Interactions between the design of short-term electricity markets in the CWE region and power system flexibility

Tom Brijs\textsuperscript{a,b,c,⇑}, Cedric De Jonghe\textsuperscript{a,b}, Benjamin F. Hobbs\textsuperscript{d}, Ronnie Belmans\textsuperscript{a,b}

\textsuperscript{a} Department of Electrical Engineering, University of Leuven (KU Leuven), Heverlee, Belgium
\textsuperscript{b} EnergyVille Research Institute, Genk, Belgium
\textsuperscript{c} Department of Civil Engineering, The Johns Hopkins University, Baltimore, MD, USA
\textsuperscript{d} Department of Environmental Health and Engineering, and the Environment, Energy, Sustainability & Health Institute, The Johns Hopkins University, Baltimore, MD, USA

HIGHLIGHTS

- Short-term electricity markets include day-ahead, intra-day, and real-time markets.
- Short-term markets are important tools to deal with variability in the system.
- A good understanding of their design is important for analyses on flexibility.
- The market design determines the rules by which flexibility providers must play.

ABSTRACT

Short-term electricity markets are generally defined as markets that take place from the day-ahead stage until physical generation and consumption. These markets include day-ahead, intra-day, and real-time balancing markets. In Europe, the first two are managed by power exchanges, while the third consists of reserve procurement and imbalance settlement and is operated by the local transmission system operator. Short-term markets are important tools to deal with net demand variability in the system, in which the need for flexibility is expressed and its provision is valorized. Due to the ongoing integration of variable renewables in the generation mix, the system’s variability is increasing as a result of the limited controllability and predictability of those resources. As such, these markets become increasingly important. The contribution of this article is a comprehensive up-to-date discussion of the key design parameters and functioning of all three short-term markets, and their impact on the demand for and supply of flexibility. An understanding of the design and its implications is useful to policy-makers who are considering changes to facilitate the integration, availability, or valorization of flexibility, while also contributing to the decision-making of flexibility investors and operators. The geographical scope is the Central Western European region, including the Belgian, French, German, and Dutch market zones.

1. Introduction

Electricity markets differ from other commodity markets because electric energy is a real-time (RT) product. Therefore, it is necessary to maintain a balance between generation and consumption at all times. Historically, electricity was generated by large centralized thermal and hydro power plants to match demand, with the help of fuel storage and pumped-hydro storage plants to compensate for the variability of consumption and the partial inflexibility of conventional power plants. The ongoing transition towards variable renewable energy sources (RES), i.e., wind turbines and photovoltaic systems, challenges the reliable operation of the power system. Their limited controllability and predictability results in an increasing need for flexibility, i.e., the ability to provide upward and downward power adjustments to compensate for temporary imbalances between generation and consumption [1,2]. At the same time, the flexibility offered by the generation side is threatened by closure of conventional power plants that are currently experiencing decreasing profitability due to lower electricity prices and a limited number of operating hours. The former is due to the close-to-zero marginal cost of RES, while...
the latter can be attributed to the merit-order effect of RES [3,4]. Further, a paradigm shift is taking place from a situation where generation was dispatched to follow inflexible demand to a situation in which flexibility is provided by both generation and consumers. However, there will be a need for electricity storage as well to fill the remaining gap, and for the further development of interconnection capacity and integration of adjacent markets to access flexible resources in neighboring regions [5].

1.1. Scope and motivation

In Europe, market players self-schedule their generation, consumption, and storage assets as a result of trading, which starts years prior to delivery and continues almost until RT. This is accomplished by a series of sequential markets, of which the earliest are the so-called unstandardized forward and standardized future markets. These markets usually continue until one day before delivery, when the power exchange holds its centrally organized day-ahead (DA) market. After clearing of the DA market, intra-day (ID) trading is possible until close-to-RT. After gate closure of the ID market, the transmission system operator (TSO) is responsible to keep the system balanced. To maintain this balance, the TSO contracts and activates reserve capacity from balance service providers (BSPs) at the procurement side, and settles imbalance positions with balance responsible parties (BRPs) at the settlement side of the balancing market [6,7]. We focus on the Central Western European (CWE) region, which, consistent with common definitions (e.g., [8–10]), includes the Belgian, French, German, and Dutch market zones.

Due to techno-economic constraints of power plants, and limited ability to foresee renewable generation, outages of power plants and grid elements, and load behavior, it is in short-term electricity markets where the need for flexibility is most apparent, and providers of flexibility are financially rewarded. These markets are generally defined as those taking place from the DA stage until physical generation and consumption in RT, i.e., including DA, ID, and RT markets. Currently these markets are becoming more important due to increasing levels of variability in the system, resulting from the ongoing integration of variable RES. A good understanding of the current design and functioning of these markets, as well as of possible future developments, is a foundation for analyzing the need for and provision of flexibility.

1.2. Context: the theory of short-term energy pricing and rewarding of flexibility

Despite the attention that the literature pays to the need for flexibility, discrete chronologic intervals (e.g., hours), ramp rate limits, and, if desired, convex approximations of commitment constraints. This linear program yields a primal solution that is not only the social least-cost solution, but also represents a market equilibrium among price-taking generators who compete to supply a fixed demand in each hour [11]. The model also yields energy prices (the Lagrange multipliers associated with the hourly energy balances) that support the optimal solution. “Support” refers to the property that each generator’s capacity and operating decisions are profit maximizing for that generator, given the prices; a generator cannot earn more profit by deviating in a feasible way from the primal solution. The only revenue earned by generators is from energy sales, showing that in theory only energy prices are needed to support optimal schedules, even when there is highly variable net demand with steep ramps. For instance, negative and positive price spikes associated with steep net load ramps will theoretically send the correct signals for operation and ultimately investment in this simplified world. Flexible generators earn more revenue and can justify their more expensive capacity costs because they can turn down and up to avoid negative price spikes and grab positive price spikes, respectively.

This result generalizes to a world of uncertainty with risk-neutral (expected profit maximizing) generators. If the generation capacity expansion model is formulated as a linear stochastic program with random (e.g., Markovian) net demand and RES generation, then it can be shown that the resulting stochastic prices support the optimal operations and capacity decisions by generators.

Therefore, justifications for the creation of flexibility products or paying separately for flexibility in addition to energy commodity prices require a rationale based on market failures. Such failures could include (1) lumpy/nonconvex investment and unit commitment decisions that are not appropriately reflected in prices; (2) price caps and floors that suppress spikes; (3) dispatch intervals (e.g., one hour) that are too long and average out spikes so that flexibility is not rewarded; or (4) failures in the investment market, e.g., political or highly risk-averse decision making. The extent to which these failures provide distorted incentives for supplying flexibility has not been quantified and is an open research topic.

In this article, we review the design and performance of existing short-term markets in the CWE region, focusing on how they reward flexibility, and where they might be reformed to provide improved incentives for providing that flexibility.

1.3. Literature review and contributions

1.3.1. Literature review on the design of short-term markets

In this section, we begin our review of how short-term markets in the CWE region incentivize flexibility by providing a comprehensive overview of the literature on the design of short-term markets in Europe. We divide the literature into three groups in the following three paragraphs. The focus of each reference is briefly highlighted, with Table 1 classifying those that consider particular market zones and short-term markets according to these two dimensions.

The first group of works discuss the short-term market design for individual market zones. Ref. [7] analyzes the occurrence of negative balancing prices in Belgium (BE), while discussing the functioning of its balancing market. Ref. [12] reviews changes to the balancing market in France (FR) to enhance transparency and competition. Ref. [13] focuses on the DA market in Germany (DE) and investigates the impact of offshore wind, and in [14] a brief overview of the functioning of the different German markets is
given. Ref. [15] analyzes the implementation of a German discrete ID auction, while discussing the functioning of both auction-based and continuous ID trading. Refs. [16–18] focus on a variety of design parameters of the German balancing market’s procurement side. Ref. [19] analyzes the German balancing mechanism while focusing on the interplay between imbalance pricing and congestion. Ref. [20] discusses three interactions between variable RES and balancing markets, and [21] studies the impact of the balancing market design on the participation of flexible consumption, both with a focus on Germany. Ref. [22] analyzes incentives to behave strategically in the German balancing market, while [23] discusses its design together with historical data. Ref. [24] reviews a general set of principles of the different markets in the Netherlands (NL), but only considers limited design parameters, while in [25] the Dutch balancing market is discussed. Ref. [26] focuses on the procurement side of the Dutch balancing market and the participation of electric vehicles. Turning to discussions of individual market zones outside the CWE region, Ref. [27] analyzes the participation of RES in the Spanish ID market, while providing an overview of the Spanish short-term markets. Ref. [28] reviews the Spanish balancing market, and identifies market attributes that may hinder RES participation, while [29] provides an overview of the procurement side of the Spanish balancing market. Ref. [30] provides a general overview of the Nordic short-term markets and analyzes the participation of wind generators, with a focus on the ID market, and [31] provides an overview of the procurement side of the Nordic balancing market.

The second group of studies focus on design of short-term markets that include multiple market zones. Ref. [32] discusses general principles of the Dutch markets, and of the German ID market, while assessing wind generator bidding strategies. Ref. [3] analyzes negative prices that occasionally occur in DA, ID, and RT markets in Belgium, France, and Germany, and generally describes these markets. In [10], the implementation of flow-based market-coupling (FBMC) in the CWE DA markets is discussed. Ref. [6] analyzes DA and ID market liquidity in France, Germany, Scandinavia, Spain, and the United Kingdom (UK), while discussing a selection of design attributes. In [33], ID trading activity and prices are analyzed for a variety of market zones in Northern Europe, while also discussing the impact of imbalance settlement rules in the Nordic countries on ID trading. Ref. [34] compares balancing market design parameters across 28 countries in Europe, while [35–40] include empirical analyses of ID and RT markets in Germany, Italy, Poland, Spain, the Netherlands, and the United Kingdom, respectively. Ref. [41] compares the settlement side of the Dutch, German, and Nordic balancing markets. In [42], a valuable discussion on a variety of design parameters for the three short-term markets is provided, based on observations from Germany, Italy, the Netherlands, Poland, Spain, and the UK, to identify aspects which may benefit from harmonization. In addition, in 2014 the Belgian, German, and Dutch TSOs published a study on potential cross-border cooperation [43], in which the design of their RT balancing markets is discussed. Finally, Ref. [44] compares the European DA market design with designs adopted in the US.

The third group of references are not zone-specific, but have a more general focus. In [45] a variety of market design issues for RES integration are discussed, while [46] analyzes the impact of RES support schemes and certain market design parameters on RES integration. Ref. [47] analyzes expected trading volumes in different European ID markets, while the general functioning of continuous and auction-based ID markets is explained. Ref. [48] discusses different design parameters for balancing markets, in order to facilitate wind integration, while [49] identifies important balancing market design parameters for both individual control areas and cross-border cooperation. Finally, Ref. [50] assesses different alternatives to allocate cross-border transmission capacity in Europe, and [51] analyzes the impact of imbalance pricing on the behavior of market participants.

### 1.3.2 Research gap in market design analyses and contributions

Although the design of short-term markets in the CWE region has been discussed before, previous works focus on individual or a limited selection of (1) design parameters, (2) sequential markets, or (3) geographical market zones. In addition, the extent to which the market design affects the needs for and rewards to flexibility is not considered. An integrated discussion of design parameters, and their interaction with the demand for and supply of flexibility, for all three short-term markets and all four CWE market zones has thus not been provided before. As such, this article answers two research questions. First, how are the markets related to flexibility, i.e., the short-term markets, designed in the CWE region? Second, how do these markets express the need for and reward the supply of flexibility? The answers to these research questions provide insight in whether flexibility is treated consistently and appropriately among the different geographical and sequential markets.

Section 2 focuses on the DA markets, while Section 3 studies the ID markets, and Section 4 encompasses the settlement and procurement side of the RT balancing markets. In each section we focus on the key design features and parameters. The structure of the discussion for each parameter is as follows: we first summarize the design for the CWE market zones, and then hypothesize and analyze the implications for flexibility. The intent is to encourage policy-makers to consider market reforms that would facilitate the integration, availability, or valorization of flexibility, and also to contribute to the decision-making of flexibility investors and operators. In addition, at the end of each short-term market’s discussion, we provide a summarizing table, and present the trading volume as a share of the consumption for 2012–2015 to give an idea of its size. Finally, Section 5 states the conclusions, and Appendix A provides a list of abbreviations.

### 2. Day-ahead markets

In the DA market, which is a double-sided blind auction facilitated by power exchanges, market players trade hourly and multi-hourly products to adapt their position from the previously held forward and future markets. These positions, resulting in scheduled output profiles, can be adjusted by submitting demand and supply quantity-price bids before DA market closure, which is at noon (12:00 pm) D-1 in the CWE region [52]. The price in a demand bid indicates the highest price a buyer is willing to pay, while the price in a supply bid indicates the lowest price at which a seller is willing to sell. The intersection of the aggregated demand and supply curve determines the market-clearing volume and price. The DA market is based on a pay-as-cleared principle, through which all cleared demand bids in a market zone pay a uni-

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Literature review classification.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day-ahead (DA)</td>
</tr>
<tr>
<td>BE</td>
<td>[3,10]</td>
</tr>
<tr>
<td>FR</td>
<td>[3,6,10]</td>
</tr>
<tr>
<td>DE</td>
<td>[3,6,10,13,14,22]</td>
</tr>
<tr>
<td>NL</td>
<td>[10,24,32,42]</td>
</tr>
<tr>
<td>Other</td>
<td>[6,27,30,42,44]</td>
</tr>
</tbody>
</table>
form market-clearing price, while all cleared supply bids in a market zone are remunerated by that same price. In the CWE region the market zones coincide with the countries, except for the German market zone, which includes Germany, Austria, and Luxembourg. However, there are plans to split the German market zone in 2018 in order to create a separate market zone for Austria [53]. While DA market trading within market zones is not constrained by the internal electric grid, its interaction with neighboring market zones is constrained due to the limited interconnection capacity. As a result, DA prices may differ between market zones when interconnection lines are congested [3,54]. Since the planned output schedules after gate closure of the DA market may lead to congestion within a zone, the TSO may be required to perform redispatch actions to clear that congestion [55].

In the remainder of this section we discuss the CWE DA markets’ order types, temporal resolution, cross-border trading, price cap and floor, and trading volumes.

2.1. Day-ahead market order types

The standardized orders in DA markets are limit orders, i.e., hourly offered or requested quantities with a certain price limit. Besides these hourly products, power exchanges may also allow so-called complex orders. In the CWE region these include block orders, linked block orders, and exclusive block orders. Block orders are used to link several hours, whose quantity may differ for each hour, on an all-or-nothing basis. That is, either the bid is matched on all hours or is entirely rejected. The acceptance of a block order depends on its bid price and the volume-weighted average DA price in the hours contained in the block. A linked block order is a block order that is part of a set of multiple block orders that have a linked clearing constraint, while an exclusive block order is part of a set of block orders of which, at most, one can be cleared [1,56]. Complex orders allow the implicit inclusion of cost non-convexities and intertemporal links. These include startup and shut-down costs, ramp rates, minimum load levels, minimum up and down times, and energy buffer dynamics [57]. In addition, they provide a means for more robust bidding under price uncertainty. However, the amount of complex orders is limited per market participant for computational reasons [58].

In addition, the introduction of so-called storage orders is currently being discussed [59]. This market product would require storage operators to provide technical parameters (i.e., stored energy at the start of the delivery day, energy and power bounds, (dis)charge efficiency, and exogenous power flows) and economic parameters (i.e., the price below which charge bids and above which discharge bids may be cleared). Given these parameters, storage participation can be decided upon by the market-clearing algorithm.

In contrast to block bids, so-called multi-part bids (which are present in some US and European (e.g., Spain) markets) allow players to explicitly include technical and economic parameters in their bids [60–62]. As block bids require available flexibility to be offered in a standardized format, which may not allow to represent all technological capabilities, we argue that block bidding systems prevent the full available capacity and flexibility from being offered to the market. By allowing current generation, storage, and consumption capacities to formulate their availability more accurately without being restricted by the rigidity of current bid types, part of the future need for flexibility can already be accommodated [42]. In contrast, it has also been argued that a design with a limited number of standardized products may prove more advantageous as transaction costs may be lower, transparency may be higher, and markets may be more easily harmonized and coupled [44]. This is a fundamental debate, and represents a difference in philosophy between what broadly can be called the European and US approaches to DA market design.

2.2. Day-ahead market temporal resolution

The DA market is based on hourly market periods, but the settlement period over which market participants are responsible to have a balanced portfolio, are quarter-hourly periods in Belgium, Germany, and the Netherlands, and semi-hourly periods in France. We argue that shorter market periods would allow for an improved alignment with the settlement period, because when market players have the opportunity to trade at a sub-hourly time scale, besides inter-hourly, this allows intra-hourly expected power variations to be dealt with by means of trading in the DA market. While not having this possibility may not be a problem for players with self-balancing capabilities, other (usually smaller) players face the financial risk of being dependent on the TSO to deal with those variations in RT by activating reserves, thereby facing the imbalance price. It is clear that this is especially the case if the ID market is based on hourly market periods as well, or if sub-hourly ID products are characterized by low liquidity.

The illustrative example provided in Fig. 1 shows how hourly DA market periods challenge players who face intra-hourly power variations, with limited self-balancing capabilities, to avoid imbalanced positions (Fig. 1a). While this player has a net balanced position, on average, over the DA market period 10:00 am–11:00 am, imbalances occur in each quarter-hourly imbalance settlement period due to discrepancies between the temporal resolutions. If the DA market period and imbalance settlement period would be aligned, such imbalance positions can be avoided (Fig. 1b).

Quarter-hourly intervals would thus shift some of the flexibility demand from the TSO in RT to the DA market, and some of the flexibility supply from self-balancing to the market and from BSPs to DA market participants. As we argue in Section 1.2, a finer temporal resolution would also improve the extent to which the value of flexibility for the system is reflected and rewarded, because the resulting price signals would represent the physics of the system more accurately [42,45]. While a finer temporal resolution would improve the valorization of flexibility in the DA market, its additional value as part of a portfolio compared to standalone flexible capacity might decrease. This can be explained as intra-hourly variations can then also be dealt with through trading, not only through self-balancing using one’s own flexible resources.

All of this leads to the hypothesis that BRPs with limited self-balancing capabilities and TSOs may thus advocate for a finer DA temporal resolution aligned with the imbalance settlement period. Contrarily, challenges may arise for players whose resources are subject to intertemporal links and non-convex costs, e.g., their start-up cost has to be recovered within the bid of a shorter period. However, the latter counter-argument can be dealt with by the availability of adequate market products to deal with such constraints (i.e., sufficiently complex orders, multi-part bidding).

2.3. Cross-border trading in the day-ahead market

Through the so-called price coupling of regions (PCR) initiative, 23 European countries’ are currently coupled through the implicit auctioning of interconnection capacity [63,64]. This means that all bids of the participating exchanges are considered in the same market-clearing algorithm to optimize the utilization of interconnection capacity available to the power exchanges. Market players only provide bids for electric energy, while interconnection capacity

---

2 These include Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the UK.
is allocated implicitly to individual bids to maximize social welfare. As a result, electric energy is exchanged in case of a price difference between geographical markets until the price difference is eliminated or until all available interconnection capacity is used. In contrast, explicit auctioning indicates that interconnection capacity is allocated to individual market players that have obtained the right to use it, after which they can use this capacity to capture price differences between market zones. Such explicit auctioning was the basis for managing limited interconnection capacity prior to the establishment of the European market-coupling-based system [1,8,65].

The interconnection capacity available for trade is challenging to determine as electric energy flows according to Kirchhoff's laws, i.e., over all parallel paths in the network, not according to commercial flows, i.e., directly from generator to consumer. In addition, this determination has to make assumptions about the within-zone distribution of generation and consumption. Since May 2015, market-coupling of the DA markets in the CWE region is based on FMBC instead of the Available Transfer Capacity (ATC) method. The ATC value on a border represents the maximum commercial exchange between the two adjacent market zones, taking into account expected market outcomes, security margins, and long-term transmission capacity nominations. The TSOs calculate this ATC value prior to the market-clearing process for each direction on each border of its control area. In contrast, in the FMBC method, a simplified representation of internal grid constraints, i.e., the collection of critical lines, is included in the market-clearing process. Prior to the market-clearing the TSOs determine the FMBC parameters that define the so-called “FMBC flow domain”, in which each boundary refers to the limit of a critical line, while during market-clearing all critical lines are taken into account. In general, FMBC is believed to result in more available interconnection capacity for trade, increased social welfare, and increased price convergence between market zones [10].

Interconnection capacity can either be allocated through physical transmission rights (PTRs) or financial transmission rights (FTRs) via long-term (i.e., yearly and monthly) auctions. At the time of writing, part of the available interconnection capacity at the French-Belgian and Dutch-Belgian borders is allocated through FTRs (since January 2016), while at the Dutch-German and French-German borders this is done through PTRs. A PTR includes the exclusive right to use part of the transmission line. In the CWE region, PTRs are subject to a use-it-or-sell-it principle, which means that if a PTR holder does not actually nominate the corresponding capacity, it is transferred to the power exchange to use in the market-clearing, in addition to the capacity not sold through long-term auctions. In case of a positive price difference between the two market zones in the direction of the PTR, the PTR holder is paid this price difference for all non-nominated capacity. Contrarily, FTRs are financial instruments, so they do not give their holder the exclusive right to use part of the line. All physical capacity subject to the FTRs is transferred to the power exchange, thereby not interfering with optimal market-clearing. What the FTR holder is entitled to is a payment equal to the price difference between the two locations for all transmission capacity subject to the FTR in case of a positive price difference between the two market zones in the direction of the FTR. This use of FTRs is consistent with the US system, where it is believed that separating financial rights from physical operation results in a more efficient use of the grid and hedging of risks [50,66,67].

Since variability is spread over a larger area through market-coupling, non-correlated power variations are smoothed and opposite power variations compensate each other. We conclude that this results in lower total flexibility needs. In addition, interconnection capacity is a “vehicle” for flexibility. Flexibility in neighboring areas can be used to compensate for the local system’s variability, allowing the sourcing of cheaper flexibility abroad. Alternatively, local flexibility providers have the opportunity to offer services to other regions as well. This reasoning shows that market-coupling thus impacts flexibility demand and supply by enlarging the relevant geographical market to trade flexibility [65].

---

**2.4. Day-ahead market price cap and floor**

The DA markets in the CWE region include a price cap of 3000 €/MWh [52]. Such a cap is usually implemented to avoid excessive pricing by generators, especially when the price elasticity of demand is rather low. Although in general price caps should be set at the value of lost load (VOLL) [69] in order to encourage investment in needed peaking plants, they can be set lower if the purpose is market power mitigation. The VOLL represents the average value that consumers attach to a unit of electric energy not supplied, and thus reflects their willingness-to-pay to avoid demand curtailment. Although country-specific estimates are available in the existing literature, and they depend on many specifics (e.g., notification time, duration, time of the day), they are typically between 2000 €/MWh and 20000 €/MWh [45,70]. Because of the large-scale integration of variable RES in the generation mix, conventional power plants currently experience diminishing profitability due to a decreasing number of operating hours and lower electricity prices [7]. The imposed price cap should thus be high enough to allow these conventional power plants to recover their fixed investment cost over a decreased number of operating hours. Although capacity mechanisms are currently also discussed and implemented in the

---

3 The price elasticity of demand refers to the relative change in demand due to a relative change in price, and is typically negative, as the demand for most commodities decreases with the price [68].
CWE region to ensure generation adequacy [71,72], such markets are out of the article’s scope as we focus on flexibility, not adequacy. It is possible for capacity markets to be designed to incent investment and retention of flexible capacity, as in California, but this is not presently done in Europe [73]. Next to system adequacy, we argue that, when set too low, price caps interfere with the market signal representing the scarcity of upward flexibility.

The price floor is set at −500 €/MWh [52]. In contrast to price caps, Ref. [45] argues there is no incentive to impose price floors in electricity markets. Negative prices occur when conventional generators are willing to pay to generate in order to avoid costly shut-downs or downward ramping, and when renewable generators are willing to pay up to the subsidies they receive in order to generate [3,7]. When set too high, we argue that price floors interfere with the market signal representing the scarcity of downward flexibility, and may result in the arbitrary curtailment of generators during excess generation periods rather than having a pricing mechanism determine who values most being kept on.

2.5. Day-ahead market trading volume and summary

The DA market plays an important role in terms of trading volume. Based on hourly traded volume data from the power exchanges [74–76], and country-specific hourly consumption data from the European network of transmission system operators for electricity (ENTSO-E) [77], Fig. 2 shows the hourly traded volume as a share of the hourly consumption for 2012–2015. The trading volume in the German and Dutch market zones represents a larger share of the consumption than that of the Belgian and French market. However, in the latter countries this share has been increasing significantly in recent years, while for Germany it has remained constant and for the Netherlands the share seems to be decreasing.

To summarize, Table 2 provides an overview of the key characteristics of the DA markets in the CWE region.

3. Intra-day markets

After the DA market-clearing, each BRP is required to submit a balanced position to the TSO for each settlement period. These so-called “nominations” provide information on the planned schedules for every individual unit, and usually differ from the accepted bids in the DA market as they take into account all transactions, including the volumes traded in the previously held long-term markets and through bilateral contracts. However, these nominations can still be adjusted through trade in the ID market based on updated information (e.g., more accurate RES generation forecasts). ID trading is the last opportunity for market-based transactions before submitted schedules become financially binding.4 After gate closure of the ID market, the TSO takes over the responsibility to keep the system balanced. It is clear that the possibility of ID trading may shift a share of the flexibility needs away from RT to the ID stage, i.e., from the TSO to BRPs, and likewise for the supply of flexibility, i.e., from BSPs to ID market players.

This section discusses the CWE ID markets’ market types, order types and temporal resolution, cross-border trading, price cap and floor, and trading volume.

3.1. Intra-day market types

While the Belgian, Dutch, and French ID markets are based on continuous trading, the German ID market includes both continuous trading and a discrete auction. With continuous trading, market participants submit supply and demand bids to a central platform, and matching bids are continuously cleared on an individual basis. Continuous trading is possible from 02:00 pm D-1 in Belgium, and from 03:00 pm D-1 in France, Germany, and the Netherlands.5 This trading can occur until close-to-RT, i.e., until 5 min to RT in Belgium and the Netherlands, and until 30 min to RT in France and Germany. The continuous trading order book is visible to all market participants, and contains all submitted bids that have not cleared yet. In addition, players can cancel submitted non-cleared bids at any time. Continuous trading bids are matched according to a price-time priority: orders are matched in order of the attractiveness of their price, with the time of submission to the central platform being a tie-breaker if there are two identical price offers. A continuous trading ID market may thus result in different prices for each trade, with the price being the price of the bid that initiated the match (i.e., the first of the two involved bids, or the “initiator”), which may be referred to as pay-as-bid. The discrete auction implemented in the German ID market since December 2014 is based on principles similar to the DA market. Players submit supply and demand bids, with gate closure at 03:00 pm D-1. These bids are then aggregated to form the supply and demand curve. The intersection determines the uniform market-clearing price [15,42,78].

We argue that there are four implications of these designs for flexibility. First, compared to a discrete auction, continuous trading may pose less risk to flexibility consumers and suppliers as they can procure and valorize flexibility immediately instead of having to wait until market-clearing, that is, if there still is a market-clearing to come. Second, when bidding truthfully, non-marginal flexibility providers may face lower remuneration for their services compared to an auction-based ID market including pay-as-cleared pricing if they are the initiator. Meanwhile, non-marginal flexibility consumers may satisfy their flexibility need at a higher cost if they are the initiator. Naturally, in such cases the (second-mover) counterparty faces more favorable prices compared to a discrete auction, but risks losing the opportunity to match bids if it waits too long. Thus, continuous trading may incentivize players to not bid truthfully, which may result in incorrect flexibility demand and supply signals. Third, a market based on continuous trading instead of discrete auctions includes a certain first-come-first-serve characteristic, as matching bids are immediately cleared, which may not lead to welfare maximization and optimal allocation of flexibility, especially in illiquid markets. Finally, an important question regarding the organization of ID auctions that is not answered yet includes the optimal number of auctions and their timing, taking into account the impact on liquidity.

3.2. Intra-day market order types and temporal resolution

The Belgian, French, and Dutch continuous ID markets include both hourly and multi-hourly (i.e., block) products, while the German ID market also includes quarter-hourly products. In contrast to the DA market, the German ID auction includes 96 quarter-hourly market periods. Block bids have not been introduced yet in that auction [15,78].

While market participants have the opportunity to update their nominations, which are submitted after clearing of the DA market, on an hourly basis, this does not allow them to tailor output schedules through market-based transactions to the temporal resolution on which their imbalance positions are calculated. The presence of quarter-hourly products in the German ID market provides this opportunity, as players are able to compensate for the misalign-

4 Submitted schedules are not physically binding, unfilled positions (i.e., imbalance positions) are settled at the imbalance price set by the TSO (see Section 4.2).

5 While hourly products can be traded from 03:00 pm D-1 in the German continuous ID market, quarter-hourly products can only be traded from 04:00 pm D-1.
ment between the DA market periods and RT settlement periods (Fig. 3), and incorporate updated information on an intra-hourly basis. As such, players with limited self-balancing capabilities can be less dependent on the TSO to balance their imbalance positions, and are thus less exposed to imbalance prices. Following this reasoning, lower reserve needs are expected for the TSO in RT. Similar to the DA market, we hypothesize that a finer temporal resolution would improve the valorization of flexibility in the ID market due to price signals, but might reduce the additional value of flexibility in a portfolio compared to standalone flexible capacity. This is because intra-hourly variations can now also be dealt with through ID trading, not just through self-balancing.

3.3. Cross-border trading in the intra-day market

In contrast to the DA market, ID markets are currently less well aligned and integrated [42], and no interconnection capacity is reserved for the ID market. Here, only the residual cross-border capacity is made available to the market. In the CWE region, continuous ID markets are, depending on the considered border, coupled through the explicit or implicit allocation of remaining interconnection capacity, which is calculated according to the ATC method. Here, explicit means that market players can obtain remaining interconnection capacity for free on a first-come-first-serve basis, after which they can engage in cross-border ID trading. If however a player does not use the obtained capacity, this unfulfilled position will be settled (i.e., penalized) at the imbalance price since allocated transmission capacity automatically means nominated capacity [79]. On the other hand, with ID implicit allocation, orders in one zone are automatically matched with orders in the neighboring zone, as long as transmission capacity is available. Unlike explicit cross-border transactions, players do not need to obtain the interconnection capacity before making a transaction. On the Dutch-Belgian border, implicit continuous capacity allocation applies [80], while on the Dutch-German and the French-Belgian border, explicit allocation is in place [81,79]. Finally, on the French-German border, implicit allocation runs in parallel with explicit allocation, the latter for OTC trading purposes [82].

With continuous trading, the value of the transmission capacity is captured by the first mover in case of explicit auctioning and by the first matching cross-border bids under implicit auctioning. In general, this fails to maximize welfare because the capacity is not necessarily allocated to the most valuable transactions. In addition, this does not provide revenues to the owner of the intercon-

---

**Fig. 2.** The hourly DA market trading volume as share of the hourly consumption, for Belgium (a), France (b), Germany, Austria, and Luxembourg (c), and the Netherlands (d).

**Table 2**

Key characteristics of the (fully harmonized) CWE region’s DA markets.

<table>
<thead>
<tr>
<th></th>
<th>Belgium</th>
<th>France</th>
<th>Germany</th>
<th>The Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power exchange</td>
<td>BELPEX</td>
<td>EPEX SPOT</td>
<td>EPEX SPOT</td>
<td>APX</td>
</tr>
<tr>
<td>Market closure</td>
<td>12:00 pm D-1</td>
<td>12:00 pm D-1</td>
<td>12:00 pm D-1</td>
<td>12:00 pm D-1</td>
</tr>
<tr>
<td>Market period</td>
<td>1 h</td>
<td>1 h</td>
<td>1 h</td>
<td>1 h</td>
</tr>
<tr>
<td>Price cap</td>
<td>3000 €/MWh</td>
<td>3000 €/MWh</td>
<td>3000 €/MWh</td>
<td>3000 €/MWh</td>
</tr>
<tr>
<td>Price floor</td>
<td>500 €/MWh</td>
<td>500 €/MWh</td>
<td>500 €/MWh</td>
<td>500 €/MWh</td>
</tr>
<tr>
<td>Block orders</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Linked block orders</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Exclusive block orders</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
nector (that could be used to lower costs for consumers or incentivize new investments in interconnection capacity), as is the case in the DA market through the sale of transmission rights. In contrast, it has been argued that the use of discrete ID auctions could facilitate the most efficient allocation of the remaining interconnection capacity and remunerate the owner similar to the situation in the DA market [15,42].

When the ID market is rather illiquid, flexibility consumers may not find a counterparty, and as such become exposed to imbalance penalties, while flexibility suppliers may not be able to valorize their flexibility. It is obvious that this can be dealt with by matching bids over a larger geographical area through cross-border ID trade.

3.4. Intra-day market price cap and floor

In the CWE region, the continuous ID market price cap is set at 9999 €/MWh, and the price floor at −9999 €/MWh, which represents a wider range compared to the DA market, while for the German ID discrete auction these price limits are more similar to the DA market at 3000 €/MWh and −3000 €/MWh, respectively. A similar reasoning as for the DA market applies with respect to the rationale behind price limits and their interaction with flexibility.

3.5. Intra-day market trading volume and summary

The ID market plays a minor role in terms of trading volume, but it is an important tool to guarantee the reliable operation of the power system: each trade may contribute to a reduction in the activation of reserves by the TSOs. Based on hourly and quarter-hourly traded volume data from the power exchanges, and country-specific hourly consumption data from ENTSO-E, Fig. 4 shows the traded volume as a share of the hourly consumption for 2012–2015. While for all four market zones the role of the ID market has increased in 2014–2015 compared to 2012–2013, a much larger share of the consumption is traded in the German ID market compared to the other market zones (note that the y-axis of Fig. 4c indicates a different scale). This may result from the fact that Germany experiences larger shares of RES generation, and of Fig. 4c indicates a different scale). This may result from the fact that market players seem to have a large interest in quarter-hourly products (i.e., both in the continuous and auction-based ID market), which take a large share of the trading volume.

Table 3 summarizes the major attributes of the ID markets in the CWE region.

4. Real-time balancing markets

After gate closure of the ID market, unforeseen differences between scheduled and actual generation or consumption can occur. These differences may originate from (1) unexpected RES generation variations, (2) unexpected consumption variations, (3) unplanned outages of generation and consumption capacity, and grid elements, (4) discrepancies between the duration of DA/ID market periods and RT settlement periods, and (5) the discretization of continuous time in discrete market periods [17].

These events are dealt with in RT by the balancing market, which is coordinated by the TSO. First, the TSO calculates the total system imbalance (SI), i.e., the demand for flexibility in RT, resulting from the aggregated individual imbalances of BRPs. The TSO then compensates for this SI by activating reserves, i.e., the supply of flexibility in RT, that may have been contracted ahead of time from market participants who provide balancing services, i.e., BSPs. In general, this demand for the activation of reserves is rather small compared to the system load, and is highly volatile and price inelastic as well [21]. The reservation and activation of reserve capacity is referred to as the procurement side of the balancing market. Afterwards, the TSO settles individual imbalances with BRPs by applying imbalance prices to their imbalance positions. This is referred to as the settlement side of the balancing market. BRPs thus “trade” balancing energy with the TSO, which in turn procures these services from BSPs [7,17,48]. BRPs can also deal with unexpected power variations by activating flexibility in their own portfolio instead of relying on the TSO to provide flexibility at the settlement side of the balancing market.

4.1. Control areas

Belgium, France, and the Netherlands each consist of one control area, each managed by one TSO, i.e., Elia, RTE, and TenneT, respectively [25]. In contrast, Germany includes four control areas, each being operated by its own TSO: Amprion, 50 Hertz, TenneT, and TransnetBW. The German TSOs cooperate to keep their control areas balanced through the so-called Grid Control Cooperation (GCC). The GCC includes four modules that were introduced one after the other, and focuses on automatic frequency restoration
Fig. 4. The hourly ID market trading volume as share of the hourly consumption, for Belgium (a), France (b), Germany, Austria, and Luxembourg (c), and the Netherlands (d).

Fig. 5. The hourly German ID market trading volume (a) continuous vs. auction-based trading, and (b) hourly vs. quarter-hourly products, both expressed as share of the total ID trading, 2015.

Table 3: Key characteristics of the (partially harmonized) CWE region’s ID markets.

<table>
<thead>
<tr>
<th></th>
<th>Belgium</th>
<th>France</th>
<th>Germany</th>
<th>Germany</th>
<th>The Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power exchange</td>
<td>BELPEX</td>
<td>EPEX SPOT</td>
<td>EPEX SPOT</td>
<td>EPEX SPOT</td>
<td>APX</td>
</tr>
<tr>
<td>Continuous trading</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Market opening</td>
<td>02:00 pm D-1</td>
<td>03:00 pm D-1</td>
<td>04:00 pm D-1</td>
<td>03:00 pm D-1</td>
<td>03:00 pm D-1</td>
</tr>
<tr>
<td>Market closure</td>
<td>5 min to RT</td>
<td>30 min to RT</td>
<td>30 min to RT</td>
<td>30 min to RT</td>
<td>5 min to RT</td>
</tr>
<tr>
<td>Market period</td>
<td>1 h</td>
<td>1 h</td>
<td>15 min</td>
<td>1 h</td>
<td>1 h</td>
</tr>
<tr>
<td>Price cap</td>
<td>9999.99 €/MWh</td>
<td>9999 €/MWh</td>
<td>9999 €/MWh</td>
<td>9999 €/MWh</td>
<td>9999.90 €/MWh</td>
</tr>
<tr>
<td>Block orders</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Discrete auction</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Market closure</td>
<td>–</td>
<td>–</td>
<td>✔</td>
<td>–</td>
<td>✔</td>
</tr>
<tr>
<td>Market period</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>15 min</td>
<td>–</td>
</tr>
<tr>
<td>Price cap</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3000 €/MWh</td>
<td>–</td>
</tr>
<tr>
<td>Price floor</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>X</td>
</tr>
<tr>
<td>Block orders</td>
<td>–</td>
<td>–</td>
<td>✔</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
reserve (aFRR), i.e., secondary control.⁷ The objective of the first module, imbalance netting, is to reduce the total activation of reserve capacity by avoiding counteracting activations. The second module includes a common reserve sizing and access to reserve capacity in other control zones in case of a local shortage. The third module procures reserves using a common market, i.e., bids are accessible to all TSOs. The implementation of the fourth module leads to a common merit-order for the activation of reserve capacity in order to activate the cheapest bids, respecting limits on the connecting transmission lines [17,23,84]. In contrast, the activation of manual frequency restoration reserve (mFRR), i.e., fast tertiary control, is currently still done in a decentralized way, but is based on common rules. This advanced inter- TSO cooperation also includes a single imbalance price over the different control zones, whose calculation is discussed in Section 4.2. We conclude that such cooperation is likely to provide similar benefits as market-coupling in the DA and ID market, and similar impacts on the need for and supply of flexibility.

4.2. Settlement side

In the settlement side of the balancing market, the BRPs’ imbalance positions and the imbalance prices are determined. As stated before, BRPs have to submit nominations for planned grid-exchanges on the plant-level to the TSO after DA market-clearing, and again after transactions in the ID market. Although this detailed plant-level information has to be reliable in order for the TSO to effectively analyze congestion on the internal grid of its control area, there is no incentive for BRPs to do this on a plant-level, since they can only be held accountable for imbalanced positions aggregated over their portfolio. Although it has not been quantified to which extent this may be a problem, a requirement for plant-level balance responsibility might ensure that TSOs receive more reliable information [42].

A BRP’s imbalance position is the difference between the nominated position after closure of the ID market and the actual net exchange of electric energy with the grid in RT. During each settlement period, a BRP can have a long, short, or balanced position. A long position indicates a positive imbalance, thereby injecting more and/or withdrawing less than planned. A short position indicates a negative imbalance, in which the BRP injects less and/or withdraws more than planned. Both long and short imbalance prices have to be determined by the TSO. BRPs with a long position in RT receive the long imbalance price, while BRPs with a short position in RT pay the short imbalance price [51].

Through the imbalance settlement, the TSO allocates the activation cost of reserve capacity (in €/MWh) to responsible BRPs, while reservation costs (in €/MW) associated with contracting reserves are recovered through grid tariffs. Imbalance prices are available shortly after RT in Belgium, France, and the Netherlands, while German imbalance prices are only published a few weeks after delivery [34,43]. As the former provides BRPs with valuable market information soon after RT, which can be used to make informed decisions with respect to their portfolio, we hypothesize that BRPs prefer this to the latter.

In the remainder of this section we discuss the settlement side’s temporal resolution, pricing rules, and price cap and floor.

4.2.1. Imbalance settlement temporal resolution

The Belgian, Dutch, and German balancing markets are based on quarter-hourly market periods, while semi-hourly periods are used in France [34]. Since imbalances vary on a continuous basis, instantaneous imbalance positions of BRPs differ from measured net imbalance position over the settlement period. As BRPs are only held accountable for the net imbalanced position over a period, players that caused large instantaneous imbalances, and thus the activation of additional reserves, may not be charged for the costs they have caused. As this issue is related to intra-settlement period imbalances, it is expected to occur less with shorter settlement periods [85]. These provide an incentive to BRPs to keep their portfolio balanced on shorter time frames, thereby most likely increasing the demand for flexibility by BRPs, either as part of their portfolio or by means of ID trading, which in turn would lower the flexibility needs of the TSO in RT.

4.2.2. Single-pricing vs. dual-pricing

Imbalance prices can either be calculated through a dual or single-pricing scheme. In a dual-pricing scheme, the imbalance price applied to BRPs’ imbalanced positions in the same direction as the SI is based on the activation cost of reserve capacity, while the imbalance price applied to BRPs’ imbalances in the opposite direction of the SI is typically based on the DA price. In contrast, within a single-pricing scheme, a uniform imbalance price, based on the activation cost of reserve capacity, is applied to all BRPs having an imbalanced position [7,17,48].

In France a dual-pricing scheme is applied, while the calculation of the imbalance price in Belgium, Germany, and the Netherlands is each based on a single price that applies to all imbalances [7,34]. However, the imbalance pricing scheme in the Netherlands is not a pure single-pricing mechanism, since in case both up and downward reserves are activated the prices differ: the long imbalance price equals the marginal activation price for downward reserve while the short imbalance price is the one for upward reserve [78].

In addition, in Belgium and the Netherlands, the imbalance price applied to short and long positions differs in the event of large imbalances, thereby moving from a single to a dual-pricing scheme. This is done by including a balance-incentivizing component, to either punish BRP imbalances in the same direction as the SI or to incentivize all BRPs to keep their balance. Although such a component is applied in Germany as well, it does not result in different short and long imbalance prices. The imbalance price is increased for all BRPs if the system is short, and decreased for all BRPs if the system is long [23]. In France, the imbalance price applied to BRP imbalances in the direction of the SI is adjusted by a multiplier 1 + K, set at 1.08 since July 2011 [86].

To lessen the need for activation of reserve capacity on the procurement side of the balancing market, BRPs can help the TSO keep the system balanced by intentionally incurring imbalanced positions in the opposite direction of the SI, which can be referred to as “passive balancing” [87]. However, we argue that with dual-pricing there is unfortunately little incentive to provide passive balancing since the DA price is applied to imbalances in the opposite direction of the SI. In contrast, single-pricing schemes incentivize BRPs to perform passive balancing. The reasoning for this proposition proceeds as follows:

- In case of a negative SI, the TSO activates upward reserve. Typically, this is activated at a higher price compared to the DA price, and a greater quantity of activated upward reserve results in a higher imbalance price, as it is selected according to a merit-order of increasing activation prices. This incentivizes BRPs to have a long position.
- In case of a positive SI, the TSO activates downward reserve. Typically, this is activated at a low price compared to the DA price, and a larger amount of activated downward reserve results in a lower imbalance price, as it is selected according to a merit-order of increasing activation prices from the TSO.

⁷ The different reserve categories are described in Section 4.3.
point of view, i.e., decreasing resulting imbalance prices when considering downward reserve capacity. This incentivizes BRPs to have a short position.

However, some regulators prefer a dual-pricing scheme (e.g., France), as this avoids BRPs to be incentivized to speculate on the direction of the SI [85]. In addition, in Germany BRPs are contractually not allowed to deviate from their nominated position to benefit from favorable imbalance prices, even though that might help the system [42]. Instead, both countries rely solely on the TSO to balance the system by activating reserve capacity, which can be justified in two ways. First, deviations based on passive balancing are not communicated to the TSO, which makes congestion management difficult due to the lack of reliable information. Second, forecasts of different players on whether to incur a long or short position for passive balancing are often based on similar algorithms. In case the algorithms suggest the wrong direction, the need for balancing services would be further increased, while when multiple BRPs provide such passive balancing and their forecasts about the system state is correct, they may turn oversupply into undersupply situations and vice versa, instead of decreasing the absolute value of the SI. In contrast to Germany where this is not allowed, and to France where the dual-pricing scheme does not provide an incentive, in Belgium and the Netherlands passive balancing is possible and allowed. Those countries believe that such behavior can serve a valuable social purpose and contributes to the valorization of flexibility.

Finally, it has to be noted that in Belgium, Germany, and the Netherlands, BRPs can still change their submitted nominations after gate closure of the ID market, even after RT, by means of a so-called OTC “day-after” market [34,43]. This trading of individual BRP imbalances has no physical meaning, but changes the accounting in the settlement process. The volume traded in this market is small (or negligible) if market participants have an idea of the level of imbalance prices, as there are no win-win situations in case of single-pricing. In case the imbalance price is positive, the player with a long position is not willing to lose its long position since he receives the imbalance price. Meanwhile, in the case in which the imbalance price is negative, the player with a short position receives an income because of its imbalance as he pays the imbalance price. In Germany, day-after trading can reduce the uncertainty until the imbalance price is known a few weeks after delivery. However, as shown in [22], there is a high predictability of the approximate level of imbalance prices, rendering this market largely irrelevant in Germany as well [20,23].

4.2.3. Marginal pricing vs. average pricing

When the imbalance price reflects the procurement cost of the activated reserve capacity, it is either based on the marginal or average activation price [7,17,48]. With marginal pricing, the imbalance price is set to the price of the marginal accepted bid, while for average pricing the imbalance price is calculated by dividing the net total activation costs of the TSO by the net activated reserve volume. The imbalance price is based on marginal pricing in Belgium and the Netherlands, while France and Germany apply average pricing [19,21,34,51].

In general, there is a widely held view that marginal pricing provides BRPs with more accurate signals into the cost to cope with their imbalances, and as such gives a greater incentive to avoid imbalanced positions [85,21]. Since in single-pricing schemes it may be profitable for BRPs to deviate from their submitted schedules for passive balancing, it is clear that profits would rise with more extreme imbalance prices. Marginal pricing leads to more extreme prices compared to average pricing in case of TSO balancing actions in one direction, while in situations where the TSO activates both up and downward reserve capacity this may not always be true. However, with average pricing, the imbalance price is often capped by the marginal activation price (e.g., Germany). Therefore, we conclude that marginal pricing may be preferred by BRPs wishing to perform passive balancing actions, while average pricing appears to be advantageous for BRPs having limited access to flexible resources to avoid the risk of being exposed to more extreme (unfavorable) imbalance prices.

4.2.4. Imbalance price cap and floor

The imbalance price in the Belgian RT balancing market includes a price cap of 3000 €/MWh, and price floor of –3000 €/MWh. Contrarily, the Dutch imbalance price is less heavily bounded, with a (theoretical) price cap and floor of 100000 €/MWh and –100000 €/MWh, respectively [25]. The German imbalance price is, after calculation according to the average pricing principle, limited by the marginal activated up and downward reserve bid. Afterwards, this capped price is compared to the average volume-weighted ID price. In case the net regulation volume (NRV) is positive (i.e., SI is negative), the ID price represents a lower limit, while if the NRV is negative (i.e., SI is positive), the ID price represents an upper limit. Any remaining reserve activation costs that are not covered by the settlement mechanism are recovered through grid fees together with the reservation costs to contract reserve capacity [19]. In France, if the SI is positive, the imbalance price for BRPs with a long position is capped by the DA price, while if the SI is negative, the imbalance price for BRPs with a short position must at least equal the DA price [86].

High imbalance prices are market signals that represent a relative scarcity of cheap upward flexibility when facing negative SIs, while negative imbalance prices signal the scarcity of cheap downward flexibility with positive SIs [7].

4.3. Procurement side

At the procurement side of the balancing market, the TSO procures and activates reserve capacity from BSPs. Besides the distinction between up and downward reserve, ENTSO-E further categorizes reserve capacity into three groups. Frequency containment reserve (FCR), i.e., primary control, is activated automatically in a matter of seconds, in response to frequency deviations for the entire synchronous zone, and needs to have the ability to be fully operational in 0.5 min. Frequency restoration reserve (FRR) is either activated automatically (aFRR), i.e., secondary control, or manually (mFRR), i.e., fast tertiary control, and restores the system frequency by restoring the balance in the control zone, thereby relieving the activated FCR. Its activation is triggered by the area control error (ACE), which is calculated as the difference between the scheduled and actual power interchange of a control area. While aFRR capacity needs to be fully operational in 5–15 min, for mFRR capacity this is in 7.5–22.5 min, both depending on the control zone. Finally, replacement reserve (RR), i.e., slow tertiary control, can be used to support or relieve the activated FRR. RR is not further discussed as it is currently not used by the Belgian, German, and Dutch TSOs. Instead, they expect BRPs to already offset part of their imbalance by means of self-balancing and trading on the ID market [43,88].

In what follows we discuss the procurement side’s reserve remuneration, contract periods, selection and activation mechanism, cross-border cooperation, and market size.

4.3.1. Reserve remuneration

In general, BSPs providing FCR only receive a reservation payment. This is because no activation payment applies since up and

---

8 We focus on frequency control, as other grid services (i.e., black-start capabilities, voltage support, and congestion management) are not within the scope of this article.
downward FCR activations are expected to compensate each other and only represent very small volumes [22,48]. However, this is not the case in France, where the DA price serves as proxy for activation payments [89]. For contracted aFRR and mFRR, both reservation and activation payments apply.

Payments are either based on a fixed regulated price, pay-as-bid pricing, or pay-as-cleared pricing.\(^9\) The reservation of FCR is remunerated pay-as-bid in Belgium, Germany, and the Netherlands, and through a regulated price in France. Both the reservation and activation remuneration of aFRR is pay-as-bid in Belgium and Germany, while in the Netherlands pay-as-bid applies to the reservation and pay-as-cleared to the activation. In France the provision of aFRR is remunerated by means of regulated prices. For mFRR the reservation is remunerated pay-as-bid in all market zones except for France, where a regulated price applies, while the activation is remunerated pay-as-bid in Belgium, France, and Germany, and pay-as-cleared in the Netherlands [17,34,51,90].

We identify several reasons for the presence of reservation payments. First, in the case of FCR, a reservation payment is needed since usually no activation payments apply. Second, reservation costs compensate BSPs’ opportunity cost to keep the contracted capacity available. Third, they lower BSPs’ risk by yielding a guaranteed income, instead of having to rely on non-guaranteed activations. Fourth, they may contribute to efficient dispatch in the presence of cost non-convexities. Fifth, capacity payments may be a means to recover costs in case BRPs cannot pass on all of their costs via activation bids because of price limits [48]. In contrast, the main disadvantage of reservation costs is the difficulty in accurately allocating them to responsible BRPs, as they are currently just spread over all system participants through grid fees.

4.3.2. Reserve contract periods

Recently, there has been a move from long-term contracting to more frequent tenders for shorter durations. Currently, FCR is contracted on a weekly basis in Belgium, the Netherlands, and Germany, while aFRR is contracted for the duration of a year in the Netherlands, and a week in Belgium and Germany. While mFRR is contracted for the duration of a year in the Netherlands, and in Belgium both monthly and yearly contracts apply, in Germany daily tenders are organized for six 4 h periods. In France, the provision of FCR, aFRR, and mFRR is mandatory [21,34,43,90]. Next to contracted reserves, in Belgium and the Netherlands (for aFRR and mFRR), and in France (for mFRR), non-contracted voluntary bids are allowed until close-to-real-time. In contrast, the German TSOs only consider contracted reserves [43].

We identify four advantages of shorter contract periods. First, they allow TSOs to size reserve needs more accurately for the upcoming period. This may lead to lower total reserve needs, as with longer periods reserve requirements may be oversized for part of the period subject to the sizing. Second, when considering the supply of flexibility, shorter periods allow BSPs to better estimate their opportunity cost and thus more accurately price their service [20,91]. Third, they also lower the availability risk for BSPs, decreasing entry barriers and fostering competition. Fourth, shorter contract periods allow to more often and accurate arbitrage between different services, thereby improving the allocation of resources to the most profitable ones at each time step. In contrast, we identify two arguments against more frequent tender periods. First, financing new investments in flexibility requires accurate expectations of (future) revenue streams, which is easier with longer contract periods. Second, TSOs may prefer longer contract periods as this guarantees the access to sufficient reserve capacity for a longer period of time [92].

4.3.3. Reserve selection and activation mechanism

FCR is contracted according to a merit-order of increasing prices in the Netherlands and Germany, starting with the lowest, and is co-optimized together with aFRR in Belgium to minimize total combined reservation costs. The latter allows BSPs to better deal with short-term operating constraints. In the Netherlands aFRR is contracted such that its reservation cost is minimized, which may include an overshoot since bids may not be fully divisible. This may lead to situations where the lowest bid is not necessarily always selected first. An identical approach applies to the reservation of mFRR in both Belgium and the Netherlands. In contrast, in Germany aFRR and mFRR are reserved according to a merit-order of increasing prices, with the lowest bid selected first [43].

While the reservation of reserve is based on the reservation price, the activation of reserve is based on the activation price. While contracted aFRR is activated simultaneously and pro rata in Belgium and France, it is activated sequentially according to a merit-order in the Netherlands and Germany. In contrast, mFRR is subject to a sequential merit-order activation in all market zones. It is important to note that while contracted aFRR is activated pro rata in Belgium, non-contracted aFRR is subject to merit-order activation [34,43,93]. In general, pro rata activation inherently results in a more flexible reserve portfolio, but as all contracted sources are activated every time the reserve product is activated, we argue that this can pose a barrier to BSPs operating so-called limited energy resources that prefer infrequent activations.

4.3.4. Cross-border reserve procurement

Given the full harmonization of DA markets in the CWE region, and plans for increasingly coupled and harmonized ID markets, the next logical step would involve the RT markets. Although some initial steps have already been taken in this direction, much more work is required. In 2012 the TSOs from multiple neighboring countries, among which TenneT NL and Elia, joined the first module of the GCC, leading to the international GCC (IGCC) [7]. Besides this, TenneT NL (2014) and Elia (2016) joined the existing FCR common procurement platform of the German, Austrian, and Swiss TSOs, for a share of their FCR obligations [42]. The participation of RTE is planned for 2017. Further international integration, along the line of all four GCC modules of the German TSOs (see Section 4.1), is expected to result in less total reserve activations, more efficient sizing, and the reservation and activation of the most efficient bids across all participating zones.

4.4. Real-time market size and summary

In general, the NRV as share of the consumption has been decreasing in recent years (Fig. 6). We explain this as follows: improved RES forecast accuracy, better profiling in the DA and ID market, increased liquidity in the ID market, increased international cooperation, and increased passive balancing by BRPs. We argue that TSOs may further incentivize the latter by reacting slower to SiS, thereby temporarily keeping the SI at a higher level to attract additional passive balancing.

Table 4 gives an overview of the key characteristics of the RT markets in the CWE region.

5. Conclusions

In the CWE region, the need for and valorization of flexibility in electric energy supply and demand is primarily expressed in the short-term markets, defined as those taking place from the DA

---

\(\text{\textsuperscript{9}}\) For a discussion on the different pricing rules in reserve markets, we refer the reader to [17].
stage until delivery, including DA, ID, and RT markets. Due to the ongoing integration of variable RES, the variability in the system is increasing, making these markets increasingly important to keep the system balanced at different time scales. A good understanding of their design, as well as of new developments, is essential for analyses of the need for and supply of flexibility. This article therefore provides a detailed overview of the design of the three short-term markets for the four market zones of the CWE region, while focusing on their interaction with flexibility.

Considering the DA market, we discuss its general functioning, and review its specific features, including available order types, temporal resolution, cross-border trading, and price caps and floors. For the ID market, a similar set of topics is considered, while also discussing the difference between its two market types, i.e., discrete auctions and continuous trading. Finally, RT balancing markets include both an imbalance settlement and reserve procurement side. Considering the former, we analyze how the BRPs’ imbalance positions and the imbalance prices are determined, while for the latter, we investigate how the TSO procures, activates, and remunerates reserve capacity from BSPs.

While other potentially desirable reforms can be identified when considering the implications for flexibility discussed throughout this article, we conclude that policy-makers should focus on four design improvements. The first one is the temporal resolution. As BRPs are only held accountable for the net imbalance position over a settlement period, and not for instantaneous imbalance positions. For the ID market, a similar set of topics is considered, while also discussing the difference between its two market types, i.e., discrete auctions and continuous trading. Finally, RT balancing markets include both an imbalance settlement and reserve procurement side. Considering the former, we analyze how the BRPs’ imbalance positions and the imbalance prices are determined, while for the latter, we investigate how the TSO procures, activates, and remunerates reserve capacity from BSPs.

While other potentially desirable reforms can be identified when considering the implications for flexibility discussed throughout this article, we conclude that policy-makers should focus on four design improvements. The first one is the temporal resolution. As BRPs are only held accountable for the net imbalance position over a settlement period, and not for instantaneous imbalance positions.
ances, quarter-hourly settlement periods in the French RT market would increase the extent to which BRPs are charged for the reserve activation costs they cause. In addition, the introduction of quarter-hourly products in the Belgian, French, and Dutch continuous trading ID markets would allow players to align hourly DA output schedules through market-based transactions to the temporal resolution of the imbalance settlement, and improve the extent to which capacity is rewarded for its flexibility. The second recommended improvement is the introduction of an ID auction for Belgium, France, and the Netherlands. As the impact on liquidity needs to be taken into account, further integration of the CWE ID markets through market-coupling naturally follows, which would also promote an efficient allocation of flexibility and interconnection capacity. Third, as passive balancing can serve a valuable social purpose and improve the valorization of flexibility, incentivizing design changes should be considered for the French and German balancing markets. Fourth, more cross-border inter- TSO cooperation should be promoted on the procurement side in the CWE region’s balancing markets, similar to the cooperation of the four German TSOs. Such cooperation is expected to contribute to more efficient reserve sizing, reservation, and activation.

In terms of relative size of the three markets, trading volume analyses show that the DA market is an important market in the CWE region. In 2015, it comprised on average 28.19% (BE), 23.48% (FR), 46.21% (DE), and 38.47% (NL) of the hourly consumption. Although the ID market still plays a minor role in terms of trading volume, i.e., in 2015 on average 0.81% (BE), 0.94% (FR), 7.95% (DE), and 0.83% (NL) of the hourly consumption, these volumes have been increasing steadily over the past few years, and are expected to keep on growing. In Germany the ID market is relatively large due to the demand for ID flexibility resulting from high RES penetrations, and because of its relatively sophisticated market design including both continuous trading and a discrete auction, and both hourly and quarter-hourly products, all of which facilitate trade. Finally, the RT balancing market’s size, measured in terms of the NRV, has seen its share of consumption decrease from 2012 to 2015, with average shares falling from 1.33% to 0.95% (BE), from 1.19% to 1.05% (FR), from 1.38% to 0.63% (DE), and from 0.49% to 0.44% (NL). This can be attributed to more accurate RES generation forecasts, improved profiling in the DA and ID market, increased liquidity in the ID market, increased international cooperation, and increased passive balancing by BRPs.

We conclude that the details of market design are crucial to the successful integration of variable RES, as they determine the rules by which flexibility providers must play, and define the opportunities for these sources to valorize flexible operations. There are at least two important topics for future research. The first is the development of models to simulate the participation of flexibility in the three short-term markets, in order to be able to quantify its value in the CWE region. The second is the use of models or statistical analyses to analyze how specific market design rules may affect this value, in order to quantitatively complement the qualitative analyses of this article and thereby provide useful advice concerning future market redesigns.

Acknowledgment

The authors would like to thank Sauleh Siddiqui for hosting the first author at The Johns Hopkins University, Research Foundation Flanders (FWO) for providing a travel grant, Glenn Plancke for support in the trading volume analyses, and Christian Linnemann for feedback on the German balancing market. The third author was partially supported by US National Science Foundation grants IB 1243482 and ECCS 1230788.

Appendix A

Table A provides a list of abbreviations and their definitions.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aFRR</td>
<td>Automatic frequency restoration reserve</td>
</tr>
<tr>
<td>ATC</td>
<td>Available transfer capacity</td>
</tr>
<tr>
<td>BRP</td>
<td>Balance responsible party</td>
</tr>
<tr>
<td>BSP</td>
<td>Balance service provider</td>
</tr>
<tr>
<td>CWE</td>
<td>Central Western European</td>
</tr>
<tr>
<td>DA</td>
<td>Day-ahead</td>
</tr>
<tr>
<td>ENTSO-E</td>
<td>European network of transmission system operators for electricity</td>
</tr>
<tr>
<td>FMBC</td>
<td>Flow-based market-coupling</td>
</tr>
<tr>
<td>FRR</td>
<td>Frequency restoration reserve</td>
</tr>
<tr>
<td>FTR</td>
<td>Financial transmission rights</td>
</tr>
<tr>
<td>GCC</td>
<td>Grid control cooperation</td>
</tr>
<tr>
<td>ID</td>
<td>Intra-day</td>
</tr>
<tr>
<td>IGCC</td>
<td>International grid control cooperation</td>
</tr>
<tr>
<td>mFRR</td>
<td>Manual frequency restoration reserve</td>
</tr>
<tr>
<td>NRV</td>
<td>Net regulation volume</td>
</tr>
<tr>
<td>OTC</td>
<td>Over-the-counter</td>
</tr>
<tr>
<td>PCR</td>
<td>Price coupling of regions</td>
</tr>
<tr>
<td>PTR</td>
<td>Physical transmission rights</td>
</tr>
<tr>
<td>RES</td>
<td>Renewable energy sources</td>
</tr>
<tr>
<td>RR</td>
<td>Replacement reserve</td>
</tr>
<tr>
<td>RT</td>
<td>Real-time</td>
</tr>
<tr>
<td>SI</td>
<td>System imbalance</td>
</tr>
<tr>
<td>TSO</td>
<td>Transmission system operator</td>
</tr>
<tr>
<td>VOLL</td>
<td>Value of lost load</td>
</tr>
</tbody>
</table>

Country codes

<table>
<thead>
<tr>
<th>Country codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE</td>
<td>Belgium</td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
</tr>
<tr>
<td>FR</td>
<td>France</td>
</tr>
<tr>
<td>NL</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
</tbody>
</table>

References


