premium calculation that has some (possibly a lot) of the flavor of the original scarcity premium.

6. Comparison of Designs

In this section we compare the performance of the different market designs for a period over which we have access to recent and detailed market data. This data corresponds to the period of September 2015 - March 2016, which was also the basis for the analysis presented in [18].

6.1. Characteristics of the Belgian System

Loads. We assume that the day-ahead demand is equal to the historically observed net demand, after removing imports. We assume an inelastic load and we ignore transmission constraints. We will assume a value of lost load equal to 8300 €/MWh, based on an estimate of the Belgian Federal Planning Bureau [28].

Generators. We consider the same mix of technologies as in previous research [1], [18]: pumped storage, blast furnace, renewable, gas-oil, LVN, coal (3 units) and combined cycle gas turbines (11 units, of which 3 are placed in strategic reserve). The marginal cost consists of the fuel cost and the CO2 emissions cost. We use the CO2 prices, emissions rates, and fuel price data from previous research [18]. The production of nuclear, wind, waste, and water are assumed to be price-inelastic, based on previous analyses [1], [18], and their production is subtracted directly from the system demand.

The fixed cost consists of startup cost and startup fuel, which we assume is incurred once per day (in the sense that for every hour that a unit is on, it must incur a cost which is 1/24 of the startup cost, so that if a unit is on for an entire day, it incurs a cost equal to its full startup cost). Additionally, we account for the minimum load fuel consumption of a generator.

Planned outages are accounted for in the data. The production capacity is scaled according to a capacity scaling factor which captures these forced outages. We ignore unplanned generator outages (assuming that they are captured implicitly in the imbalance scenarios), and simply set the real-time power generation capacity equal to the day-ahead capacity.

We set the ramp rate of all units except for CCGT units and pumped hydro units equal to zero. We assume that the fast ramp capacity is equal to half of the slow ramp capacity, which is equal to the 15-minute ramp rate of units. This is due to the fact that we associate fast capacity to secondary reserve, which is assumed to have a response time of 7.5 minutes. We associate slow capacity to tertiary reserve, which is assumed to have a response time of 15 minutes.

Pumped hydro. We account for pumped hydro resources by adding the six generators and four pumps that are located at the Coo pumped hydro facility in Belgium. In production mode, there exist three pairs of 144/215 MW generators, and three pairs of 145/200 MW pumps. We then use the total water storage capacity of Coo (8,450,000 m^3), and the head height (245 m for each of the two pumped hydro reservoirs) to compute the total energy storage capacity.

We employ a separate model for pumped hydro resources, which we do not develop here in order to avoid overburdening the notation, and since it is not required for describing the equilibrium formulation. For the sake of simplicity, we assume that pumped hydro resources do not offer reserve, otherwise it would be necessary to employ a multi-stage stochastic program in order to properly account for the random activation of reserve during the day. This random activation could result in binding operational constraints, and also would use up water which has value that depends on market prices. In order to focus the paper on the formation of reserve prices under scarcity pricing, we do not attempt to model this level of complexity since it would distract from the main purpose of the report. The assumption that pumped hydro resources are not contributing to reserve is of minor significance, since pumped hydro resources are still allowed to increase their production in real time up to the level of the maximum production capacity under tight system conditions. We note that our assumption that hydro resources do not commit capacity in reserve auctions is not far from empirical data. We have access to historical data from 2017, and note that pumped hydro resources contribute during some days to secondary reserve capacity, but only to a limited extent.

System operator. In order to design the ORDC, we need to account for the fact that fast and slow reserve can substitute for each other. As explained in [3], the valuation of the system operator for fast reserve capacity can be derived as follows:

$$V^{R,F}(r^F; r^{S,0}) = (VOLL - \widehat{MC}(\sum_g p_g)) \cdot (0.5 \cdot LOLP_{7.5}(r^F) + 0.5 \cdot LOLP_{15}(r^{S,0} + r^F))$$

$$V^{R,S}(r^S; r^{F,0}) = (VOLL - \widehat{MC}(\sum_g p_g)) \cdot 0.5 \cdot LOLP_{15}(r^S + r^{F,0})$$

where r^F is the amount of fast reserve capacity, r^S is the amount of slow reserve capacity, $r^{F,0}$ and $r^{S,0}$ are reference values for these capacities[32], $VOLL$ is the value of lost load, $\widehat{MC}(\sum_g p_g)$ is a proxy of the marginal cost of the marginal unit[33], and $LOLP_t$ is the loss of load probability given the uncertainty that the system is facing in the following t minutes.

We consider an operating reserve demand curve which is identical in the day ahead and real time. In computing $LOLP_{7.5}$, we assume perfectly correlated increments of uncertainty, following the statistical analysis of [18]. We use the parameters of the imbalance distribution shown in Table 6 in order to calibrate loss of load probabilities for every season and every four-hour block of every season.

Season	Hours	Mean [MW]	Standard deviation [MW]
Winter	1, 2, 23, 24	29.5	165.4
	3-6	23.6	147.8
	7-10	16.6	181.3
	11-14	-20.9	224.1
	15-18	8.1	162.4
	19-22	9.8	147.2
Spring	1, 2, 23, 24	28.4	147.9
	3-6	42.3	131.3
	7-10	27.8	151.3
	11-14	68.4	174.9
	15-18	69.0	161.5
	19-22	9.0	134.3
Summer	1, 2, 23, 24	20.1	133.1

[32] These are reference values of system capacity around which we linearize the marginal benefit to the system of additional capacity. A reasonable choice for the case of Belgium could be the same requirements as in the previous studies, namely 350 MW of the *R3 production* tertiary reserve product and 140 MW of secondary reserve.

[33] We use 25 €/MWh for this study, although this can be refined to more closely approximate the real-time system lambda.

	3-6	42.5	111.5
	7-10	25.8	132.1
	11-14	34.8	154.4
	15-18	47.1	140.3
	19-22	13.5	108.8
Fall	1, 2, 23, 24	29.2	138.7
	3-6	28.9	105.9
	7-10	-11.2	142.8
	11-14	18.5	164.9
	15-18	0.2	142.8
	19-22	-10.8	147.2

Table 6: Mean and standard deviation used for the estimation of LOLP15.

In the case of demand-side tertiary reserve product, it is straightforward to introduce it to the models. Namely, we augment the demand function for slow reserve, and we place a limit on the amount of reserve that demand response can offer which corresponds to the amount of *ICH* capacity. Since demand response reserve capacity is typically available at zero opportunity cost (if a load is consuming, it can offer its consumption as demand response capacity), any extra reserve demand that is requested by the system will always be served first by demand response. Thus, the effect of adding extra reserve demand and at the same time increasing the amount of reserve that can be satisfied by demand are two effects that cancel each other out.

Uncertainty. The overall 15-minute uncertainty in the system is characterized by the parameters of Table 6. These parameters are based on the imbalance data of 2017. In order to derive real-time demand scenarios, we use this data in conjunction with the data of figure 8 of [21]. In that figure, we observe the distribution of dynamic sizing requirements. We will assume that these requirements correspond to a factor that inflates real-time system imbalance. We have eight possible scenarios of "inflation", where the inflation factor corresponds to first-stage uncertainty. In order to define scenarios for the second stage, we multiply this inflation factor by a discretized normal distribution of imbalances, which is calibrated using the data of Table 6. The inflation factors and their corresponding probabilities can be observed directly in figure 8 of [21].

The transition probabilities from the second to the third stage are chosen so that we capture outliers (2 scenarios with probability 0.1% each) and we discretize the remaining mass of the distribution in evenly spaced "buckets" of mass.

Strategic reserve. We assume that strategic reserve capacity can contribute in real time at a very high cost, which is still below VOLL but above the marginal cost of the most expensive unit. This corresponds to a total capacity of 375 MW (Esche) + 485 MW (Seraing) + 385 MW (Vilvoorde), equal to 1245 MW in total. Our justification is that the commitment of strategic reserve capacity is the last resort before load shedding[34].

Other. We assume that reserve capacity is cleared daily, despite the fact that in the period of the case study the reserve auctions were weekly. The rationale for this choice is that, in practice, generators can trade their reserve obligations even after the week-ahead auction.

[34] http://www.elia.be/~/media/files/Elia/Products-and-services/Strategic-Reserve/SFR-2017-18_fr_final.pdf

6.2. Tested Designs

We will present our analysis as an evolution from the existing Belgian market design (REP) to a two-settlement system (SCV). We will report day-ahead prices for energy and reserves and the profits of CCGT units and loads as our key performance indicators. Our focus on CCGT units and loads is due to the motivation of the study for examining the incentives to invest in flexible capacity, but we remark that the model is set up in order to compute all relevant economic indicators, including system welfare. Of course, system welfare and the degree to which the market sends signals for adequate investment are closely linked.

Before presenting the results, we proceed to discuss the various market designs that we consider[35]. Each model represents an evolution with respect to the existing market design. We introduce one incremental change to the Belgian market design (REP) at a time, until we arrive to a two-settlement system (SCV), following the structure of sections 3-5. We then analyze the impact of each change on market prices and generator profits. The path that we consider is not the only one from REP to SCV, but is rather based on our assessment of increasingly disruptive changes.

REP-0.1. The REP-0.1 model represents our first proxy of current Belgian market operations. The clearing of reserve precedes the clearing of energy in the day-ahead market. No virtual trading is allowed. There is no market for real-time reserve capacity. Deviations are penalized at 10% of the marginal cost of generators. The day-ahead demand for reserve capacity is set at the Belgian reserve requirements: 140 MW for secondary reserve (aFRR), and 350 MW for tertiary reserve[36] (mFRR). Compared to the existing status of the Belgian market, this would require a transition to a uniform price for balancing, meaning that upward and downward activation should be priced at the same value, which would be the imbalance price. The rationale here is that the demand side of the real-time energy market (the entities causing imbalances) should be paid the same price as the supply side of the real-time energy market (the entities offering upward or downward activation). At this stage, we can also insert an energy adder, although this is not expected to have any effect, neither on system dispatch, nor on market clearing prices[37].

---

[35] We have additionally considered one more design, REP-0.1-ORDC, for which we do not present results in order not to disrupt the flow of information in the report. This design can be seen as an intermediate step between RCP-0.1 and REP-0.1, and its main attribute is the introduction of a day-ahead ORDC. Concretely, the REP-0.1-ORDC model is an evolution of the Belgian market design whereby inelastic day-ahead demand for reserve is replaced by a day-ahead operating reserve demand curve. The day-ahead ORDC is designed according to the procedure of section 6.1. This evolution is minimally disruptive, and generally in line with the revamping of day-ahead reserve procurement processes that is ongoing in Belgium, and which is related to the dynamic sizing of day-ahead reserve requirements [21].

[36] Note that this is the requirement for tertiary reserve from production (R3Prod). In particular, (i) we are ignoring the tertiary reserve requirements covered by demand. This has no effect on the model, since the requirements for tertiary reserve offered by demand can easily be covered by scheduled demand, and demand is not eligible for covering R3Prod requirements. Also, (ii) we are ignoring R3 dynamic profile. The definition of R3 dynamic profile is idiosyncratic (https://www.elia.be/~/media/files/Elia/Products-and-services/ProductSheets/S-Ondersteuning-net/S8_The-tertiary-reserve-Dynamic-Profile.pdf) and deviates from the simple definition used in this report of a resource that is capable to respond within 15 minutes.

[37] Effectively, generators can adjust their bids in order to internalize the adder (for example, an adder of 1 €/MWh which generators can anticipate will imply that generators will simply adjust their energy bid by 1 €/MWh down, and therefore the same exact outcome will prevail in terms of both dispatch as well as market clearing price as if the adder did not exist in the first place).

Note that we allow generators to offer different levels of reserve capacity from one hour to the next, even if reserve is cleared daily. This is confirmed by the data that has been provided to us by ELIA, where we see that generators nominate different amounts of reserve capacity for different periods of the day, even though the reserve auctions are cleared daily. We have tested the alternative of imposing the same reserve capacity from the same individual unit over all hours of the day, and the impact on the results has been observed to be negligible.

REP-0.1-Inelastic. The REP-0.1-Inelastic model represents our second proxy of the Belgian market. In this model, the following constraint has been introduced in the real-time market:

$$r_g^{DA} - r_{g,\omega'}^{RT} \leq 0.$$

This constraint requires that the real-time reserve capacity be at least at the level of the day-ahead reserve capacity. Note that $r_{g,\omega'}^{RT}$ corresponds to the excess headroom that is left over in a unit *after* an imbalance has been cleared within a given imbalance interval. Whether or not this constraint should be imposed depends on whether or not the system operator demands that units make their reserve capacity available, even *after* they have been activated for clearing an imbalance. The argument for removing this constraint is that a generator that has supported the system within an imbalance interval should not be held accountable for reserve capacity shortfall at the *end* of the imbalance interval. The argument for keeping the constraint is that the end of one imbalance interval signifies the beginning of a new imbalance interval, and therefore the system should be prepared, anew, to balance real-time uncertainty. The REP-0.1 model corresponds to the first point of view (not including the constraints). The REP-0.1-Inelastic model corresponds to the second point of view (including the constraints). Together these models envelope our proxy of current Belgian market operations.

RCP-0.1. This model emerges from the introduction of a real-time operating reserve capacity product in the Belgian electricity market. The demand function for operating reserve is identical to the day-ahead demand function. This would concretely imply that free bids should be paid for the real-time capacity that they make available at the reserve price determined by the real-time ORDC, and that generators should pay for / buy back the reserve capacity that is activated in real time.

RCP. This model emerges from the removal of administrative penalties for real-time deviations from day-ahead positions. This relates to the treatment of balancing as the spot market, rather than a service that maintains day-ahead net positions. This evolution effectively moves away from the notion of balancing responsible parties having to maintain their day-ahead trading positions, and may therefore be a more disruptive evolution compared to the settlement changes introduced in the previous steps. In terms of practical operations, the implementation of this measure could be facilitated by a transition to central dispatch in real time, whereby the system operator issues instructions to individual generators and posts real-time prices that are consistent with these instructions. This should be contrasted to the current paradigm, whereby balancing responsible parties cannot support the system needs (e.g. by increasing their net injection when the system is short) unless they speculate on the imbalance price value. This implies a risk for BRPs which would be alleviated in central dispatch, since in central dispatch the set point is determined by the system operator, and is consistent with the needs of the system (as expressed in the real-time price) as well as the profit maximizing behavior of individual resources.

RCV. This model is the evolution of RCP whereby we lift the physical constraints on the trading of reserve in the day-ahead market. We impose the requirement that any reserve that is contracted in the day-ahead reserve auction must be backed up by the commitment of generators in the day-ahead time frame. This corresponds to the day-ahead nomination of generators which is applied in the Belgian market, according to which reserve commitments must be backed up by physical capacity in the day-ahead time frame. On the other hand, energy is traded freely (i.e. without the backing of physical assets) in the day-ahead time frame, which corresponds to a departure from the current practice of the Belgian market, at least in principle.

SCV. This model is the evolution of RCV, whereby the energy and reserve markets are cleared simultaneously in the day ahead. Nevertheless, we maintain constraints that limit virtual trading on reserves in the day-ahead time frame. For example, a generator cannot commit reserve beyond its ramp rate limit, and cannot commit reserve unless a resource is committed. This is in line with US design, whereby virtual trading in reserve markets is precluded by the fact that reserve can only be made available by non-virtual traders.

6.3. Prices

In what follows, we present average results of our models and historically observed energy and reserve prices (both day-ahead as well as real-time). The model results are the average values over the full scenario tree described previously, whereas the historical realizations can be thought of as sample realizations over this tree.

Month	SCV	RCV	RCP	RCP-0.1	REP-0.1	REP-0.1 inelastic	Hist. DA	Hist. RT
1	41.00	41.00	41.00	41.12	35.61	56.39	52.50	39.51
2	31.17	31.17	31.17	31.31	30.68	32.63	55.41	61.04
3	46.88	46.88	46.88	47.05	30.60	66.23	43.12	36.57
4	37.44	37.38	37.36	37.49	28.77	51.50	35.94	33.31
5	41.25	41.25	41.25	41.32	27.17	63.36	32.61	29.48
6	21.74	21.71	21.73	21.87	19.61	26.42	25.39	21.80
7	21.14	21.12	21.13	21.23	20.74	21.44	27.13	25.11
Average	34.37	34.36	34.36	34.48	27.60	45.42	38.87	35.26

Table 7: Energy price (€/MWh) for the models considered in the comparison.

We present the energy prices for the different designs in Table 7. The average real-time prices are identical. Note that the historical average day-ahead price is greater than the average historical real-time price, indicating a risk premium associated with day-ahead trading. The REP-0.1 and REP-0.1-inelastic models, which are the closest proxies to the current Belgian design, are enveloping the historically observed day-ahead and real-time energy prices. Recall that the two models differ in terms of whether or not reserve capacity is required to be available after the activation of reserve within an imbalance interval. It is evident that this requirement may have a very significant impact on prices.

The REP-0.1-inelastic model effectively corresponds to imposing an inelastic ORDC in the real-time market. As we will explain later, this results in an over-valuation of reserve capacity. On the other extreme, the REP-0.1 completely removes the ORDC from the real-time market. This results in an under-valuation of reserve. The advantage of an ORDC which is designed on the basis of loss of load probability is that the valuation of reserve self-adjusts to reasonable levels, where we will explain “reasonable” later in the profit analysis.

Note that the price reduction that occurs in the REP-0.1 model cannot be arrested by the introduction of a day-ahead ORDC, which is what we model in REP-0.1-ORDC. This can be explained by the KKT conditions of the REP models. Effectively, in removing the ORDC from the real-time market, this market design creates a disconnect between the real-time value of capacity (as reflected in real-time energy prices) and the day-ahead value of reserve (as reflected in day-ahead reserve prices). Introducing a day-ahead ORDC cannot amend this defect.

The price reduction observed in the energy price of REP-0.1 and REP-0.1-ORDC can be arrested by the introduction of a real-time market for reserve capacity. This restores the connection between real-time energy prices and the real-time value of reserve capacity. This explains the increase in energy prices which is observed in the RCP-0.1 and RCP models. The transition to virtual trading has a negligible effect on prices for the risk-neutral case. The transition to simultaneous clearing of energy and reserves has a more notable, however still minor, effect on prices. This effect is due to the fact that, in the case of simultaneous clearing, agents are allocating their resources with more information at hand.

The energy prices are largely linked to reserve prices, due to no-arbitrage conditions. Therefore, in order to understand the energy prices, we focus on understanding reserve prices, which we analyze below. Note that, in the following tables, the day-ahead and real-time prices produced by our stochastic equilibrium models were almost identical. We therefore limit ourselves to presenting the day-ahead prices. This convergence between day-ahead and real-time prices is likely possible to attribute to our assumption that all agents are risk-neutral.

Month	SCV	RCV	RCP	RCP-0.1	REP-0.1	REP-0.1 inelastic	Hist. DA
1	17.60	17.60	17.60	17.62	1.37	34.00	10.90
2	9.01	9.04	9.03	9.08	1.25	11.08	8.67
3	22.73	22.73	22.73	22.83	1.25	42.07	11.79
4	21.16	21.34	21.31	21.35	2.90	34.87	10.57
5	24.50	24.52	24.51	24.49	1.07	46.30	9.25
6	8.77	8.83	8.87	8.81	1.11	13.49	7.69
7	6.39	6.42	6.40	6.36	0.97	6.55	8.28
Average	15.74	15.78	15.78	15.79	1.42	26.90	9.59
Average (RT)	15.56	15.69	15.65	15.15	N/A	N/A	N/A

Table 8: Day ahead prices of fast reserve (€/MW-h) for the models considered in the comparison.

The price for fast reserve is shown in Table 8 and for slow reserve in Table 9. We present reserve prices for all models, as well as historical prices. The data source for the historical reserve price data is ELIA[38]. Note that, in contrast to our model, in reality there is no substitutability between secondary and tertiary reserves in the Belgian market.

There are occasional differences between day-ahead and average real-time prices, but they are minor. For the REP models, this price effectively reflects the marginal cost of activating resources in the day-ahead time frame, and is not necessarily reflective of the real-time value of reserve in keeping loss of load probability in check. Note that this remains true even when we introduce ORDC in the day-ahead time

[38] http://www.elia.be/en/suppliers/purchasing-categories/energy-purchases/Ancillary-services/Ancillary-Services-Volumes-Prices

frame. For the RCP, RCV and SCV models, the price of reserve is driven by the demand side, and relates to the value of reserve in keeping a low loss of load probability.

Month	**SCV**	**RCV**	**RCP**	**RCP-0.1**	**REP-0.1**	**REP-0.1 inelastic**	**Hist. DA**
1	10.82	10.83	10.86	11.00	1.37	34.00	4.66
2	6.32	6.34	6.34	6.43	1.25	11.08	4.66
3	13.73	13.75	13.75	14.01	1.25	42.07	4.66
4	14.34	14.21	14.44	14.52	2.90	34.78	4.66
5	16.77	16.76	16.79	16.83	1.07	46.30	4.71
6	6.90	6.87	6.91	6.93	1.11	13.49	6.08
7	5.02	5.00	5.02	5.00	0.97	6.55	7.46
Average	10.56	10.54	10.59	10.67	1.42	26.90	5.27
Average (RT)	10.70	10.54	10.52	10.17	N/A	N/A	N/A

Table 9: Day ahead prices of slow reserve (€/MW-h) for the models considered in the comparison.

6.4. Profits

In Table 10 we present the profit results for the 8 CCGT units that were active in the market during the test period. Note that three other CCGT units were available on strategic reserve, and were therefore not actively producing power in the energy market. The profit results that we report are discounted by the capacity of the generators. It is worth comparing these results to the capital investment cost of a typical CCGT unit. We assume an overnight cost of 676 $/kW (EIA 2012 estimate), an exchange rate of 0.88 €/$, annual discounting at a rate of return r, and T years of investment lifetime. We consider a range of r from 8 to 12%, and an investment lifetime of 25 to 30 years. This gives CCGT investment costs ranging from 6.03 €/MWh to 8.66 €/MWh.

	SCV	RCP	RCP	RCP-0.1	REP-0.1	REP-0.1 inelastic
G1	*7.37*	*7.37*	*7.37*	*7.40*	**2.59**	16.15
G2	20.68	20.66	20.68	20.79	15.07	31.80
G3	*8.06*	*8.06*	*8.06*	*8.09*	**2.64**	19.03
G4	12.04	12.04	12.04	12.08	**3.84**	28.62
G5	21.07	21.05	21.07	21.18	15.45	32.26
G6	*8.30*	*8.29*	*8.30*	*8.32*	**2.66**	19.42
G7	21.45	21.43	21.45	21.56	15.82	32.57
G8	20.58	20.56	20.58	20.69	14.93	31.67

Table 10: Generator profits (€/MWh) for the models considered in the comparison.

We have indicated the entries of Table 10 according to how they compare to the running investment cost of a typical CCGT unit. Generators that are below 6.03 €/MWh are indicated in bold font, and correspond to generators that are unable to recover fixed costs, even under optimistic assumptions about fixed costs. Generators that are in the range of 6.03 - 8.66 €/MWh are indicated in italic font, and correspond to units that earning a profit within the range of investment costs. These units are breaking even. Units indicated in normal font are earning a profit above, 8.66 €/MWh, and are therefore covering investment costs even under pessimistic investment requirements.

We observe that the two REP envelope models, REP-0.1 and REP-0.1-inelastic, cover a wide range of generator profits. Therefore, the extent to which generator capacity *after* activation must correspond to reserve capacity committed in the day-ahead market can shift a unit from making losses to earning excessive profits. Removing this requirement altogether, which is the case in the REP-0.1 (and REP-0.1-ORDC) models, places 4 out of 8 units in a non-viable financial position. The introduction of a real-time market for reserve capacity (RCP and RCP-0.1) restores 3 of these units to breaking even, and 1 of them to covering its investment costs comfortably. In a risk-neutral environment, this is the key market design change, with the introduction of virtual trading and simultaneous clearing of energy and reserve having a secondary impact on the profitability of units.

In Table 11 we present the profits of generators after having corrected for unaccounted fixed costs. Concretely, for every period when a generator is partially active ($0 < y_{g,\omega} < 1$), we compute the difference between the full minimum load cost of the generator and the cost incurred by only activating the generator up to $y_{g,\omega}$. This is an underestimation of the impact of fixed costs on profits, since we disregard startup costs. The effect of this unaccounted fixed cost on generator profits is notable only in the case of generators that are not breaking even. By contrast, generators that are earning comfortable financial profits are also fully committed. Regardless, the effect of this unaccounted fixed cost is minor (less than 1 €/MWh for all units) and does not have an impact on the financial viability of any units.

	SCV	**RCV**	**RCP**	**RCP-0.1**	**REP-0.1**	**REP-0.1 inelastic**
G1	*6.99*	*6.98*	*6.99*	*7.01*	**2.36**	15.93
G2	20.68	20.66	20.68	20.79	15.07	31.80
G3	*7.74*	*7.73*	*7.74*	*7.77*	**2.41**	18.87
G4	10.68	10.68	10.68	10.66	**3.46**	27.98
G5	21.07	21.05	21.07	21.18	15.45	19.96
G6	*7.78*	*7.75*	*7.78*	*7.77*	**2.47**	19.22
G7	21.45	21.43	21.45	21.56	15.82	32.57
G8	20.58	20.56	20.58	20.69	14.93	31.67

Table 11: Generator profits (€/MWh) for the models considered in the comparison, with corrected fixed costs.

We conduct a similar analysis for the implications of scarcity pricing for the profitability of loads. We summarize our results in Table 12. The table reads as follows: (i) The second column presents the decrease in the profits of loads, relative to REP-0.1, under the assumption that loads do not offer any reserve to the market. Note that this is the *total* increase in the consumer bill from the introduction of scarcity pricing, divided by the total average demand during the study, which amounts to 7442 MW. (ii) The third column is the monthly increment in profit that loads enjoy by offering an additional MW of ramp capacity into the system. This increment is the result of their ability to offer the additional capacity for secondary and/or tertiary reserve[39]. (iii) The fourth column is the amount of reserve capacity that the loads would need to offer to the reserve market in order to offset their losses from the increase in energy prices which results

[39] Concretely, we compute this value by introducing 1 MW of reserve that can be offered by loads and measure the difference in profits compared to not being able to offer that MW (in the sense of increasing the ramp rate of loads from 0 MW/min to 1/60 MW/min). In doing so, we do not perturb the market equilibrium (in the sense of the reserve price shifting due to the introduction of the load).

from scarcity pricing. Any amount of capacity above this level would result in the market design in question creating a net benefit for loads, relative to the design with the lowest electricity price (REP-0.1).

	Load profit decrease (€/MW-month)	**Δ Profit / Δ Reserve (€/MW-month)**	**Break-even reserve capacity (MW)**
REP-0.1	-	-	-
REP-0.1-inelastic	5,676	52,819	926.4
RCP-0.1	5,227	57,154	680.6
RCP	5,111	66,010	576.2
RCV	5,109	66,126	575.0
SCV	5,120	65,877	578.4

Table 12: Load profitability under the different designs.

As expected, the introduction of scarcity pricing reduces the profits of loads under the assumption that loads cannot offer reserves and monetize their flexibility in the reserve market. This can be seen by observing, in the second column, that the REP-0.1 design, which is the design with the lowest energy prices and one of the two proxies of the current Belgian market design, results in the highest profits for loads. On the other extreme, the REP-0.1-inelastic design entails the greatest energy prices and the least profit for loads. The introduction of scarcity pricing in the RCP models lifts the prices of both energy and reserve. If loads cannot offer reserve to the market, this clearly entails a net loss. On the other hand, the introduction of the real-time reserve market under the RCP designs also revives the reserve price, and the third column of the table indicates the amount of reserve that loads would need to offer to the market in order to be able to offset the losses that they incur from the increase of energy price that results from scarcity pricing.

The ratio of the second and third column of Table 12 indicate that, for every MW of load in the system, the additional expense that results from the introduction of scarcity pricing can be recuperated by making approximately 7.8% (for RCP, RCV, and SCV) to 10.7% (for REP-0.1-inelastic) of that capacity available in the reserve market. Any additional load capacity that can be made available in the reserve market stands to gain from the introduction of scarcity pricing.

Reserve prices drive the profitability of flexible resources. Therefore, getting real-time reserve prices right is a fundamental aspect of sound market design in an environment of large-scale renewable energy integration, where flexible resources are needed for supporting system security.

A major difficulty with the absence of a real-time reserve market is that it becomes difficult to value reserve precisely. Using the no-arbitrage conditions of the stochastic equilibrium model that we have developed for this study, the back-propagation of the day-ahead price of reserve capacity when we are not forced to carry any reserve capacity after activation (REP-0.1 model) can be expressed as follows:

$$\lambda^{R,DA} = \beta_g^{G,DA} + \mathbb{E}\left[\alpha_{g,\omega}^{G,DA}\right],$$

where $\beta_g^{G,DA}$ corresponds to ramping scarcity in the day-ahead market and $\alpha_{g,\omega}^{G,DA}$ corresponds to capacity scarcity in the day-ahead market. This signal is too *weak* to signal scarcity in the system.

On the other hand, when we are forced to carry the full amount of reserve after activation (REP-0.1-Inelastic model), the scarcity signal is too *strong*:

$$\lambda^{R,DA} = \beta_g^{G,DA} + \mathbb{E}\left[\alpha_{g,\omega}^{G,DA}\right] + \mathbb{E}\left[\gamma_{g,\omega'}^{G,DA}\right],$$

where $\gamma_{g,\omega'}^{G,DA}$ corresponds to the requirement of carrying the day-ahead reserve after activation.

The real-time ORDC automates this calculation in a self-correcting fashion, and arbitrage propagates this price to the day-ahead market, thereby signaling investment in reserve capacity in case of tight system conditions:

$$\lambda^{R,DA} = \beta_g^{G,DA} + \mathbb{E}\left[\alpha_{g,\omega}^{G,DA}\right] + \mathbb{E}\left[\lambda_{\omega'}^{G,R,RT}\right].$$

7. Implementation

In this section we describe the implementation of scarcity pricing in Belgium. We first explain, on a conceptual level, how many adders are required. We then discuss the data that is required for their computation. We explain how this data is used in the scarcity pricing formula, and provide some examples that illustrate how scarcity pricing is used for the settlement of market participants. These formulas were used by ELIA[40] for the back-simulation of scarcity pricing in the Belgian market in 2017 [11].

7.1. The Three Adders

Scarcity pricing requires the introduction of three adders in the real-time market:

- The adder for 7,5' reserve capacity: remuneration paid for *standby* (non-activated) secondary reserve capacity in excess of what a generator has been cleared for in previous markets. For example, if 1 MW of secondary reserve capacity shows up in real time, whereas they had no obligation to show up, that 1 MW is paid the fast reserve capacity adder. This also works in the opposite direction: if a resource is providing 1 MW less of secondary capacity than what they had promised in the day-ahead market, they are charged the fast reserve capacity price times the 1 MW.
- The adder for 15' reserve capacity: remuneration/charge paid for *standby* (non-activated) tertiary reserve capacity in excess of what a generator had been cleared for in previous markets.
- The adder for energy: how the system real-time marginal price should be changed in order to pay for changes in real-time energy, compared to the day-ahead / intraday set-point (e.g. for activated upward energy).

The formulas for these adders are presented here for ease of reference. They are repeated later, and the logic behind their computation is explained.

The adder for 7.5' reserve capacity is derived as follows:

$$\lambda RRT_t^F = \frac{T_1}{T_1 + T_2}\left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1}(R_{T_1})$$

$$+\frac{T_2}{T_1 + T_2}\left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1+T_2}(R_{T_1+T_2})$$

The adder for 15' reserve capacity is derived as follows:

$$\lambda RRT_t^S = \frac{T_2}{T_1 + T_2}\left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1+T_2}(R_{T_1+T_2})$$

The adder for the energy price is equal to λRRT_t^F.

[40] Note that, as ELIA explains in their report, their calculation did not account for the effect of back-propagation.

7.2. Required Data

The data which is required for the implementation of scarcity pricing can be described as follows.

- Historical data of system imbalances (Imb)
- Static data for computation of the adder
  - Value of lost load ($VOLL$)
  - Response time for secondary reserve (T_1)
  - Response time for tertiary reserve (T_2)
- Strategic reserves
  - Technical minimum (SRPMin_g)
  - Technical maximum (SRPMax_g)
- Ramp rate ($RRSR_g$)
- Dynamic data for computation of the adder
  - Real-time system incremental cost, approximated by imbalance price (λImb_t)
  - Available demand response reserve capacity (DR)
  - Conventional units
    - Available capacity of units that can offer secondary reserve ($PMax_g$)
    - Scheduled set-point of units (pF_{gt})
    - On/off status of units (u_{gt})
    - Ramp rate of units (RR_g)
  - Hydro units
    - Technical max of pumped hydro units in production mode (PHPMax_g)
    - Scheduled set-point of pumped hydro units in production mode ($pPHF_{gt}$)
    - Ramp rate of pumped hydro units in production mode ($RRPH_g$)
- Payment settlement
  - Generator forward/nominated production (pF_{gt})
  - Load forward consumption[41] (dF_{gt})
  - Generator and load forward reserve commitment (rF_{gt})
  - Generator real-time production (pRT_{gt})
  - Load real-time consumption[42] (dRT_{gt})
  - Generator and load real-time reserve provision (rRT_{gt})

The formulas used in the following exposition are based on two preceding studies that were performed on behalf of the CREG [1], [18]. In these studies, we use reserve activation data in order to estimate imbalances, because we did not have direct data for imbalances.

CREG provided the following data for estimating the activation of upward capacity:

- VolumeFCRup: amount of activated upward FCR capacity
- R2up: amount of activated secondary reserve capacity

---

[41] As clarified by ELIA, the TSO does not have nominated schedules available for loads. Normal demand is hidden behind the BRP, and does not need to indicate forward positions. The TSO charges demand on the basis of the net position of the BRP. Resources that are offering reserves (e.g. big industry) must indicate positions individually.

[42] According to ELIA, this information will probably only be known per BRP. In the past, settlement was performed one year after real time.

- IChMwh: amount of ICh activated capacity

The CREG provided the following data regarding downward activation:

- VolumeFCRdown: amount of activated downward FCR capacity
- R2down= amount of activated secondary reserve capacity

The total positive imbalance, indicated $ImbP$, was estimated on the basis of the available data as follows.

$$ImbP = \text{VolumeFCRup} + \text{R2up+IChMwh}$$

The rationale of the above formula is that we measure the total upward imbalance as the sum of the upward reserve capacities that were activated in order to cover this imbalance.

The total negative imbalance, indicated $ImbN$, was estimated on the basis of the available data as follows.

$$ImbN = \text{VolumeFCRdown} + \text{R2down}$$

The rationale of the above formula is that we measure the total downward imbalance as the sum of the downward reserve capacities that were activated in order to cover this imbalance.

The total imbalance, indicated (Imb), is computed as follows.

- If $|ImbP| > |ImbN|$ then $Imb = ImbP$
- Else $Imb = -|ImbN|$

The rationale of the above is that we record, for every balancing interval, the worst-case imbalance in either the upward or the downward direction as the imbalance of the system for the balancing interval in question.

This imbalance data can then be used in order to fit a distribution of system imbalance. In practice, a Gaussian distribution of system imbalance is used in Texas. In order to specify a Gaussian distribution, it is necessary to determine its mean and standard deviation.

7.3. Constructing the Price Adders

The total scarcity adder is composed of two components. The first component captures the capacity value of resources that can respond within the time frame of secondary reserves, while the second component captures the capacity value of resources that can respond within the time frame of tertiary reserve response. The concept is illustrated in the following figure.

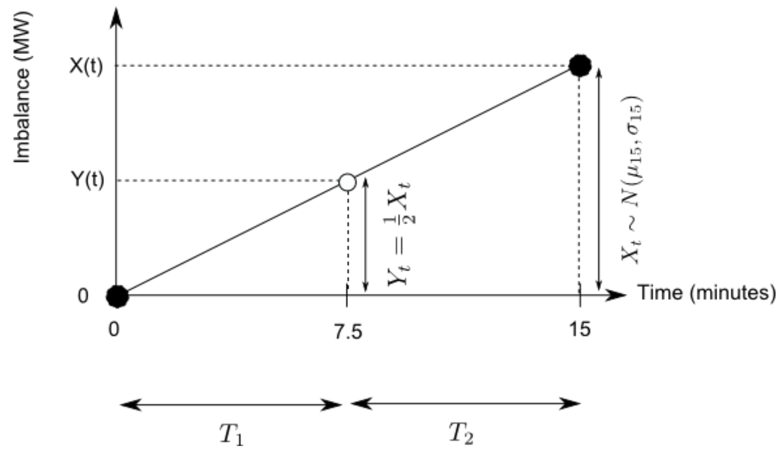


Figure 18: Decomposition of an imbalance interval into a part that corresponds to the response time of secondary reserves, and a part that corresponds to the response time of tertiary reserves.

We denote the duration of the first interval as T_1, and the duration of the second interval as T_2. Assuming that the overall imbalance that is observed within the interval is a result of a gradual evolution of imbalance towards its final value, we arrive to the following formula:

$$Y(t) = \frac{T_1}{T_1 + T_2} X(t),$$

where $Y(t)$ denotes the imbalance at T_1 minutes within the imbalance interval, and $X(t)$ denotes the imbalance at $T_1 + T_2$ minutes within the balancing interval.

The ORDC adder corrects both the energy price, as well as the reserve price. A different component of the adder can apply to secondary and tertiary reserve. Regarding the influence of the scarcity adder on reserve price, [3] provides the following formula for nesting reserve payments to fast and slow-moving reserves.

- Resources that contribute towards 7.5’ reserve capacity requirements receive the following reserve capacity price:

$$\lambda RRT_t^F = \frac{T_1}{T_1 + T_2}\left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1}(R_{T_1})$$
$$+ \frac{T_2}{T_1 + T_2}\left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1+T_2}(R_{T_1+T_2})$$

- Resources that contribute towards 15’ reserve capacity requirements receive the following reserve capacity price:

$$\lambda RRT_t^S = \frac{T_2}{T_1 + T_2}\left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1+T_2}(R_{T_1+T_2})$$

Note that resources which have a 7.5’ response time contribute also to 15’-minute reserve capacity requirements, and therefore receive both adders. The adder λRRT_t^F is also added to the imbalance price in order to settle real-time energy, as we explain in section 7.6.

Note that the argument within the LOLP functions is the leftover capacity *after* we have activated resources in order to clear the imbalance of the present interval[43]. For the quantity Imb_t we use the system imbalance (SI) of a given interval, as opposed to the net regulation volume (NRV) which can be thought of as a delayed version of SI. The subtraction of the term $\widehat{MC}\left(\sum_g p_{g,t}\right)$ captures the fact that there is a tradeoff between shedding load and dispatching the marginal unit to serve load.

The following data which is used in the adder formulas can be assumed to be constant over time, and we propose specific values that have been used in previous work [1], [18].

- Value of lost load: 8300 €/MWh (VOLL estimated by the Belgian federal planning bureau [28])
- Duration of the first part of the imbalance interval: T_1 = 7.5 minutes
- Duration of the second part of the imbalance interval: T_2 = 7.5 minutes

The following data, which is used in the adder formulas, changes over time depending on the conditions of the system.

- A function introduced by Hogan as the incremental cost for meeting an additional increment in demand: $\widehat{MC}\left(\sum_g p_{g,t}\right)$. For this value, one could employ the merit order function of the system when transmission constraints are ignored. Previous studies for Belgium [1], [18] have employed the imbalance price, i.e. $\widehat{MC}\left(\sum_g p_{g,t}\right) = \lambda Imb_t$. The rationale behind this choice is that the system imbalance price should be a reasonable proxy of the incremental cost for meeting an additional increment in demand for Belgian real-time operations.
- The amount of reserve capacity that can respond within T_1 minutes. This is denoted as R_{T_1}, and its computation is described in detail in the following paragraphs.
- Reserve capacity that can respond within $T_1 + T_2$ minutes. This is denoted as $R_{T_1+T_2}$, and its computation is described in detail in the following paragraphs.

Having described the calculation of the ORDC adders, we now discuss how we calculate the amount of excess capacity R_{T_1} and $R_{T_1+T_2}$. What we describe resembles closely the computation of the Available Regulation Capacity (ARC) which is performed by ELIA.

The amount of excess capacity in the system depends on whether or not we have strategic reserve available. For months without strategic reserve, the available secondary reserve capacity is computed as follows:

$$R_{T_1} = CGCap_{T_1} + HCap_{T_1} + DR$$

The components in this formula correspond to the following:

- $CGCap_{T_1}$ is the available capacity of committed resources within T_1 minutes
- $HCap_{T_1}$ is the available capacity of pumped hydro resources within T_1 minutes

[43] In ERCOT, the computation of the adder takes place every 5 minutes, and the system is able to keep better track of the instantaneous reserve capacity after activation. For systems with larger imbalance intervals (e.g. 15 minutes), such as Belgium, we recommend the use of the net reserve capacity after imbalances, since this a better reflection of the stress on the system.

- DR is the available demand response capacity, where previous studies [18], [1] have assumed 27 MW of primary demand response reserve and 261 MW of secondary demand response reserve.

The logic of the preceding formula is that it adds up all the available capacity from conventional thermal plants, hydro plants, and demand response.

The formula for $CGCap$ is given by

$$CGCap_{T_1} = \min(\sum_{g \in GC} (\text{PMax}_g - pF_{gt}) \cdot u_{gt}, T_1 \cdot RR_g)$$

The data used in this formula are the following:

- GC is the set of resources that are available for offering secondary reserve, meaning resources that are cleared in reserve auctions as well as free bids
- PMax_g [MW] is the technical max of a resource
- pF_{gt} [MW] is the scheduled set-point of the unit
- u_{gt} is the on/off status of the unit
- RR_g [MW/min] is the ramp rate of the unit

The logic of the preceding formula is that it measures the amount of excess capacity that is available by a thermal resource. If the generator is on, then this capacity is capped by the technical maximum of the unit, or its ramp rate within the required response time, which is why we have different amounts of capacity (and different adder components) for a horizon of T_1 minutes versus a horizon of $T_1 + T_2$ minutes.

The formula for $HCap$ is given by

$$HCap_{T_1} = \min(\sum_{g \in PH} (\text{PHPMax}_g - pPHF_{gt}), T_1 \cdot RRPH_g)$$

The data used in this formula are the following:

- PH is the set of pumped hydro units
- PHPMax_g [MW] is the technical max of pumped hydro units in production mode
- $pPHF_{gt}$ [MW] is the scheduled set-point of pumped hydro units in production mode
- $RRPH_g$ [MW/min] is the ramp rate of pumped hydro units in production mode

The logic of this formula is that it measures the amount of excess capacity that can be made available by hydro units, which is capped by the production capacity of these units and their ramp rate.

For the computation of the amount of excess capacity that can be made available in $T_1 + T_2$ minutes, the same idea applies as for the formula which was used for measuring the capacity that can be made available in T_1 minutes:

$$R_{T_1+T_2} = CGCap_{T_1+T_2} + HCap_{T_1+T_2} + DR,$$

where

$$CGCap_{T_1+T_2} = \min(\sum_{g \in GC} (\text{PMax}_g - pF_{gt}) \cdot u_{gt}, (T_1 + T_2) \cdot RR_g)$$

and

$$HCap_{T_1+T_2} = \min(\sum_{g \in PH} (\text{PHPMax}_g - pPHF_{gt}), (T_1 + T_2) \cdot RRPH_g)$$

For months with strategic reserve, the following capacity is added to *both* R_{T_1} and $R_{T_1+T_2}$ [18]:

- 358.4 MW of demand response
- Strategic reserve units for which we did not have unit-specific technical data. This includes the capacity of Angleur (50 MW), and the capacity of Izegem (20 MW).
- Strategic reserve units for which we have unit-specific technical data.

Since we are adding the full capacity of demand response and strategic reserve units for which we do not have unit-specific data, we are making the optimistic assumption that these resources can provide upward balancing corresponding to their full capacity within T_1 minutes.

Regarding units for which we have technical data, we add the following capacity to R_{T_1} :

$$SRCap_{T_1} = \min\left(\sum_{g \in SR} (\text{SRPMax}_g - \text{SRPM}in_g), T_1 \cdot RRSR_g\right)$$

where

- SR is the set of strategic reserve units for which we had unit-specific technical data (Seraing, Vilvoorde, Esche-sur-Alzette)
- $RRSR_g$ is the ramp rate of strategic reserve units
- SRPMin_g is the technical minimum of strategic reserve units

A similar formula applies for adjusting $R_{T_1+T_2}$ in order to account for the contribution of strategic reserve:

$$SRCap_{T_1+T_2} = \min(\sum_{g \in SR} (\text{SRPMax}_g - \text{SRPM}in_g), (T_1 + T_2) \cdot RRSR_g)$$

7.4. Settlement

In this section we discuss the settlement in the energy and reserve market. As explained in the report, the energy price, including the adder, applies to all BRPs and BSPs. The reserve pricing applies to BSPs providing reserves, as well as free bids that show up in real time.

The following data is required for settlement purposes:

- The forward price for energy [€/MWh] (e.g. the day-ahead price), which we denote as λPF_t
- The forward price for reserve [€/MW-h] (e.g. day-ahead or month-ahead reserve *capacity* price), which we denote as λRF_t
- The real-time price for energy [€/MWh] (imbalance price), which we denote as λPRT_t
- The real-time price for reserve [€/MW-h], which we denote as λRRT_t. Note that such a price does not exist currently in Belgium, meaning that reserve capacity is currently not traded in real time in the Belgian market.

The revenues of a generator in the day-ahead and real-time energy and reserve market can be computed as follows. Consider a generator which is scheduled for pF_{gt} MW of production and rF_{gt} MW of reserve capacity in a given imbalance interval t. Suppose that the generator produces pRT_{gt} MW in real time, and makes available rRT_{gt} MW of reserve capacity in real time. The revenues of the producer are computed as follows:

- In the day ahead, the producer earns $\lambda PF_t \cdot pF_{gt} + \lambda RF_t \cdot rF_{gt}$
- In real time, the producer earns $\lambda PRT_t \cdot (pRT_{gt} - pF_{gt}) + \lambda RF_t \cdot (rRT_{gt} - rF_{gt})$, which means that the real-time prices apply to the changes in the day-ahead position. Note that this includes a settlement of the reserve capacity position in real time.

The revenues of a load in the day-ahead and real-time energy and reserve market can be computed as follows. Consider a load which is scheduled for dF_{gt} MW of load and rF_{gt} MW of reserve capacity in a given imbalance interval t. Suppose that the load consumes dRT_{gt} MW in real time, and makes available rRT_{gt} MW of reserve in real time. The revenues of the load are computed as follows:

- In the day ahead, the load earns $-\lambda PF_t \cdot dF_{gt} + \lambda RF_t \cdot rF_{gt}$
- In real time, the load earns $-\lambda PRT_t \cdot (dRT_{gt} - dF_{gt}) + \lambda RF_t \cdot (rRT_{gt} - rF_{gt})$

7.5. Putting It All Together with Examples

In this section, we provide a number of concrete examples that illustrate the application of the scarcity adder on the settlement of generators and loads under different circumstances. For the illustration of this example, we use the data that was used in the ELIA scarcity pricing incentive report [11]. We will focus on one of the most severe 15-minute intervals in the year, namely November 29th, at 6pm – 6:15pm. This period corresponds to a highly stressed condition of the system, where there is a significant depletion of hydro capacity and a significant forecast error which causes major imbalance and places the system under stress. Concretely, we will consider the following values for our examples:

- The parameters for the imbalance distribution are as follows: the mean is equal to 0.24 MW, and the standard deviation is equal to 142.8 MW
- Available reserve in T_1 minutes: R_{T_1}= 366.5 MW
- Available reserve in $T_1 + T_2$ minutes: $R_{T_1+T_2}$= 1013.5 MW
- System lambda: λImb = 310.0 €/MWh
- VOLL = 8300 €/MWh
- T_1 = 7.5 minutes
- T_2 = 7.5 minutes
- Imbalance of 673.5 MW

Assuming that the marginal unit is providing secondary reserve, the energy price with the adder equals λImb+λRRT_t^F. Then the contribution of the scarcity adders are as follows:

- The contribution of the adder to secondary reserve capacity prices amounts to
$$0.5 \cdot \left(VOLL - \widehat{MC}\left(\textstyle\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1}\left(R_{T_1} - 0.5 \cdot Imb\right) = 1264.9 \text{ €/MWh}$$
- The contribution of the adder to tertiary reserve capacity prices amounts to
$$0.5 \cdot \left(VOLL - \widehat{MC}\left(\textstyle\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1+T_2}\left(R_{T_1+T_2} - Imb\right) = 34.3 \text{ €/MWh}$$

This creates the three following adders:

- Adder for capacity deliverable in 7.5':

$$\tilde{\lambda}RRT_t^F = 0.5 \cdot \left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1}\left(R_{T_1} - 0.5 \cdot Imb\right) + 0.5 \cdot \left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1+T_2}\left(R_{T_1+T_2} - Imb\right) = 1299.2\ €/MWh$$

- Adder for capacity deliverable in 15':

$$\lambda RRT_t^S = 0.5 \cdot \left(VOLL - \widehat{MC}\left(\sum_g p_{g,t}\right)\right) \cdot LOLP_{T_1+T_2}\left(R_{T_1+T_2} - 0.5 \cdot Imb\right) = 34.3\ €/MWh$$

- Adder for energy:

$$\tilde{\lambda}RRT_t^F = 1299.2\ €/MWh$$

7.5.1. Benefits of the Mechanism for BSPs

We now consider the settlement of a generator. We use the following inputs for our example:

- The generator used in our example can provide secondary reserve
- The day-ahead energy price is λPF_t = 20 €/MWh
- The day-ahead secondary reserve price is $\tilde{\lambda}RF_t$ = 65 €/MWh
- The real-time secondary reserve price is $\tilde{\lambda}RRT_t^F$ = 1229.2 €/MWh; note that this is entirely due to the contribution of the adder
- The real-time energy price is λPRT_t = 1539.2 €/MWh; note that this is equal to λImb+$\tilde{\lambda}RRT_t^F$
- The capacity of the generator is PMax = 125 MW

The following table describes the settlement of a BSP, when the adder is not applied, in the absence of the mechanism. Thus, this corresponds to the current setup of the Belgian market, where a real-time reserve market is absent, hence there is no real-time settlement of reserve capacity.

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
Day-ahead energy	$\lambda PF_t \cdot pF_{gt}$	λPF_t =20 €/MWh	pF_{gt} = 0 MW	0
Day-ahead reserve	$\tilde{\lambda}RF_t \cdot rF_{gt}$	rF_{gt} = 65 €/MWh	rF_{gt} = 25 MW	1,625
Real-time energy	$\lambda PRT_t \cdot (pRT_{gt} - pF_{gt})$	λPRT_t = 300.0 €/MWh	$pRT_{gt} - pF_{gt}$ = 125 MW	37,500
Total				39,125

Table 13: Settlement of a generator without an adder: forward reserve awarded, deployed.

The following table describes the settlement of the generator when the mechanism is applied, for the case where the forward reserve is deployed in real time.

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
Day-ahead energy	$\lambda PF_t \cdot pF_{gt}$	λPF_t =20 €/MWh	pF_{gt} = 0 MW	0

Day-ahead reserve	$\tilde{\lambda}RF_t \cdot rF_{gt}$	rF_{gt} = 65 €/MWh	rF_{gt} = 25 MW	1,625
Real-time energy	$\lambda PRT_t \cdot (pRT_{gt} - pF_{gt})$	λPRT_t = 1,529.2 €/MWh	$pRT_{gt} - pF_{gt}$ = 125 MW	191,150
Real-time reserve	$\tilde{\lambda}RRT_t \cdot (rRT_{gt} - rF_{gt})$	$\tilde{\lambda}RRT_t$ = 1,229.2 €/MWh	$rRT_{gt} - rF_{gt}$ = -25 MW	-30,730
Total				162,045

Table 14: Settlement of a generator with an adder: forward reserve awarded, deployed.

Note that, in responding with 100 MW above what was promised (in the form of reserve capacity in the day-ahead market), the generator supports the system in an imbalance interval of severe system stress. The proposed mechanism rewards the generator handsomely for this contribution.

7.5.2. Additional Capacity Decreases the Adder

The following example illustrates the self-correcting behavior of the adder when additional reserve capacity enters the system, and relieves the system from stress. Namely, consider the case of the ELIA scarcity pricing incentive report [11] for the most severe 15-minute interval in the year, namely November 29th, at 6pm – 6:15pm, where the 7.5-minute capacity is 1257.5 MW instead of 1090.0 MW. In this case, the adders are as follows:

- Adder for capacity deliverable in 7.5':

$$\tilde{\lambda}RRT_t^F = 45.1\ €/MWh$$

- Adder for capacity deliverable in 15':

$$\lambda RRT_t^S = 34.3\ €/MWh$$

- Adder for energy:

$$\tilde{\lambda}RRT_t^F = 45.1\ €/MWh$$

Note that, by introducing 167.5 MW of capacity into the system, the fast reserve capacity adder (and consequently the energy adder) immediately recedes.

7.5.3. Impact of the Mechanism on BRPs

From the point of view of a balancing responsible party, the only difference to the previous example is that the BRP is not offering reserve. Moreover, the BRP is encouraged to aid the system by moving in the right direction. Let us consider a concrete example where the BRP (which corresponds to a single generator) has a net positive injection of 100 MW in the day-ahead market, and that the generator has a PMax of 125 MW. Suppose, first, that the BRP remains in balance in real time (i.e. its physical production balances the short position that the BRP has taken in the day-ahead energy market). Then we have the following settlement.

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
Day-ahead energy	$\lambda PF_t \cdot pF_{gt}$	λPF_t =20 €/MWh	pF_{gt} = 100 MW	2000
Real-time energy	$\lambda PRT_t \cdot (pRT_{gt} - pF_{gt})$	λPRT_t = 1,529.2 €/MWh	$pRT_{gt} - pF_{gt}$ = 0 MW	0
Total				2000

Table 15: The settlement of a BRP which stays in balance in real time under scarce system conditions.

On the other hand, suppose that the BRP actually supports the system by taking a long position: the physical production of the BRP is equal to the PMax, 125 MW, which exceeds the short position that the BRP has taken in the day-ahead energy market. The settlement then proceeds as follows:

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
Day-ahead energy	$\lambda PF_t \cdot pF_{gt}$	λPF_t =20 €/MWh	pF_{gt} = 100 MW	2000
Real-time energy	$\lambda PRT_t \cdot (pRT_{gt} - pF_{gt})$	λPRT_t = 1,529.2 €/MWh	$pRT_{gt} - pF_{gt}$ = 25 MW	38,230
Total				40,230

Table 16: The settlement of a BRP which supports the system under scarce system conditions.

It is clear that the BRP stands to gain by supporting the system in taking a long position when the system is short on power. However, if the dispatch of the BRP is decentralized, then this could be a risky strategy, since the BRP would need to guess the direction (and magnitude) of the system imbalance.

7.5.4. Benefits for Loads

The following example considers a load which provides reserve services. This example illustrates how **the mechanism can be beneficial for loads**. We consider the same setup as before. The following table describes the settlement of the load when the mechanism is *not* applied.

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
Day-ahead energy	$-\lambda PF_t \cdot dF_{lt}$	λPF_t =20 €/MWh	dF_{lt} = 20 MW	-400
Day-ahead reserve	$\tilde{\lambda} RF_t \cdot rF_{lt}$	rF_{lt} = 65 €/MWh	rF_{lt} = 0 MW	0
Real-time energy	$-\lambda PF_t \cdot (dRT_{lt} - dF_{lt})$	λPRT_t = 300.0 €/MWh	$dRT_{lt} - dF_{lt}$ = 0 MW	0
Total				-400

Table 17: Settlement of a load without the mechanism.

The following table describes the settlement of the load when the mechanism is applied.

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
Day-ahead energy	$-\lambda PF_t \cdot dF_{lt}$	λPF_t =20 €/MWh	dF_{lt} = 20 MW	-400
Day-ahead reserve	$\tilde{\lambda} RF_t \cdot rF_{lt}$	$\tilde{\lambda} RF_t$ = 65 €/MWh	rF_{gt} = 0 MW	0
Real-time energy	$-\lambda PRT_t \cdot (dRT_{lt} - dF_{lt})$	λPRT_t = 1,529.2 €/MWh	$dRT_{lt} - dF_{lt}$ = 0 MW	0
Real-time reserve	$\tilde{\lambda} RRT_t \cdot (rRT_{lt} - rF_{lt})$	$\tilde{\lambda} RRT_t$ = 1,229.2 €/MWh	$rRT_{gt} - rF_{gt}$ = 20 MW	24,584
Total				24,184

Table 18: Settlement of a load with the mechanism.

In this example, the load has only bought power in the day-ahead energy market, and has not sold any reserve in the day-ahead market. The load consumers, in real time, the amount of power that it procured in the day-ahead market. In doing so, the load offers upward reserve capacity to the system: if needed, the load can back down by 20 MW. In the presence of a real-time market for reserve capacity, this generates a significant revenue for the load. This sends the signal to loads for installing equipment that can qualify them for providing secondary reserve capacity to the system.

7.5.5. Superior Remuneration of Faster Capacity

The following example illustrates that faster-responding capacity is better remunerated. Consider the same example as in section 7.5.4, but suppose that load is only eligible for tertiary reserve capacity. Then the settlement of the loads can be described as follows.

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
Day-ahead energy	$-\lambda PF_t \cdot dF_{lt}$	λPF_t =20 €/MWh	dF_{lt} = 20 MW	-400
Day-ahead reserve	$\lambda RF_t \cdot rF_{lt}$	λRF_t	rF_{gt} = 0 MW	0
Real-time energy	$-\lambda PRT_t \cdot (dRT_{lt} - dF_{lt})$	λPRT_t = 1,529.2 €/MWh	$dRT_{lt} - dF_{lt}$ = 0 MW	0
Real-time reserve	$\tilde{\lambda} RRT_t \cdot (rRT_{lt} - rF_{lt})$	λRRT_t = 34.4 €/MWh	$rRT_{gt} - rF_{gt}$ = 20 MW	686
Total				286

Table 19: Settlement of a load that can only offer tertiary capacity.

In this case, the profitability of the load has decreased substantially compared to the case of the previous table. This is due to the fact that load is only eligible for tertiary reserve.

7.6. Implementation Details

In this subsection, we discuss certain implementation details that relate to the implementation of scarcity adder computations.

7.6.1. ORDCs and their Relation to the Adders Computed by ERCOT and ELIA

The demand curves for operating reserves, which follow page 23 of [3], are expressed as follows:

$$V^{R,F}(r^F; r^{S,0}) = (VOLL - \widehat{MC}(\sum_g p_g)) \cdot (0.5 \cdot LOLP_{7.5}(r^F) + 0.5 \cdot LOLP_{15}(r^{S,0} + r^F))$$

$$V^{R,S}(r^S; r^{F,0}) = (VOLL - \widehat{MC}(\sum_g p_g)) \cdot 0.5 \cdot LOLP_{15}(r^S + r^{F,0})$$

These formulae, which describe the operating reserve demand curves, are also used for computing the adders themselves (this can be justified by assuming an interior solution for the system operator problem, meaning that the valuation of the system operator for reserve sets the price for reserve).

The operating reserve demand curves implied by these formulas which were used in the stochastic equilibrium analysis of the previous section are shown in Figure 19 in blue font. The black markers indicate the present hard reserve requirements of ELIA before activation.

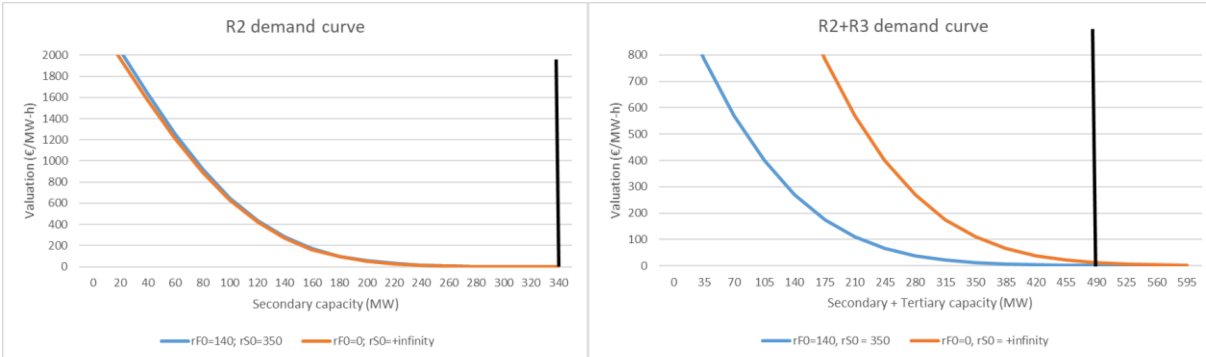


Figure 19: The operating reserve demand curves that were used for the stochastic equilibrium model for $r^{F,0}$ = 140 MW and $r^{S,0}$ = 350 MW (in blue font) and the demand curves when we use $r^{F,0}$ = 0 MW and $r^{S,0}$ = $+\infty$ MW (orange font).

The reserve prices that ERCOT uses (and that were also used in the ELIA report [11]) can be obtained from the above formulas if we set $r^{S,0} = +\infty$ and $r^{F,0} = 0$. This would reproduce the reserve prices that are reported in slide 34 of [6]. The corresponding demand functions are presented in Figure 19, and are given by the following formulae:

$$V^{R,F}(r^F; +\infty) = \left(VOLL - \widehat{MC}\left(\sum_g p_g\right)\right) \cdot \left(0.5 \cdot LOLP_{7.5}(r^F) + 0.5 \cdot LOLP_{15}(+\infty + r^F)\right)$$

$$= 0.5 \cdot \left(VOLL - \widehat{MC}\left(\sum_g p_g\right)\right) \cdot LOLP_{7.5}(r^F)$$

$$V^{R,S}(r^S; 0) = 0.5 \cdot (VOLL - \widehat{MC}(\sum_g p_g)) \cdot LOLP_{15}(r^S)$$

If the marginal unit is offering fast reserve, we then recover the energy adders reported in slide 34 of [6].

In slide 46, the ERCOT manual [6] explains that the energy price is adjusted according to the online reserve adder. In the formulae used in the ELIA report [11] (equation 1, page 11), we have proposed the same formula.

The equilibrium formula that we have developed for the present study illuminates why this is a reasonable approach. The key formulas of the equilibrium formulation are the following complementarity conditions of the generator:

$$0 \leq p_{g,\omega}^{RT} \perp C_g - \lambda_\omega^{RT} + \alpha_{g,\omega}^{G,RT,+} - \alpha_{g,\omega}^{G,RT,-} \geq 0, g \in G, \omega \in \Omega$$

$$0 \leq r_{g,\omega}^{F,RT} \perp \alpha_{g,\omega}^{G,RT,+} + \beta_{g,\omega}^{G,F,RT} - \lambda_\omega^{R,F,RT} - \lambda_\omega^{R,S,RT} \geq 0, g \in G, \omega \in \Omega$$

$$0 \leq r_{g,\omega}^{S,RT} \perp \alpha_{g,\omega}^{G,RT,+} + \beta_{g,\omega}^{G,S,RT} - \lambda_\omega^{R,S,RT} \geq 0, g \in G, \omega \in \Omega$$

Ideally, the real-time price is the solution of the stochastic equilibrium problem. The computation of this value is not practical under real-time operations. By making a number of assumptions (we explain these below), we can arrive at a practical calculation that is suitable for real-time operations.

Consider an interior solution for $p_{g,\omega}^{RT}$, i.e. $p_{g,\omega}^{RT} > 0$. Then $\alpha_{g,\omega}^{G,RT,-} = 0$, therefore

$$\lambda_{\omega}^{RT} - C_g = \alpha_{g,\omega}^{G,RT,+}.$$

The coefficient $\alpha_{g,\omega}^{G,RT,+}$ is effectively the adder to the marginal cost of the marginal unit, which determines the real-time energy price. This is the key quantity that we are interested in. The question is how this adder behaves as a function of the fast reserve capacity price $\lambda_{\omega}^{R,F,RT}$ and the slow reserve capacity price $\lambda_{\omega}^{R,S,RT}$.

In order to analyze this question further, let us assume that adders result from limited capacity, as opposed to limited ramping, which is to say that $\beta_{g,\omega}^{G,F,RT} = \beta_{g,\omega}^{G,S,RT} = 0$. Typically, the marginal unit is either offering fast reserve, or slow reserve, but not both. In the former case, and under the previous assumption, the real-time energy price is the marginal cost of the marginal unit plus the price of fast reserve capacity (in order to preclude arbitrage), while in the latter case it is the marginal cost of the marginal unit plus the price of slow reserve capacity (in order to preclude arbitrage). As a practical recommendation, we propose to set the energy price equal to the system lambda plus the fast reserve capacity price, in order to allow the energy price to become non-zero when *either* the $LOLP_{7.5}$ *or* the $LOLP_{15}$ becomes high. In this way, we will pick up the shortage in the energy price when R2 or R3 capacity is scarce. The implicit assumption in doing so is that the marginal unit is always a unit that is offering fast reserve capacity on the margin. This is already the approach that has been adopted in the ELIA study [11].

In order to illustrate our approach, we present the results for the first scenario of the first time step of our case study. We note that in this case, the real-time energy price is 57.19 €/MWh, which is the marginal cost of the most expensive unit that is producing strictly below its technical maximum plus the price of slow reserve capacity. This unit is in fact producing slow reserve as well as fast reserve. Its fast ramp constraint is binding, therefore there is no reason to expect that the real-time price should be driven by the fast reserve price. Instead, the slow reserve ramp rate constraint is non-binding, which explains why the real-time price is the system lambda plus the slow reserve capacity price. Since the resolution of this complementarity system is too complex for real-time operations, we propose instead as a workable approximation to augment the system lambda (the marginal cost of the marginal unit) by $\tilde{\lambda}_{\omega}^{R,F,RT} = \lambda_{\omega}^{R,F,RT} + \lambda_{\omega}^{R,S,RT}$. In the case of our example, this would result in a real-time price of 60.31 €/MWh, instead of 57.19 €/MWh.

7.6.2. Reference Horizon

The premise of deriving an ORDC based on loss of load probability is a two-stage stochastic program, as described in the appendix of [3]. The first stage of the program is the dispatch decision, then uncertainty is revealed in terms of capacity shortfall, and in the second stage the system reacts by balancing the shortfall through adjustments in generator production or load shedding. Throughout this process, it is assumed that the commitment of units remains constant. The system presumably 'resets' at the end of this two-stage program, in the sense that what happens beyond this horizon is independent of the decisions that take place within this horizon.

The question that we are interested in addressing in this section is how the horizon of this two-stage program (which we denote as Δ_k) relates to the activation time of reserves[44]. Consider a reference unit of

[44] These issues are also discussed in pages 17, 18 of [16]. For example, in page 18: "PJM proposes to base its reforms on a 30-minute look ahead for uncertainty for the synchronized and primary reserve requirements and a 60-minute look ahead for the 30-minute reserve requirement".

time (e.g. one second) and suppose that the total disturbance that the system experiences within the horizon is a sum of fully correlated disturbances that take place in the reference time unit (i.e. a sum of one-second perfectly correlated disturbances). We use Δ_0 to denote the duration of the reference time unit. Assuming that the reserve capacity that a system can make available is only limited by ramp rates, we would have that the loss of load probability over the horizon Δ_k is:

$$\mathbb{P}[\Delta_k \geq \mathrm{R}_k] = \mathbb{P}[k \cdot \Delta_0 \geq k \cdot \mathrm{R}_0] = \mathbb{P}[\Delta_0 \geq \mathrm{R}_0]$$

Thus, under these assumptions, the loss of load probability is invariant with respect to the length of the horizon, and depends purely on the ramp rate that is available and the distribution of Δ_0.

In practice, we may have deviations from these idealized assumptions:

1. $\Delta_k \neq k \cdot \Delta_0$ if Δ_k is not the sum of fully correlated increments.

2. $\mathrm{R}_k \neq k \cdot \mathrm{R}_0$ if the system has limited reserve capacity, i.e. if reserve capacity is limited by generator technical maximum instead of ramp rates.

In the range of a few seconds, the system behaves more in line with the assumptions that $\Delta_k = k \cdot \Delta_0$ and $\mathrm{R}_k = k \cdot \mathrm{R}_0$. Thus, the LOLP does not depend on the choice of k. In the range of a few minutes or an hour, this is a decreasingly accurate approximation of reality, and therefore the LOLP calculation does not remain invariant with respect to k.

When the LOLP is dependent on Δ_k, the first question to answer is what an appropriate choice of Δ_k should be. One hour seems to be the maximum acceptable length, since for either US or European systems, unit commitment decisions are updated every hour. One hour is the horizon of choice for the ORDC calculations in ERCOT. In the ELIA report [11], we have assumed that Δ_k equals fifteen minutes. This is justified by the assumption that the system 'resets' at the beginning of every balancing interval.

The approach that is adopted in ERCOT differs slightly from what has been tested for Belgium. The differences are illustrated in Figure 20. In ERCOT (left), the adder is recomputed every 5 minutes (with every new run of SCED[45]), the look-ahead horizon is 1 hour, and the slow reserves are assumed to be fully available within 30 minutes. In Belgium (right), the adder is computed every 15 minutes (at every imbalance interval), the look-ahead is 15 minutes, and the slow reserves are assumed to be available within 7.5 minutes. Note that this assumption is in line with the moment at which tertiary reserve should be **mobilized** (tertiary reserve is supposed to unload secondary reserve, i.e. it should be mobilized within 7.5 minutes, at which point secondary reserve is fully deployed), but is optimistic in terms of the time by which it should be **fully activated** (since the tertiary reserve is normally fully responsive in 15 minutes, not 7.5 minutes).

[45] Note that the fact that SCED is re-run every five minutes means that the amount of reserve is updated frequently, and accurately tracks the level of stress in the system in the LOLP calculations. With less frequent calculations of the adder (e.g. 15 minutes) it is important to approximate the amount of reserve throughout the imbalance interval. We do so in our calculations by computing the amount of reserve in the system as the reserve available *after* activation for covering imbalances.

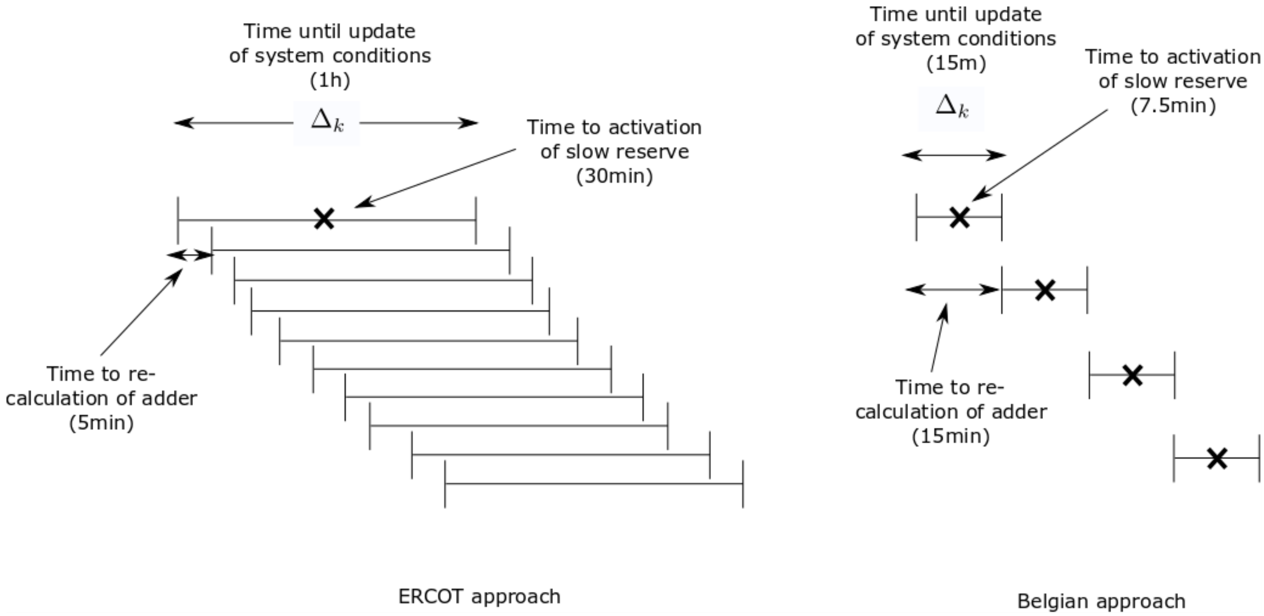


Figure 20: The rolling calculation of scarcity adders.

In what follows, we perform some sensitivity analyses in order to better understand how the choice of horizon affects the adder. These sensitivities use a 30-minute window for Δ_k, and assume that the slow reserve is fully available in 15 minutes, which follows the strict specifications of the product. In doing so, we consider the case where the imbalance increments are perfectly correlated, and the case where they are perfectly independent (note that previous work [18] has shown that imbalance increments are better approximated as being perfectly correlated, whereas ERCOT assumes perfectly independent imbalance increments, which is more justifiable when the increments are longer – 30 minutes in the case of ERCOT).

Note that a larger horizon implies more uncertainty in a 30-minute horizon than in a 15-minute horizon, and therefore the loss of load probability is higher for a *given* amount of reserve capacity in a 30-minute horizon than in a 15-minute horizon. However, also note that the 30-minute time window implies that more reserve capacity will be available. In particular, if the available reserve capacity is only limited by ramp rate, with double the time window we can count on double the capacity. The sensitivity analyses that have been conducted explore the impact of the following factors on the adders (all formulas are based on a 30-minute look-ahead):

- Computation of the adder before / after the clearance of the imbalance
- 30-minute imbalances are the sum of two perfectly correlated / independent 15-minute imbalance increments
- The 30-minute capacity is equal / twice as much the 15-minute capacity, corresponding to a purely energy / ramp-rate constrained system

In the following table we show the results for November 29th, 2018. The calculations have similarly been performed for all tight days of the system (29/11, 6/11, 12/12, 10/12) but are only reported for November 29th.

	R3 (all ramp constrained)	**R2 (all ramp constrained)**	**R3 (all capacity constrained)**	**R2 (all capacity constrained)**

Default	34.3 / 0 / 1	45.1 / 0 / 1	N/A	N/A
Correlated imbalances + reserve after activation	0 / 0 / 0	1.5 / 0 / 1	171.5 / 3.8 / 4	173.0 / 3.8 / 4
Independent imbalances + reserve after activation	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
Correlated imbalances + reserve before activation	0 / 0 / 0	0 / 0 / 0	0.1 / 0.1 / 0	171.5 / 3.8 / 4
Independent imbalances + reserve before activation[46]	0 / 0 / 0	0 / 0 / 1	0 / 0 / 0	0 / 0 / 0

Table 20: The (a) maximum price, (b) second highest price, and (c) number of occurrences of a price of at least 1 €/MW-h on November 29, 2018 for R2 and R3 reserve prices.

We observe that the approach of expanding the horizon to half an hour increases the elasticity of the demand function if we assume perfectly correlated imbalances. The notable results are highlighted in green font. These non-zero adders occur when (i) we compute the adders on the leftover capacity after balancing, and when (ii) the system is purely capacity constrained. The maximum spike is higher than the one that results from the sensitivity case of ELIA [11], however the spike computed in [11] only lasts for one period, whereas the price adders reported in the green font of the previous table occur for 4 time periods, which indicates a higher elasticity of the demand curve, which is a desirable feature.

7.6.3. What Counts as Fast Reserve

In its recent study [11], ELIA performed a sensitivity analysis, depending on whether R3 non-CIPU should be considered as part of the tertiary reserve capacity or not. The ELIA report found a large sensitivity of the adder on this assumption. This finding is not surprising, in light of the demand curves of Figure 19. For example, in the November 29 incident of the ELIA report ([11], page 45) the available 7.5-minute reserve capacity in the reference case is 536 MW, whereas it amounts to 366 MW in the sensitivity (before clearing imbalances). This implies a difference of 170 MW in available secondary reserve, which is exactly half the secondary reserve of the system.

In its analysis, ELIA considered the following resources as contributing towards the various types of reserve:

[46] Note that here the leftover reserve capacity is equal to the original reserve capacity minus the 15-minute observed imbalance. This contradicts the assumption about perfectly correlated imbalances, however the alternative would be to presume a 30-minute imbalance that was not actually observed in practice. We choose the first option.

	R2	CIPU coordinable	ICH	R3 CIPU standard	R3 CIPU flexible	R3 non-CIPU standard	R3 non-CIPU flexible	Hydro
R3 capacity		√	√	√	√	√	√	√
R2 capacity	√	0.5		0.5	0.5	0.5	0.5	0.5
R2 capacity (sensitivity)	√	0.5		0.5	0.5			0.5

Table 21: Resources included in the ELIA study as contributing towards R2 and R3.

In 2019, the following changes are considered:

- ICH is discontinued, so it is no longer a separate entry in the table
- R2 is included in R3 capacity
- 50 MW of inter-TSO reserve are added to the firm capacity of the system

The resources that contribute to each type of capacity for the 2019 scarcity pricing incentive are thus summarized in the following table.

	R2	CIPU coordinable	R3 CIPU standard	R3 CIPU flexible	R3 non-CIPU standard	R3 non-CIPU flexible	Hydro	Inter-TSO (50 MW)
R3 capacity	√	√	√	√	√	√	√	√
R2 capacity	√	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table 22: Resources considered in 2019 as contributing towards R2 and R3.

8. Conclusions

Our analysis underscores the importance of establishing a well-functioning real-time market as a necessary condition for rewarding reserve services adequately in a regime of power system operations that requires significant levels of flexibility. The most important measure in this direction is to introduce a market for real-time reserve capacity. Concretely, we propose the introduction of a scarcity adder for reserve capacity which is payable to standby real-time reserve capacity, and which also uplifts the Belgian imbalance price. The calculation of this adder requires the so-called Available Reserve Capacity (ARC), which is measured in real time by the system operator, and which is used for computing the loss of load probability that is required for computing the scarcity price. The Belgian system operator has already published a report [11] where the ARC is used for computing scarcity prices. These adders should affect not only activated energy bids, but also free bids and other standby capacity which is available in real time, even if not activated.

The implementation of a real-time market for reserve capacity in Belgium, as described throughout the report, is the least ground-breaking measure along the chain of evolutions that are indicated in our analysis. Nevertheless, as indicated by our analysis, it delivers the greatest benefits in a market with risk-neutral agents.

Coincidently, our analysis indicates that, in a risk neutral setting, the more disruptive measures of introducing virtual trading and co-optimizing the trading of reserve and energy in the day-ahead time frame also have a lesser influence on prices and the profitability of flexible resources. A transition to explicit virtual trading would require a radical overturn of the European view of real-time markets as a service, which remains endemic in European policy debates, and is therefore likely to be a time-consuming effort. A future transition to co-optimization of energy and reserves is also likely to be challenging, as it raises computational challenges related to the uniform pricing approach that is adopted in European day-ahead markets based on paradoxically rejected bids, and requires the engagement of various stakeholders, including transmission system operators and market operators. Our conclusion, therefore, is that the most realistic and effective first step for the implementation of scarcity pricing in Belgium is the introduction of a real-time market for reserve capacity.

We note that our observations about the importance of virtual trading and simultaneous clearing of energy and reserve are only valid in a risk-neutral setting. As we have demonstrated through the analysis of small illustrative examples, the impact of virtual trading is likely to have a material impact on price convergence in the absence of risk-neutral agents. In conducting our large-scale numerical simulation of the Belgian market, we are inevitably limited to a risk-neutral model in order to be able to cast the problem equivalently as a stochastic program. In future work, we are interested in extending this large-scale numerical simulation to the case without risk-neutral agents. At present, the application of the resulting model to large-scale simulations is only possible for the risk-neutral case. We are currently exploring numerical methods for scaling the risk-averse version of the model to realistic instances.

Acknowledgements

The authors appreciate helpful discussions with William Hogan, Alain Marien, José Gonzalez Pastor, Steven Mertens, Jacques Gheury, and Andreas Tirez, as well as with attendants at the 2019 Workshop on Electricity Systems of the Future at the Isaac Newton Institute in Cambridge University, the 2019 NTNU PhD winter school in energy systems and markets, the 2018 INFORMS Annual Meeting, and the 2018 Energy Day at CORE.

References

[1] A. Papavasiliou and Y. Smeers, "Remuneration of Flexibility under Conditions of Scarcity: A Case Study of Belgium," *The Energy Journal,* vol. 38, no. 6, pp. 105-135, 2017.

[2] W. W. Hogan, "On an ‘energy only’ electricity market design for resource adequacy," Center for Business and Government, JFK School of Government, Harvard University, 2005.

[3] W. W. Hogan, "Electricity scarcity pricing through operating reserves," *Economics of Energy and Environmental Policy,* vol. 2, no. 2, p. 65–86, 2013.

[4] ERCOT, "ERCOT Market Education, Basic Training Program, Module 6: Real-Time Operations".

[5] ERCOT, "ERCOT Market Education, Basic Training Program, Module 2: Day-Ahead Operations".

[6] ERCOT, "ERCOT Market Training: Purpose of ORDC, Methodology for Implementing ORDC, Settlement Impacts for ORDC".

[7] ELIA, "Access Responsible Party Contract".

[8] ELIA, "Ancillary services: Volumes and prices".

[9] ELIA, "ELIA SystemRegles de Fonctionnement du Marché Relatif à la Compensation des desequilibres quart-horaires - Entrée en vigueur au 01/01/2018".

[10] ELIA, "General Framework for Tertiary Control by Generators".

[11] ELIA, "Study report on scarcity pricing in the context of the 2018 discretionary incentives," 2018.

[12] ELIA, "The CIPU contract: a set framework for taking part in the high-voltage grid management".

[13] J.-M. Glachant and M. Saguan, "An Institutional Frame to Compare Alternative Market Designs in EU Electricity Balancing," 2007.

[14] M. Madani and M. Van Vyve, "Computationally efficient MIP formulation and algorithms for European day-ahead electricity market auctions," *European Journal of Operational Research,* vol. 242, no. 2, pp. 580-593, 2015.

[15] P. Economics, "2017 State of the Market Report for the ERCOT Electricity Markets," 2018.

[16] S. P. William Hogan, "PJM Reserve Markets: Operating Reserve Demand Curve Enhancements," 2019.

[17] A. Flamm and D. Scott, "Electricity Balancing Significant Code Review - Final Policy Decision," Office of Gas and Electricity Markets, 2014.

[18] A. Papavasiliou, Y. Smeers and G. Bertrand, "An Extended Analysis on the Remuneration of Capacity under Scarcity Conditions," vol. 7, no. 2, 2018.

[19] J. E. Parsons, C. Colbert, J. Larrieu, T. Martin and E. Mastrangelo, "http://www.mit.edu/~jparsons/publications/20150300_Financial_Arbitrage_and_Efficient_Dispatch.pdf," MIT Center for Energy and Environmental Policy Research, 2015.

[20] W. W. Hogan, "Virtual bidding and electricity market design," *The Electricity Journal,* vol. 29, no. 5, pp. 33-47, 2016.

[21] K. De Vos, N. Stevens, O. Devolder, A. Papavasiliou, B. Hebb and M.-D. Johan, "Dynamic Dimensioning Approach for Operating Reserves: Proof of Concept in Belgium," *Energy Policy,* vol. 124, pp. 272-285, 2019.

[22] S. S. Oren, "Design of Ancillary Services Markets," in *Hawaii International Conference on System Sciences*, 2001.

[23] "Potential cross-border balancing cooperation between the Belgian, Dutch and German electricity transmission system operators," 50Hertz Transmission GmbH, Amprion GmbH, Elia System Operator NV, TenneT TSO B.V., TenneT TSO GmbH, TransnetBW GmbH, 2014.

[24] G. Oggioni, R. Dominguez and Y. Smeers, "Reserve procurement and flexibility services in power systems with high renewable capacity: effects of integration on different market designs," 2019.

[25] E. SPOT, GME, HEnEx, N. Pool, OMIE, OPCOM, OTE and TGE, "EUPHEMIA Public Description - PCR Market Coupling Algorithm," 2018.

[26] Federal Energy Regulatory Commission, "Uplift in RTO and ISO Markets," 2014.

[27] I. Aravena, "Analysis of renewable energy integration in transmission-constrained electricity markets using parallel computing," PhD thesis, Université catholique de Louvain, 2018.

[28] D. Devogelaer, "Belgian blackouts calculation: a quantitative evaluation of power failures in Belgium," Belgian Federal Planning Bureau, 2014.

[29] H. Scarf, "The allocation of resources in the presence of indivisibilities," *Journal of Economic Perspectives,* vol. 8, no. 4, pp. 111-128, 1994.

Appendix: Glossary

The following glossary provides short definitions of specific terms that are used in the Texas and Belgium market, and points out correspondences whenever relevant.

Adjustment period (Texas): a process following the day-ahead market and before reliability unit commitment in Texas day-ahead operations, where schedules of individual generators are adjusted in order to allow for self-commitment and outages.

Automatic frequency restoration reserve / aFRR (Belgium): Synonym, and most recent terminology for, secondary reserve.

Available regulation capacity / ARC (Belgium): A function operated by ELIA which computes the amount of capacity which can be made available for responding in the upward and downward direction within 15 minutes.

Balancing responsible party / BRP (Belgium): Entity in the Belgian market which is responsible for arriving to real time with a forward financial position that exactly matches its net physical position.

Balancing service provider / BSP (Belgium): Entity in the Belgian market that offers secondary and/or tertiary reserves.

Base point deviations (Texas): These are deviations of resources during their real-time dispatch from the energy and ancillary services set-points that have been instructed by the system operator.

Continuous intraday market (Belgium): An energy market that operates after the day-ahead auction and until 45 minutes before real time, with a continuous matching of bids on a bilateral first-come-first-serve basis.

Current operating plan / COP (Texas): The hourly on/off, technical minimum, technical maximum, and ancillary service obligation schedule of individual generators in the Texas day-ahead market. This is the analog of nominations in the Belgian market.

Coordination of the Injection of Production Units / CIPU contract (Belgium): A legacy classification of conventional units (as opposed to newer renewable or demand-side resources) in the Belgian system, along with an associated set of rules that govern the operation of these units.

Day-ahead reliability unit commitment / DRUC (Texas): The day-ahead process that includes TSA and RUC.

EUPHEMIA (Belgium): The algorithm that is used for clearing the European day-ahead power exchange.

Energy bids (Texas): Demand-side bids in the ERCOT day-ahead market that are submitted by QSEs for buying energy.

Energy-only offers (Texas): Supply-side bids in the ERCOT day-ahead market that are submitted by QSEs for selling energy.

Free bids (Belgium): Bids for upward and downward regulation which are submitted to the Belgian real-time market by resources that have not pre-committed their capacity as reserve.

Frequency control reserve / FCR (Belgium): Synonym, and most recent terminology for, primary reserve.

Independent system operator / ISO (Texas): ERCOT, the entity which operates the electric power system and electricity market of Texas, including the day-ahead and real-time reserve and energy markets.

Intraday price coupling of regions / IDPCR (Belgium): A continuous auction that trades energy after the day-ahead market and before real time.

Load frequency control / LFC (Texas): Automatic control that is sent every 4 seconds to resources that are providing regulation in the Texas market, triggered by frequency deviations. This is the analog of primary reserve in the Belgian system.

Locational marginal prices / LMP (Texas): Marginal value of locational power balance constraint in the SCED.

Manual frequency restoration reserve / mFRR (Belgium): Synonym, and most recent terminology for, tertiary reserve.

Minimum contingency level / MCL (Texas): The minimum amount of reserve capacity, below which the ORDC adder produces a real-time price equal to VOLL.

Market information system / MIS (Texas): The information technology platform that is used in the ERCOT market in order to map postings related to market operations.

Net regulation volume / NRV (Belgium): The energy that ELIA dispatches in order to cope with system imbalance.

Nomination (Belgium): The day-ahead procedure whereby the set-point, technical maximum and quantity of offered reserve of individual resources are declared by the owners to the system operator.

Paradoxically rejected bids / PRB (Belgium): Bids in the day-ahead power exchange which may be rejected, even if activating them would result in a profit for these resources.

Pool (Texas): The organization of the day-ahead market in Texas whereby the non-convex costs and operating constraints of generators are represented explicitly in the day-ahead market bids.

Portfolio bidding (Belgium): The practice whereby market participants enter the day-ahead power exchange with a bid representing a portfolio of resources, as opposed to an individual generator or load.

Power exchange (Belgium): EPEX Spot, the entity which operates the Belgian day-ahead energy market, and the actual operation of trading energy in the day-ahead time frame.

Primary reserve /R1 (Belgium): Reserve in the Belgian market that needs to react within 3 seconds. This is the analog of regulation in the Texas market.

Proactive balancing (Europe): The notion that BRPs freeze their schedules hours in advance of real time, with the TSO taking over balancing of the system from that point onwards.

Qualified scheduling entities / QSE (Texas): Market entities that manage generation resources and load resources.

Reactive balancing (Europe): The notion that BRPs should be responsible for balancing their perimeter right up to real time operations, with the TSO providing advance indicators that can help BRPs balance their perimeter, and with the TSO only handling any remaining imbalances.

Reliability unit commitment (Texas): A process which is executed in the day-ahead time frame after the Texas day-ahead market in order to commit additional units beyond those committed by the day-ahead market, in case the ISO assesses that this is needed in order to ensure reliable operations.

Reserve price adders (Texas): The adder computed by the ORDC methodology. This corresponds to the Belgian scarcity adder.

Real-time online reserve price adder / RTORPA (Texas): The amount of reserve capacity that can be made available in a horizon of 30 minutes, as measured every five minutes by the results of a SCED run.

Real-time offline reserve price adder / RTOFFPA (Texas): The amount of reserve capacity that can be made available in a horizon of 60 minutes, as measured every five minutes by the results of a SCED run.

Real-time reserve price for online reserve / RTRSVPOR (Texas): The average of RTORPA over a 15-minute interval, used for settlement purposes.

Real-time reserve price for offline reserve / RTRSVPOFF (Texas): The average of RTRSVPOFF over a 15-minute interval, used for settlement purposes.

Real-time settlement point prices (Texas): The result of combining locational marginal prices with reserve price adders, which is used for paying activated reserves.

Reliability must run / RMR units (Texas): Resources that are required to run in real time for reliability reasons, independently of the outcome of the day-ahead market.

Reservation (Belgium): The procurement of reserve capacity by the TSO in auctions that take place before the day-ahead energy market.

Responsive reserve service / RRS (Texas): Reserve that needs to be made available within 30 minutes in the Texas market. This is the analog of tertiary reserve in the Belgian market, in the sense that it is the slowest type of operating reserve.

R3 flexible (Belgium): A type of tertiary reserve product offered in the Belgian market which has less stringent delivery conditions than standard tertiary reserve.

R3 standard (Belgium): The reference tertiary reserve product that is offered in the Belgian market.

Secondary reserve / R2 (Belgium): Reserves that need to be activated within 7.5 minutes in the Belgian market. This is similar to responsive reserves in the Texas market, in these sense that this is the fastest operating reserve.

Security constrained economic dispatch / SCED (Texas): A real-time dispatch model that is run in the Texas market every five minutes.

Self-commitment (Texas): The decision to commit a unit independently of the result of the day-ahead market. Self-commitment typically takes place in the adjustment period after the day-ahead market, and resources that are self-committed are not guaranteed a make-whole payment for their fixed and startup costs.

Shift factor (Texas): Output produced by the network security analysis function of ERCOT, which is added to the system lambda in order to determine the LMP.

Strategic reserve (Belgium): An emergency measure used in Belgium for keeping units that are intended to be mothballed as available backup capacity in order to overcome adequacy issues during winter months.

System imbalance: The discrepancy between injections and offtakes of power which produce deviations from reference frequency.

Tertiary reserve / R3 (Belgium): Reserves that need to be activated within 15 minutes in the Belgian market. This is similar to the Texas non-spinning reserve, in the sense that it is the slowest operating reserve.

Three-part supply offers (Texas): Supply-side bids in the ERCOT day-ahead market that are submitted by QSEs for selling energy and are associated to individual generators.

Transmission security analysis / TSA (Texas): Part of the DRUC process which generates input for RUC by screening contingencies.

Transmission service provider / TSP (Texas): TSPs are responsible for operating and monitoring transmission resources (lines, transformers, buses).

Transmission system operator / TSO (Belgium): ELIA, the entity which operates the Belgian electric power system.

Two-settlement system (Texas): An accounting system for treating day-ahead financial transactions followed by physical real-time injections/withdrawals of power. The two-settlement system can be viewed in two identical ways. (i) Agents buy out their *entire* financial position at the real-time price, and are also paid the real-time price for their entire physical production/withdrawal. (ii) Equivalently, agents are paid the real-time price for the difference between their physical injection/withdrawal and their position in the forward day-ahead market.