On Congestion Management in European Electricity Markets

Dissertation

zur Erlangung des Doktorgrades
Dr. rer. pol.
der Fakultät für Wirtschaftswissenschaften
der Universität Duisburg-Essen

vorgelegt von
Simon Voswinkel

aus
Attendorn, Deutschland

Betreuer:
Prof. Dr. Christoph Weber
Lehrstuhl für Energiewirtschaft

Essen, 17.05.2023

Erstgutachter: Prof. Dr. Christoph Weber

Zweitgutachter: Prof. Dr. Dominik Möst

Tag der mündlichen Prüfung: 19.12.2023

DOI: 10.17185/duepublico/81543

URN: urn:nbn:de:hbz:465-20240305-063201-0

Danksagung

Diese Dissertation ist während meiner Zeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Energiewirtschaft der Universität Duisburg-Essen entstanden. Sie wäre ohne die Mitwirkung und Unterstützung vieler Menschen nicht möglich gewesen.

Mein tiefster Dank gilt Professor Christoph Weber, welcher mir die Freiheit und Flexibilität ermöglicht hat, bei der Themenwahl der Paper ganz meinen Interessen nachzugehen. Auch für die unzähligen Anregungen, spannenden Diskussionen und das wertvolle Feedback möchte ich mich herzlich bedanken. Ein besonderes Dankeschön geht zudem an Professor Dominik Möst, welcher als Zweitgutachter zur Verfügung stand.

Meinen weiteren Ko-Autoren möchte ich für die fruchtbare Zusammenarbeit danken. Danke an meinen langjährigen Bürokollegen Jonas Höckner für die hervorragende Zusammenarbeit im Kontext von enera und unserer Paper, an Björn Felten und Tim Felling für die Einführung in die Komplexität des Flow-Based Market-Couplings, an Gerald Blumberg für spannende Einblicke aus der Praxis und an Michael Bucksteeg für seine umfangreichen Ratschläge und sein Feedback zum Abschluss des Promotionsprojektes.

Abschließend möchte ich meiner Partnerin Lisa Brotte danken, die während der gesamten Zeit mit ihrem Rückhalt und ihrer Unterstützung an meiner Seite war. Zudem danke ich meinen Eltern, welche das Fundament für alles gelegt haben.

Contents

Abbreviations

CACM Capacity Allocation and Congestion Management

CEE Central Eastern Europe

CNEC Critical Network Element and Contingency

CWE Central Western Europe

DSO Distribution System Operator

EHV Extra High Voltage

ENTSO-E European Network of Transmission System Operators for Electricity

FBMC Flow-Based Market Coupling

GSK Generation Shift Key

HV High Voltage

ISO Independent System Operator

LP Linear Program

LV Low Voltage

MIP Mixed Integer Program

MV Medium Voltage

NEMO	Nominated Electricity Market Operator
NLP	Non-Linear Program
OPF	Optimal Power Flow
PCR	Price Coupling of Regions
PTDF	Power Transfer Distribution Factor
RAM	Remaining Available Margin
RDpot	Redispatch Potential
RES	Renewable Energy Sources
SDAC	Single Day-Ahead Coupling
TSO	Transmission System Operator

Chapter 1

Introduction

The transition to clean and renewable energy sources (RES) poses challenges to electricity markets and the electricity transmission and distribution infrastructure. On the one hand, existing conventional power plants, powered by lignite, coal, and gas as well as nuclear power plants (in Germany) are replaced by (mostly) wind and solar power plants. On the other hand, mobility, heating, and many industrial processes will rely on electricity generated from RES for decarbonization in what has become known as sector coupling. This increases the total electricity demand and changes the temporal and spatial pattern of the demand, which puts further strain on the electricity grids and on the electricity markets that rely on the grid to function. An important challenge arising from these changes are congestions in the electricity grid. In order to secure stable grid operation, these congestions must be managed to avoid power flows that exceed the physical capacity of grid elements and would ultimately damage them.

This thesis explores different aspects of congestion management in European electricity markets. First, fundamental mechanisms and methods will be touched upon in this introduction in Section 1.1. Building on these concepts, Section 1.2 introduces research questions that this thesis addresses. At the core of this thesis are then five articles (Chapters 2 to 6), which investigate various topics related to these research questions. Finally, the conclusion in Chapter 7 highlights the contributions of the five articles to the research questions.

1.1 Background and motivation

This section starts by introducing the challenges faced by electricity grids (Section 1.1.1) and gives an overview of basic electricity market design alternatives and how they relate to congestion management (Section 1.1.2). Subsequently, the focus is narrowed to the European market design (Section 1.1.3), the costs of congestion management (Section 1.1.4), and how electricity markets with congestion management can be modeled to explore the interactions and evaluate different design choices (Section 1.1.5).

1.1.1 Challenges for the electricity grids

The traditional approach to electricity production, transmission, and distribution involved generating electricity at large, centralized power plants located near population centers. With the shift towards renewable energy sources such as wind and solar, this model is changing to one in which electricity is generated by smaller, distributed units like wind turbines and solar panels. Additionally, sector coupling significantly increases electricity demand. As a result, the expansion of renewable energy must not only replace the energy previously provided by conventional power plants, but also meet the rising electricity demand.

For the electricity grids, this poses challenges at different levels. Electricity grids in Germany operate with different voltages depending on their task (cf. Weber et al. 2022, pp. 136–138). Extra high voltage (EHV) grids (transmission grids, operated by transmission system operators (TSOs)) are used to transport electricity over long distances. Historically, this is where most large power plants fed in their electricity production. From the substations of the EHV grid, high voltage (HV) grids further distributed the electricity. Medium voltage (MV) grids distributed the energy into residential streets where individual homes receive low voltage (LV). These distribution grids are managed by distribution system operators (DSOs)

The challenges posed to the electricity grid differ by voltage level (cf., e.g., Pearson et al. 2022). The EHV grid was designed for centrally located large power plants. However, the placement of especially wind turbines requires a different structure. The geographic center of wind energy production lies in the north of Germany: The proximity to the sea with high wind speeds, a flat geography, and relatively low population density make it an attractive area for the construction of wind farms. Furthermore, the North Sea and the Baltic Sea will host large offshore wind farms, taking advantage of consistently high wind speeds. However, the load centers mostly lie in the south and west of Germany, thus requiring the transport of large amounts of electricity where it was formerly not required.

The distribution (MV and HV) grids were designed to transport electricity towards the customers. However, they were not designed for the large increase in electricity demand by households, who charge their electric vehicles at home and heat their homes with heat pumps powered by electricity. Additionally, rooftop solar modules feed into the distribution grids, which were not designed to transport electricity *away* from consumers. Wind turbines feeding into distributional grids exacerbate the problem.

Expanding the grid is therefore necessary to meet the changing demands for electricity (E-Bridge et al. 2014; 50Hertz Transmission GmbH et al. 2021). However, this process is time-consuming and costly as it involves large-scale infrastructure projects with significant investment costs. Additionally, it is not practical or efficient to expand the grid to accommodate every single megawatt-hour of energy generated by renewable sources such as wind and solar power plants, especially during infrequent production peaks.[1] As a result, effectively managing congestion on the grid is crucial for its reliable and efficient operation.

1.1.2 Electricity market design and congestion management

There are two main ways to organize electricity spot markets. Nodal pricing, as employed, among others, in large parts of the United States, and zonal pricing, as employed in Europe. These market designs differ in their approach to managing the finite capacity of the electricity grid.

In its general form, nodal pricing assigns a price to every node in the system (Schweppe et al. 1988; Hogan 1992). A node can in general correspond to any point where infeed or withdrawal from the grid is possible but mostly corresponds to power stations or substations in transmission grids. This nodal price then corresponds to the cost of providing an extra megawatt-hour of electricity at that specific location. Nodal prices take into account congestion and may differ from node to node. Power plants submit their bids and operating constraints to a central authority (often called an Independent System Operator (ISO)), which determines the dispatch to minimize costs of production, subject to the constraints of the electricity grid. This ensures a market result that is feasible for the grid, i.e., not subject to further congestion management.

[1] In Germany, for example, grid expansion plans are allowed to assume that up to 3% of yearly infeed from wind onshore and solar will be curtailed (§11 EnWG).

Zonal pricing, on the other hand, is organized on a more aggregate level (Bjørndal and Jørnsten 2001). The electricity grid is partitioned into zones, also called price zones or bidding zones. Within each bidding zone, there is no constraint on trading imposed by the grid and prices do not vary between nodes within a bidding zone. Instead, a uniform price is determined for each bidding zone. Trade between bidding zones is possible, subject to (approximated) grid constraints. This system does not ensure a market result feasible for the grid, therefore requiring further congestion management after market clearing to ensure safe grid operation. This is called redispatch, referring to the re-dispatching of generators to remove overloads on grid elements.

1.1.3 The European electricity market design

Since the liberalization of the energy sector, European spot electricity markets have continued to evolve towards a tighter coupling of the formerly independent markets in each country, which now form bidding zones in a zonal market design. These bidding zones are coupled in line with EU regulation *Guideline on Capacity Allocation and Congestion Management* (CACM) (European Comission 2015).[2] The most important wholesale market for electricity is the day-ahead spot market, and it is this market that is the subject of all analyses regarding zonal markets within this thesis.[3] European day-ahead spot markets are subject to the *Single Day-ahead Coupling* (SDAC). In the SDAC, Nominated Electricity Market Operators (NEMOs) – i.e., power exchanges – collect individual bids by market participants for power generation and withdrawal. The bids are then matched across NEMOs by the Price Coupling of Regions (PCR) algorithm called *Euphemia*. In this matching process, cross-border capacities for electricity trade are implicitly allocated in a way that maximizes welfare (NEMO Committee 2020; ENTSO-E 2023). In large parts of the EU, this implicit allocation of cross-border capacities is performed via the so-called *Flow-Based Market Coupling* (FBMC), which has been in effect since 2015. First introduced for the area of Central Western Europe (CWE), comprised of Germany, Austria, France, the Netherlands,

[2] While the bidding zones mostly correspond to country borders, this configuration is not mandated by the CACM guideline and is subject to review.

[3] The term *electricity markets* in the context of this thesis is used synonymously with wholesale electricity spot markets.

Belgium, and Luxembourg, FBMC has since been expanded to include Central Eastern Europe (CEE), i.e., Poland, the Czech Republic, Slovenia, Slovakia, Hungary, Romania, and Croatia. CWE and CEE form the CORE capacity calculation region since June 2022 (JAO 2022).

Bidding zones within FBMC coincide with the borders of the participating countries. The only exceptions are Germany and Luxembourg, which together form a common bidding zone. This zone also included Austria before a bidding zone split in 2018. In the market clearing with FBMC, the individual bids by market participants are collected and matched in a two-sided multi-unit auction to maximize total welfare, subject to a set of constraints that approximate the physical transmission grid[4]. As FBMC is part of a zonal market design, the detailed nodal power flow equations are aggregated to the zonal level. To perform this aggregation, certain assumptions about the state of the power system are required. Firstly, a pre-selection of relevant transmission lines and transformers is made, resulting in a set of so-called Critical Network Elements and Contingencies (CNECs).[5] Secondly, each CNEC has an associated Remaining Available Margin (RAM). This is the part of the thermal capacity of the network element that is available for cross-border trade. Thirdly, zonal Power Transfer Distribution Factors (PTDFs) reflect the change in power flow on a particular network element induced by a change of exports from a specific zone. The zonal PTDFs are approximated from precise nodal PTDFs with the help of Generation Shift Keys (GSKs). GSKs describe changes in generation within a bidding zone associated with changes in the net trade balance of the zone. The selection of CNECs with their associated PTDFs and RAMs and the used GSKs can be summarized as *flow-based parameters*. The flow-based parameters provide approximations of the physical constraints of the grid and are calculated by TSOs based on their expectation of the eventual market result. In summary, the market clearing algorithm maximizes welfare subject to the constraints stating that the power flows over CNECs induced by the net position of each bidding zone do not exceed the RAMs of the respective CNECs.

The rationale behind limiting commercial transactions is to ensure the feasibility of the market outcome with respect to the physical constraints

[4] For what follows cf. Schönheit et al. (2021b)

[5] A CNEC refers to the combination of a critical network element (CNE) with a contingency, e.g., the failure of a different transmission line. The associated parameters in this case reflect a grid topology where this outage has occurred.

of the transmission grid. There are, however, several factors that result in a remaining need for ex-post adjustments, i.e., redispatch (Felten et al. 2021). FBMC, by design, only *approximates* the physical constraints of the transmission grid. The zonal PTDFs represent the effect that the total net position of each prize zone has on any given CNEC, regardless of where exactly within each zone the electricity is generated or consumed. The GSKs, i.e., the translation of changes in the zonal net position to individual nodes, are determined by the TSOs beforehand and used to calculate the zonal PTDFs. Furthermore, any zonal market design requires the bidding zones themselves to be internally free of congestion to function optimally, because internal congestion cannot be adequately considered within zonal market coupling. This requires appropriately delimited bidding zones, a requirement that is not fulfilled: As stated above, the bidding zones within the FBMC correspond to countries. But with grid expansion lagging, especially Germany features a considerable amount of internal congestion. Adjusting bidding zones to better reflect structural congestion, e.g., splitting the German bidding zone, would address the problem but faces significant political opposition and, like a move to nodal pricing, has to date not been a politically feasible option (Eicke and Schittekatte 2022; Felling et al. 2023). Consequently, redispatch is likely to play an important part also in the future operation of FBMC.

1.1.4 The costs of congestion management

Congestions in the electricity grid are costly. As far as grid constraints are considered in the market, they may force an outcome where expensive generators must be utilized instead of cheaper alternatives because of the location of the generators. This increases the total costs of production, decreasing overall welfare. Because this happens as part of the market process, however, additional costs are directly and implicitly bourne by the market participants themselves: Cheaper generators do not get to produce (and do not earn a contribution margin), load assets are subject to higher or lower prices, depending on their location in the grid. Both nodal pricing and zonal pricing generally exhibit these effects, but only a part of congestion is internalized in zonal markets.

Overloads that would result from transactions in the zonal market must be prevented by using redispatch, i.e., (mostly ex-post) congestion management. Two types of redispatch may be distinguished, namely cost-based and market-based redispatch. (Weber et al. 2022, pp. 336–337). In cost-based redispatch, TSOs directly request changes in the planned dispatch of generation units. The units used for redispatch are either compensated if they must increase their production or are required to reimburse the saved operational costs if they must decrease their production.[6] In market-based redispatch, generation and demand units submit bids for either increases or decreases in generation or demand.[7] In both methods, the additional costs for increasing generation in one location always outweigh the savings for decreasing generation in another location, otherwise this changed schedule would have already been a result of the original market procedure (as it would lead to an overall cheaper outcome). The additional costs incurred in this way are not directly borne by any market party, and must therefore be recovered by some dedicated refinancing mechanism. In Germany, the costs for redispatch (including curtailment of renewables) are projected to have reached 2.3 billion Euros in 2021 (Bundesnetzagentur 2022). These costs incurred by the TSOs are reimbursed by being added to the network fees, which are part of the electricity price paid by end users.

The specific network fees for each end user depend on the region. The German transmission grid is divided between four TSOs, but their network fees have been aligned since the beginning of 2023. However, at the end of 2021, over 840 different grid operators were responsible for the German distribution grids, each of them setting their own specific network fees (Bundesnetzagentur and Bundeskartellamt 2022). As the number of congestions in distribution grids is increasing, fairly attributing congestion costs to individual grid operators (and ultimately the end users connected to their grid) continues to become more important. Furthermore, the mechanism by which costs are allocated can itself set incentives that adversely impact the efficiency of redispatch. Until 2021, costs for curtailment of renewables in Germany were incurred by the system operator that initiated the curtailment, rather than by all the system operators that benefit from it (§14 and §15 EEG 2017). This incentivizes system operators to wait as long as possible for other system operators to take

[6] Increases and decreases in generation must always be balanced to maintain the overall equilibrium of supply and demand.

[7] Both cost-based and market-based redispatch exhibit their own set of problems, cf. Hirth et al. (2019)

measures, which may prevent cooperation – thus further underlining the need for an appropriate mechanism to allocate the costs.

1.1.5 Modeling electricity markets and congestion management

To investigate, understand, and assess the interactions of different design options for electricity markets and congestion management, modeling is an invaluable tool that is used extensively in this thesis. The field of modeling energy systems is diverse. Pfenninger et al. (2014) differentiate between *energy system optimization models*, *energy system simulation models*, *power systems and electricity market models*, and *qualitative and mixed-methods scenarios*. Energy system optimization and simulation models feature a large scope, not only focusing on electricity but also considering for example the supply and markets for input factors such as fuel and CO_2 emission certificates. These models can be used to predict the future evolution of the energy system – or provide the means analyze how best to shape it. On the other hand, power systems and electricity market models feature a narrower scope and focus on electricity markets while following a bottom-up approach to model the demand and supply of electricity. These models have a high temporal and spatial resolution and allow for the detailed analysis of the interdependencies of, for example, market design choices and a changing generation pattern. Within these models, a high level of detail for the technical parameters of individual power plants as well as time series data for the infeed of renewable energy sources and general demand is used to approximate the results that would be obtained in the market they are employed to model. These results include electricity prices, the dispatch of power plants, and by extension derived outcomes such as greenhouse gas emissions.

Ventosa et al. (2005) classify electricity market models as either single-firm optimization models or equilibrium models, where the optimization problems (for profit maximization) of individual firms are coupled to arrive at a global equilibrium where supply equals demand. As they note, a perfectly competitive market can be modeled as a cost-minimization or net benefit maximization problem. In this case, the individual optimization problems of the individual firms are replaced by a single optimization problem, where the bids reflect the marginal costs of the individual firms. This approach corresponds to the category of power systems and electricity market models in Pfenninger et al. (2014) and it is this approach that all models used in this thesis follow.

Modeling in the context of this thesis refers to the simulation of the electricity markets to obtain results close to those that would be obtained in practice given the underlying assumptions. As described in Section 1.1.3, in much of Europe, the Euphemia algorithm couples the individual markets by matching the bids for power generation and demand in a two-sided multi-unit auction to maximize welfare. In the case of FBMC, the capacity for cross-border trade is implicitly allocated by considering approximated grid constraints in the optimization problem. This can be modeled as described by Ventosa et al. (2005) by solving the optimization problems of the individual firms against the background of an overall equilibrium. However, this approach can be simplified when assuming perfect competition – an assumption that may be justified by the fact that the first aim of the CACM regulation is "promoting effective competition in the generation, trading and supply of electricity" (European Comission 2015, Article 3).

As noted, a perfectly competitive electricity market can be expressed as a single optimization problem. In the case of Euphemia, the objective function maximizes total welfare, which suggests that a similar approach may be taken in modeling. However, for practical purposes, electricity demand is often assumed to be price inelastic.[8] In that case, the welfare maximization problem corresponds to a cost minimization problem, where the system costs, consisting mainly of marginal costs of electricity-producing power plants, are minimized (Weber et al. 2022, p. 242).

Congestion management can be incorporated into such an optimization problem by including constraints that model the underlying electricity grid. In the practical implementation of FBMC as discussed in Section 1.1.3, the flow-based parameters are added to the optimization problem in Euphemia. Considering these constraints in the electricity market model gives the *zonal problem*. In the case of nodal pricing, the zonal constraints are replaced and the model is extended by adding the power flow equations as constraints, resulting in an optimal power flow (OPF) calculation (cf., e.g., Zimmerman et al. 2011), also called the *nodal problem*.

[8] Elastic demand can be converted to inelastic demand by modeling the elastic portion as negative generation.

Since the nodal problem considers the relevant grid constraints in detail instead of the zonal approximations and in addition to the other constraints already considered in the zonal problem (such as technical parameters of power plants), its solution is the cheapest way to satisfy demand given the grid constraints in addition to the existing non-grid-related constraints of the zonal problem.[9] It is for this reason that nodal pricing is commonly used in the scientific modeling literature as the benchmark to compare design alternatives against (e.g., in Bjørndal and Jørnsten (2001), Bjørndal and Jørnsten (2007), Oggioni and Smeers (2013), and Felling et al. (2023)).

The nodal problem considers detailed grid constraints, ensuring that the result is feasible for the grid. Besides modeling nodal pricing, this property is also used to model redispatch – the adjustment of the zonal result to ensure feasibility for the grid. Conceptually, redispatch can be modeled by using the zonal solution as the starting point for the nodal problem (which also minimizes system costs), where it will be adjusted to conform to the nodal grid constraints. However, this approach will always lead to a solution that is equivalent in costs to that which would be reached by nodal pricing, as it is the cheapest way to satisfy all constraints (as explained above).[10] This outcome conflicts with the objective behind organizing markets zonally – to let the market determine the optimal dispatch and only adjust it where necessary using redispatch. Additionally, executing redispatch after and outside of the day-ahead market, but before the closure of other marketing opportunities such as the intraday market leads to opportunity costs and other additional costs, such as the cost for alternative heat supply – a fact that was also recognized by German courts, who allowed these additional costs to be reflected in payments under the cost-based redispatch regime (OLG Düsseldorf 2015; Felling et al. 2023). To reflect the increased costs and the objective of the TSOs to minimize redispatch volume, the objective function of the redispatch problem is hence modified by including penalties.[11]

[9] For limitations of this assumption see Chapter 7.3.

[10] There may be multiple solutions with the same costs. Consequently, different starting values may lead to different solutions. In this case, the redispatch solution may not be equal to the nodal solution arrived at from a different starting point, but it will nevertheless be nodally optimal.

[11] See also Chapter 3 for more details.

Models that work according to the principles described in this section are used in this thesis to contribute to answering multiple research questions that concern electricity markets and congestion management. These research questions are formulated in the following section.

1.2 Research Questions

This thesis consists of five articles that address multiple research questions in the context of electricity markets and congestion management. These research questions can be divided into methodological questions and substantive questions. The methodological questions are answered by introducing new methodologies that may be used to subsequently answer substantive questions.

1.2.1 Methodological Questions

As indicated in Section 1.1.3, there are multiple stages to Flow-Based Market Coupling, the mechanism for implicitly allocating cross-border transmission capacity in many of Europe's coupled electricity markets. To analyze the effects of policies and market design on electricity markets and congestion management and to allow for consistent comparisons between the results, markets and congestion management must be modeled consistently and in an integrated approach. As described in Section 1.1.5, optimization models are frequently used to model electricity markets. Specifically, a subgroup called linear programs may be used. This raises the first research question:

1. How can different market designs be modeled with linear programming and what are the strengths and weaknesses?

A crucial part of operating electricity grids is congestion management (cf. Sections 1.1.3 and 1.1.4). Flow-Based Market Coupling, with its approximation of grid constraints between bidding zones, aims to incorporate a part of congestion management into the electricity market. Still, redispatch is required to avoid overloads of the grid infrastructure. This results in significant extra costs, which are external to the electricity market and must therefore be allocated ex-post. The German distribution grids alone are operated by over 800 grid operators (cf. Section 1.1.4). In this context, the present thesis addresses the following question:

2. How can the use of Shapley values and linear programming contribute to achieving a fair allocation of congestion costs?

1.2.2 Substantive Questions

The fundamental difference between zonal pricing as employed in Europe and nodal pricing as employed in the United States can be traced back to the trade-off between (1) enabling as much free and frictionless trading as possible, but externalizing the task to ensure the market result is physically executable within stable grid operation (zonal pricing) and (2) completely internalizing congestion management, thereby fragmenting the electricity market to a considerable extent.[12] The framework developed in answering the first question raised above provides the means to analyze some of the implications of this trade-off, formulated in the following third research question:

3. What are the welfare implications of the trade-off between increasing trading possibilities and accounting for grid constraints?

Finally, in light of current and future developments and challenges – such as the rising share of renewables – an efficient operation of the electricity market must be ensured. This is addressed in the fourth and final research question:

4. What contributes to the efficient operation of zonal electricity markets with congestion management?

1.3 Structure of the thesis

The remainder of this thesis consists of the five articles forming the core of this thesis followed by a concluding chapter. Each article corresponds to a chapter and is laid out as published. In the final Chapter 7, the focus will return to the research questions raised in Section 1.2. There, the contributions of this thesis to the research questions will be summarized.

[12] At the extremes this corresponds to the fully unbundled and fully integrated systems described by Wilson (2002).

Market distortions in flexibility markets caused by renewable subsidies – The case for side payments
by Jonas Höckner, Simon Voswinkel, and Christoph Weber
published in 2020 in Energy Policy 137, corresponding to Chapter 2

Given that extensive renewable expansion is required for the energy transition to succeed, combined with the fact that congestion management is required for stable operation of the power grid under these circumstances, efficiently integrating renewables into congestion management is essential. This is complicated by the subsidies that are employed to facilitate investments of renewables, because, as laid out in the article, these subsidies distort the market results if a market-based approach is used for redispatch. The article investigates this problem analytically by first demonstrating the market distortions that would occur without further action, and subsequently introducing the concept of "side payments". These side payments are used to eliminate the distortive effects of subsidies by externalizing them to monetary flows outside of the market.

Flow-Based Market Coupling – What Drives Welfare in Europe's Electricity Market Design?
by Simon Voswinkel, Björn Felten, Tim Felling, and Christoph Weber
HEMF Working Paper No. 08/2019, corresponding to Chapter 3

The following two chapters deal with market design and congestion management more generally. First, Chapter 3 introduces an integrated model spanning all stages of Flow-Based Market Coupling as well as modeling nodal market clearing. The model is applied to the area of Central Western Europe and investigates the effect of the chosen GSK method, the effect of an imperfect delimitation of bidding zones, and the effect of measures aimed to increase cross-zonal trade, such as minRAM and the removal of internal CNECs from the flow-based parameters.

Improving flow-based market coupling by integrating redispatch potential – Evidence from a large-scale model
by Michael Bucksteeg, Simon Voswinkel, and Gerald Blumberg
submitted to Energy Policy in May of 2023, corresponding to Chapter 4

Chapter 4 extends the model from Chapter 3 by implementing so-called *redispatch potential* (RDpot) into the zonal market clearing. The concept of RDpot aims to increase the capacities available for cross-zonal trade by incorporating the knowledge about available redispatch measures into the zonal market clearing, allowing higher exchanges to take place if generators are available to return the underlying dispatch to a state feasible for grid operation. The article investigates multiple methods for determining potential redispatch measures that should be made available to the zonal market and analyzes the effects on system costs, redispatch volumes, zonal exchanges, and other important metrics.

Sharing congestion management costs among system operators using the Shapley value
by Simon Voswinkel, Jonas Höckner, Abuzar Khalid, and Christoph Weber
published in 2022 in Applied Energy 317, corresponding to Chapter 5

The final two articles deal with the problem of allocating the costs that arise from redispatch. First, Chapter 5 investigates the Shapley value, a concept known from game theory, as a method to allocate congestion management costs to congested grid elements. Its main contribution is the development of two methods that simplify the calculation of the Shapley value in this context – a necessary step towards the practical use of the Shapley value as it is very computationally expensive to calculate. The effectiveness of these methods is demonstrated on a small benchmark grid for one grid load case.

Simplifying the computation of Shapley values for allocating congestion costs in large power grid models
by Simon Voswinkel
submitted to Applied Energy in May of 2023, corresponding to Chapter 6

Finally, Chapter 6 builds on the previous contribution in Chapter 5 by applying the more effective method to a realistic grid dataset, the transmission grid of Germany, consisting of over 700 nodes and over 1300 transmission lines and transformers. The method is applied to a year of hourly data. Within the 8784 hourly snapshots, the number of potential overloads after the zonal clearing ranges from zero up to 22. The algorithm decreases the time to calculate the Shapley values for all time steps from weeks to hours by saving 98.68 % of the calculations required otherwise. The results are furthermore analyzed statistically to identify the drivers of the observed computational benefits across the 8784 hourly snapshots. The results of the statistical analysis show that the savings achievable with the simplification algorithm are mainly driven by the number of overloaded grid elements after the zonal clearing together with the number of elements that are still fully utilized after performing redispatch.

Chapter 2

Market distortions in flexibility markets caused by renewable subsidies – The case for side payments

by Jonas Höckner, Simon Voswinkel, and Christoph Weber
published in 2020 in Energy Policy 137

Energy Policy 137 (2020) 111135

Contents lists available at ScienceDirect

Energy Policy

journal homepage: http://www.elsevier.com/locate/enpol

Market distortions in flexibility markets caused by renewable subsidies – The case for side payments☆

Jonas Höckner*, Simon Voswinkel, Christoph Weber

House of Energy Markets and Finance University of Duisburg-Essen, Universitätsstr. 12, 45117, Essen, Germany

A R T I C L E I N F O

JEL classification:
Q41
Q48

Keywords:
Congestion management
Smart markets
Flexibility markets
Side payments
Renewables

A B S T R A C T

Strongly increasing costs of congestion management have provoked a discussion in Europe about new approaches to solve grid congestions in a more efficient way. One approach is to design flexibility markets. In this paper we focus on the effects of subsidies for renewable energy on the market outcome of a flexibility market. We show that subsidies can cause market distortions and lead to an inefficient selection of flexibility options to solve grid congestions. We propose the implementation of side payments together with price caps and uniform pricing to achieve an efficient market design. Ultimately choosing between flexibility markets with and without side payments involves a tradeoff between minimizing system costs and maximizing renewable infeed. Our analysis provides the framework for a conscious political choice on that subject.

1. Introduction

Due to a rapid increase of distributed renewable resources and delayed network expansions, the number of network congestions and the congestion management costs have increased drastically over the last years. This particularly applies to Germany (see Fig. 1), where high subsidies and a priority feed-in for renewable energy have been stipulated in the renewable energy law (EEG) which have led to a significant increase in renewable generation capacities. The German transmission grid is not yet capable of transporting huge amounts of wind power from the North to the load centers in the South. Additionally, high shares of renewables are connected to the distribution grid, which causes critical backflows to the transmission grid in periods of high renewable infeed.

Given the ambitious goals of extending renewable capacities, which will be primarily connected to the distribution grid, congestions on lower grid levels are expected to increase significantly. Thus, it is important to find new methods of congestion management that grid operators on all grid levels can apply.

A recently discussed approach are regional flexibility markets on which system operators can procure flexibility to alleviate grid congestions. In this paper we analyze the effects of subsidies for renewable energy sources on the outcome of such regional flexibility markets, by modeling the opportunity costs of participating assets. We show that subsidies can cause market distortions and lead to an inefficient selection of flexibility options. We also discuss how external payments may solve this problem. The discussion is rooted in the specific German context in order to avoid lengthy discussions of the broad variety of regulations in place, yet the results obtained can also be transferred to other legislations with renewable support schemes, zonal markets and congestion management issues in the distribution grid.

The paper is organized as follows: In section 2, we give a brief literature overview of congestion management approaches in transmission and distribution grids. Section 3 describes the assumptions underlying the model of this paper. Opportunity costs of flexibility are modelled in section 4 with an emphasis on variable renewable energy sources. Within this framework, distortions caused by EEG subsidies and the resolution of these distortions by implementing side payments are discussed in section 5. An illustrative example is given in section 6, which demonstrates the problem of market distortions caused by EEG subsidies and how side payments can prevent these distortions. In section 7, we discuss the results of the illustrative example and the assumptions underlying the model. Section 8 concludes and practical recommendations for future regulation are derived.

☆ The authors are solely responsible for the contents which do not necessarily represent the opinion of the House of Energy Markets and Finance nor the views of the entire enera consortium.

* Corresponding author.
E-mail addresses: jonas.hoeckner@uni-due.de (J. Höckner), simon.voswinkel@uni-due.de (S. Voswinkel), christoph.weber@uni-due.de (C. Weber).
URL: http://www.hemf.net (J. Höckner).

https://doi.org/10.1016/j.enpol.2019.111135
Received 1 May 2019; Received in revised form 19 November 2019; Accepted 20 November 2019
Available online 4 December 2019

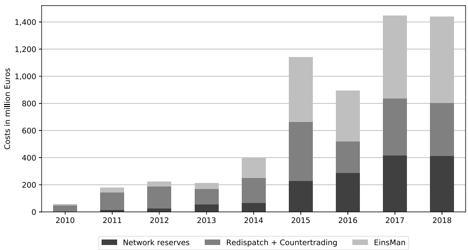


Fig. 1. Congestion management costs in Germany (bdew, 2018; Bundesnetzagentur, 2019).

2. Literature review

The design of electricity markets has been subject to many discussions over the last decades. Different approaches to handle grid congestions in transmission grids caused by limited line capacities in particular provoked debates between academics (Weibelzahl, 2017). The most frequently compared concepts are nodal pricing, as it is prevalent in the United States, and zonal pricing, which underlies the European market model.

Many experts assess nodal pricing (also referred to as locational marginal pricing (LMP)) as the optimal solution, because transmission constraints are directly reflected in the market outcome by differentiated prices at the various grid nodes (Bjørndal and Jørnsten, 2001; Ehrenmann and Smeers, 2005; Hogan, 1992; Neuhoff et al., 2013; Schweppe et al., 1988).

In contrast, the European electricity markets are based on a zonal market model, where there are one or at most a few price zones per country. In a zonal market, intrazonal congestions cannot be addressed efficiently by market splitting and grid operators need alternative congestion management methods like redispatch or countertrading to prevent congestions (Bjørndal and Jørnsten, 2007; Vries and Hakvoort, 2002). Several publications have analyzed the market results of zonal market models and compared them to a nodal market design (Bjørndal et al., 2003, 2018; Bjørndal and Jørnsten, 2007, 2001; Ehrenmann and Smeers, 2005). Different forms of congestion management in a zonal market design have been discussed as well (Holmberg and Lazarczyk, 2015; Oggioni and Smeers, 2013). Bjørndal et al. (2018) discuss a hybrid congestion management model, where a country with nodal pricing is embedded in a zonal pricing system.

Most recently, Weibelzahl and Märtz (2018) analyzed the effects of storage facilities on electricity prices under different congestion management methods including nodal and zonal pricing. Sarfati et al. (2019) compare the efficiency of nodal and zonal pricing assuming imperfectly competitive markets.

Whereas congestion management on the level of transmission grids has been discussed for decades now, congestion management in distribution grids moved to the center of the debates rather over the last years, primarily driven by the increasing share of distributed energy resources. For DSOs, short-term congestion management usually comprises three different approaches: reconfiguration of the distribution system, direct load control and market-based mechanisms (Fotouhi Ghazvini et al., 2019; Liu et al., 2014). Market-based mechanisms are thereby assumed to maximize the social welfare by harnessing the benefits of demand-side flexibility (Fotouhi Ghazvini et al., 2019).

Most approaches of implementing market-based congestion management in distribution grids are based on locational price signals. A common concept is to implement distributional locational marginal prices (DLMP) inspired by nodal pricing in transmission grids (Fotouhi Ghazvini et al., 2019; Heydt et al., 2012; Huang et al., 2015; Sotkiewicz and Vignolo, 2006).

Yet besides concepts based on locational price signals, various other approaches have been also developed to address congestion management and the use of flexibilities in the distribution grid. This broad range of concepts is frequently discussed under the name “smart markets”, cf. e.g. in Germany (Bundesnetzagentur, 2011; Ecofys and Fraunhofer IWES, 2017).[1] According to Ecofys and Fraunhofer IWES (2017), two broad categories of smart markets may be distinguished (cf. Fig. 2): on the one hand those where flexibility is procured by the grid operator and on the other hand approaches where quotas of available grid capacities are allocated to grid users and possibly traded by them on a secondary market.

The project “Proaktives Verteilnetz” (proactive distribution grid) aims to demonstrate a quota based smart market (Wiedemann, 2017). Applying new approaches for distribution system state estimation, the grid operator determines an individual and non-discriminatory power range per retail company. Thereupon each retail company can decide individually how to fulfill these restrictions most efficiently. The concept of a secondary trading platform for these flexibility calls is evaluated as well.

Another approach to utilize flexibilities for congestion management in distribution grids is to implement local flexibility markets or platforms to coordinate the flexibility demand of the system operator and the existing flexibility providers. Several approaches of those flexibility markets have been designed and partly implemented as part of different research projects. A literature review of international research concerning flexibility markets is provided by Villar et al. (2018).

As examples of markets where flexibility is procured explicitly, the German E-Energy projects[2] implemented manifold market platform solutions with the intention to integrate new decentral players to an innovative market setting (Karg et al., 2014). While “eTelligence” designed regional products that could be utilized by grid operators as ancillary services, the project “E-DeMa” developed a market to acquire new demand side flexibilities (Agsten et al., 2013; Koch et al., 2012). Furthermore, “MeRegio” designed a hybrid model which enables grid operators to tender the elimination of a certain grid congestion on a market platform (Karg et al., 2014).

Apel et al. (2014) conceptually designed a flexibility market “RegioFlex” that enables system operators to access flexibilities in the system to prevent congestions. Currently, several projects of the SINTEG[3] program further develop and demonstrate regional flexibility platforms in order to test new approaches to ensure secure grid operation with high shares of intermittent power generation (BMWi, 2016). The showcase enera, for example, implements a regional flexibility platform based on local order books in cooperation with EPEX Spot in order to set up regional ancillary services to manage grid congestions.

A further approach is the concept of a flexibility clearing house, which acts as a third party to provide a platform to coordinate flexibility

[1] These markets can in principle be utilized by transmission grid operators as well and extend existing transmission grid services. The concept is in accordance with the strongly discussed proposal by the European Commission to use market-based mechanisms to resolve grid congestions (cf. Article 12 in European Commission (2016); Hirth and Schlecht (2018)).

[2] The E-Energy projects were funded by the Federal Ministry of Economics and Technology (BMWi) and Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) and comprised six flagship initiatives (eTelligence, E-DeMa, MeRegio, moma, RegModHarz and Smart Watts), which focused on the ICT-based energy system of the future.

[3] “The Smart Energy Showcases - Digital Agenda for the Energy Transition” (SINTEG) is funded by the Federal Ministry of Economics and Technology (BMWi) of Germany and aims to set up large-scale showcase regions for developing and demonstrating model solutions that can deliver a secure, efficient and environmentally compatible energy supply with electricity being generated to a large extent from volatile sources such as wind or solar. The showcase projects are enera, WindNODE, C/sells, Designnetz and NEW 4.0.

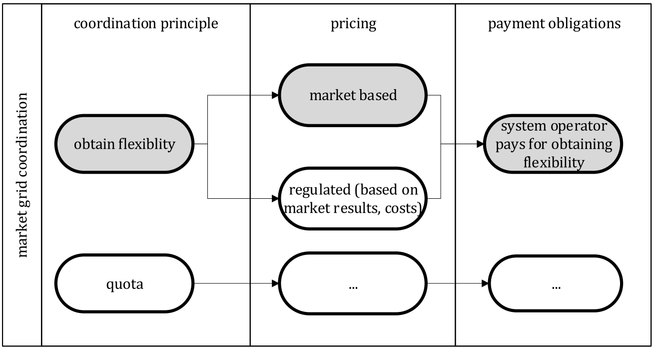


Fig. 2. Key categories and potential specifications of smart markets based on Ecofys and Fraunhofer IWES, 2017, p. 24.

demand and supply (Heussen et al., 2013; Zhang et al., 2013). The flexibility market relies on aggregators that pool flexibility options and provide standardized ancillary services to the DSO. On the market several standardized products can be traded and the main goal is to reduce transaction costs by coordinating DSOs and flexibility providers (Heussen et al., 2013). Zhang et al. (2013) introduce three possible trading setups, i.e. bilateral contracts, auctions and supermarkets,[4] which are evaluated as part of the Danish national project iPower. Torbaghan et al. (2016) introduce a local market framework that allows a local market operator to adjust the energy program of its specific market area before they are forwarded to the wholesale market. To achieve that, the local flexibility market acts as an intermediary on which the local market operator can efficiently select flexibilities that are offered by aggregators located in its market area.

Our paper contributes to the existing literature on congestion management in transmission and distribution grids by examining regional flexibility markets that can be applied by system operators to perform congestion management more efficiently. To achieve that, we study regional flexibility markets that are open to all types of technologies, i.e. demand-side and generation-side flexibilities. Generation-side flexibilities can comprise the curtailment of renewable power plants, which then compete with demand-side flexibilities on the flexibility market. Additional remunerations by renewable support schemes, representing opportunity costs of renewable flexibilities, play a decisive role on the resulting market outcome.

This is the key research focus of the present paper. We examine the distorting effects of renewable support schemes on the outcome of flexibility markets based on a simple stylized model and derive recommendations in order to prevent such market distortions. This complements the research strand of congestion management in zonal markets by improving congestion management methods to resolve intrazonal congestions. Additionally, we contribute to the literature on flexibility markets as in Villar et al. (2018) by deriving explicit market design recommendations for regional flexibility markets.

3. Market characteristics, bidding behavior and regulatory setting

In order to show the effects of subsidies on the outcome of a flexibility market, assumptions about the characteristics of the market, the bidding behavior of participating units and the regulatory and institutional settings are necessary. These assumptions are detailed in the following sections.

3.1. Market characteristics

We assume a regional flexibility market on which grid operators buy the flexibility they need to manage congestions from market participants, who in turn adjust the dispatch of their assets to conform to the trades they make on the market. Participation is open to any asset already able to participate in spot markets but is not mandatory. That also includes all variable renewable energy sources (v-RES) under a so-called direct marketing regime (cf. section 3.3.3), irrespective of size. Further, every participant can set bid offers freely, with the system allowing free price formation as it is recommended by Ecofys and Fraunhofer IWES (2017) for wind dominated network areas. Therefore, assets are in competition with each other. The grid operator chooses the best options among the offers based on the merit-order. For our analysis, we assume a “naïve” pay-as-bid mechanism, where flexibilities bid their opportunity costs and are remunerated according to their bid. This assumption is discussed in detail in section 7.2.1.

We define flexibility as the deviation of actual power infeed (or withdrawal) from a previously planned baseline. Participants sell their willingness to deviate from the previously planned baseline to the grid operator and in turn must then adjust the actual infeed or withdrawal from their assets. Additionally, participants whose bids are accepted need to buy or sell the energy on the spot market to keep the system balanced (cf. Fig. 3).

In terms of interactions with other markets, we assume that the regional flexibility market takes place in parallel to the intraday market. As participation in the flexibility market is voluntary, participants are free to pursue other options. One such option is participation in the reserve power market, where auctions happen on the previous day. Participants must only ensure that the flexibility they sell on the flexibility market has not already been marketed on the reserve power market, as coinciding adjustments to deliver balancing energy do not count as flexibility.

3.2. Bidding behaviour

In any market, a market participant who has no obligation to participate will only conclude a trade if its (expected) profits exceed profits obtained under other marketing alternatives. Hence, any flexibility provider in a flexibility market will base its bid into the market on opportunity costs, i.e. foregone profits from alternative operation and trading strategies. We assume that market participants bid these

[4] A supermarket means that the aggregator offers different ancillary services based on its portfolio and the DSO can select the ancillary services it needs.

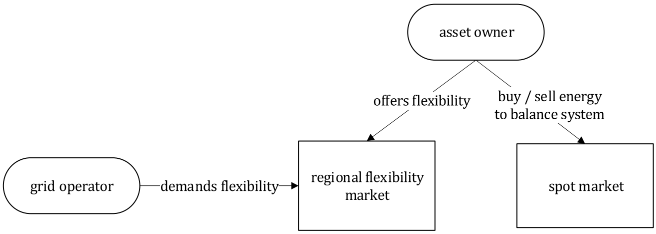


Fig. 3. Interaction between grid operator, asset owner and flexibility/spot markets.

opportunity costs when offering flexibility (see section 4 for a detailed derivation of opportunity costs).

Dealing with v-RES implies handling uncertainties. The generation from v-RES is characterized by two types of uncertainty: (1) There is uncertainty about the volume of energy which will be produced at any given time (and consequently the volume of flexibility that can be offered) and (2) there is uncertainty about the exact values of the components included in the subsidies (notably the so-called market premium). For our analysis we assume that the first type of uncertainty does not have an impact on the value of the bid itself, but only on the volume that is offered. It follows that in order to bid in the market, the participants must somehow determine the volume to offer. This process is not part of our analysis, we only assume that it has taken place. The second type of uncertainty would alter the bids themselves. Here we assume perfect foresight, but discuss this assumption, its consequences and possible remedies in section 7.2.2.

As detailed in section 3.1, participants are free to act in markets other than the flexibility market, as long as this does not interfere with delivery of sold flexibility. The only necessary assumption about activity on other markets is that the volumes offered in the flexibility market have not been sold elsewhere.[5]

3.3. Regulatory and institutional setting

Opportunity costs of v-RES plants under the EEG regime are determined by a multitude of factors. The following sections detail the regulatory framework concerning the traffic light concept, feed-in tariffs and the contracts governing the relationship between asset owner and direct marketers.

3.3.1. The traffic light concept

To describe the complex interaction between market and grid sphere, a traffic light concept has been proposed in Germany (bdew, 2013; Bundesnetzagentur, 2011). The green phase designates situations without binding grid constraints, so that no market interventions by the grid operator are required. In contrast, the red phase indicates an actual congestion which requires direct action by the grid operator. German legislation enables the transmission grid operators to take these coercive measures and request the modification of conventional power plant schedules, that are initially based on the unrestricted market results. When these measures are exhausted and the congestion could not be averted, grid operators can curtail renewable infeed using Einspeisemanagement (EinsMan) as an ultimate measure. In Germany, compared to the set of measures that are available to transmission system operators (TSO), the distribution system operator's (DSO) options to manage congestions are limited. Besides topological measures, DSOs are basically restricted to EinsMan.[6] The idea behind the yellow traffic light is to enable new market-based approaches like regional flexibility markets to efficiently prevent predicted congestions.

3.3.2. Feed-in tariffs and remuneration by the grid operator

Most v-RES plants in Germany are subsidized in the EEG system. As of January 1st, 2016, all newly installed v-RES plants above 100 kW nameplate capacity in the EEG system have to sell their energy directly on the spot market (§19 EEG, so-called "direct marketing") as opposed to earlier systems, where marketing of the energy was done by the responsible transmission system operator. To market the energy on the spot market, aggregators can be employed to minimize transaction costs for the individual asset owners. As it stands, only directly marketed v-RES plants with or without subsidies in the EEG system can participate in a regional flexibility market of any kind. As by far the biggest part of v-RES plants in Germany are subsidized in the EEG system, we will only consider v-RES plants that are in direct marketing, remotely controllable, and subsidized in the EEG system. The EEG system guarantees a feed-in tariff determined by the German Renewable Energy Act (§19 EEG in combination with §20 EEG). The guaranteed EEG tariff of a specific asset depends on several factors such as commissioning year and technical specifications.

The remuneration for these power plants consists of two parts: spot market revenues and the so-called market premium (MP), with the market premium intended to fill the gap between the guaranteed EEG remuneration and spot market revenues. The market premium is determined monthly for each asset as the difference between its specific EEG feed-in-tariff and a monthly market value (*MMV*). The MMV is an index calculated as a volume weighted average spot price that is supposed to represent the average spot revenues made by selling the electricity generated by the specific v-RES type (e.g., wind, solar) at the spot market. Both the market premium and the monthly market value are determined ex-post, as market data is necessary for the calculation.[7]

3.3.3. Contracts between direct marketers and asset owners

We assume there is a contract between asset owner and direct marketer, where the asset owner receives remuneration in the height of the monthly market value from the direct marketer. Together with the market premium received from the grid operator, the asset owner thus receives the full feed-in tariff for each produced MWh of electricity. The direct marketer carries the market risk but can also keep all market income exceeding the *MMV*. We further assume that the grid operator only pays the market premium for *actually*, not *potentially* produced electricity; the same holds for the direct marketer who will not pay the asset owner if infeed is curtailed by the grid operator. However, the grid

[5] A power plant may, for example, have sold half of its possible downward adjustment as secondary reserve on the balancing market, which leaves the other half for marketing on the flexibility market.

[6] A new method which allows grid operators to contract controllable loads in the distribution grid has been recently introduced. Yet the formation of the statutory framework currently continues.

[7] For the implications of these uncertainties see section 7.2.2.

operator reimburses the asset owner for these lost revenues in case of EinsMan curtailment. If the direct marketer curtails the power plant outside EinsMan, it must reimburse the asset owner, as the grid operator will not pay the market premium for a voluntarily curtailed power plant.

For the most part in our analysis, we treat the direct marketer and the asset owner as a single entity. The terms of contract between both parties only become relevant when talking about the reimbursements by the grid operator. In this case, it is important to note that we use as working assumption that the grid operator does not have an obligation to reimburse the direct marketer, but only the asset owner.[8] We will point out the difference in the relevant parts of this paper.

4. Modelling the opportunity costs of flexibility

As stated in section 3.2, we assume that flexibility providers bid their opportunity costs when offering flexibility. Especially opportunity costs of variable renewable energy sources are essential to consider, because they represent in principle a large potential of flexibility in the future power system. In periods of excess production, curtailing renewable infeed is always an option – it is obviously not beneficial in terms of CO_2 emissions, but might be more cost efficient than other alternatives. Thereby, the market premium, which is an important part of the governmentally guaranteed remuneration, can induce market distortions and therefore has to be examined in more detail. The next sections will discuss the opportunity costs of different types of assets and in detail those of variable renewable energy sources under the EEG-regime.

4.1. Opportunity costs of demand-side flexibility

For demand-side assets participating in the regional flexibility market, one must assume that care has already been taken to optimize their dispatch on the spot market. Therefore, any action taken to assist the grid operator in solving a potential problem (e.g. a looming congestion), will move the asset away from its optimal dispatch. Opportunity costs of demand-side flexibility are hard to quantify. They often include the costs of electricity at different times, marginal utility of dispatch and operational costs such as personnel.

For the purposes of this paper, we assume that shiftable loads have no additional costs to postpone its withdrawal from the grid, which is in line with Steuer (2017). The only costs that arise are caused by the spread of electricity prices between originally planned and shifted time interval. Power to gas plants, in contrast, have to pay levies and taxes for additional electricity consumption but also obtain the value of gas produced. Additionally, losses of efficiency have an impact on the opportunity costs.

4.2. Opportunity costs of generation-side flexibility

Opportunity costs for generation-side flexibility are more easily determined than for demand-side flexibility. For increasing as well as decreasing dispatch they will be largely determined by fuel costs, offset by the revenue generated when selling (or buying) this additional (or now missing) energy on the spot market.

As with demand-side flexibility in section 4.1, we abstract from actual costs and only assume that some sort of cost will be associated with a modification of their dispatch and hence demanded on the regional flexibility market.

4.3. Opportunity costs of variable renewables under the EEG regime

4.3.1. Unrestricted scheduling (green light)

If no congestions have been identified, the v-RES plant can feed in all generated electricity without restrictions. The direct marketer will receive the revenues (R_{Normal}) from the sale of electricity due to direct marketing at the wholesale market (p_{DA}) and pays the *MMV* to the asset owner. Additionally, the grid operator will pay a market premium to the asset owner for every MWh fed into the network, such that the revenues of the v-RES plant sum up to the warranted EEG tariff (cf. eq. (1)).

$$R_{Normal} = P_{DA} + MP \tag{1}$$

The monthly market value does not feature in this equation, because it is paid by the direct marketer and received by the asset owner. It is therefore an internal cash flow and not of direct interest in this analysis. Given the green phase, the market premium is paid out of the so-called EEG account for which cash inflows are collected through the EEG levy from the electricity customers. These details are of considerable relevance when the overall system costs of flexibility markets are to be assessed (see section 7.1.1).

4.3.2. Curtailment by EinsMan (red light)

In case of a grid congestion, regulation allows the grid operator to curtail wind energy infeed as an ultimate measure (EinsMan, cf. section 3.3.1). In doing so, regulation stipulates that the grid operator who is responsible for the curtailment has to compensate the asset owner for lost revenues (*EEG*).[9] Under the assumption that the direct marketer sold the predicted wind power infeed on the spot market for a price P_{DA}, additional costs arise for balancing because the balancing group manager has to offset imbalances caused by the unscheduled curtailment due to EinsMan operations. This can be achieved by purchasing energy at the intraday market or paying for balancing energy. In practice, there is often no time to balance at the intraday market so that costs for balancing energy incur (P_{BC}). Consequently, the revenues and therefore opportunity costs ($R_{EinsMan}$) per MWh can be described as

$$R_{EinsMan} = P_{DA} - P_{BC} + EEG = P_{DA} - P_{BC} + MP + MMV \tag{2}$$

The *MP* and *MMV*, as part of equation (2), are paid by the grid operator as compensation for lost revenues directly to the asset owner, bypassing the direct marketer. The *MP* is paid as compensation, because payment of the "normal" *MP* depends on *actually* produced energy. The *MMV* has to be paid because the asset owner does not get remuneration for the curtailed energy from the direct marketer, and the *MMV* therefore constitutes lost revenues (see section 3.3.3).

4.3.3. Flexibility market (yellow light)

The auctions of the platform are held intraday with a lead time long enough for flexibility providers to modify their mode of operation and offset imbalances at the intraday market (for a price P_{ID}). If the v-RES plant already sold the power on the day ahead spot market (P_{DA}) and then successfully offers voluntary curtailment in the flexibility market (for a price P_{FM}), the revenues (R_{FM}) sum up to equation (3).

$$R_{FM} = P_{FM} - P_{ID} + P_{DA} \tag{3}$$

To participate in the flexibility market, the direct marketer has to receive at least the opportunity costs of the marketing alternative. In case of a congestion, the alternative would be EinsMan and the opportunity costs would be determined as explained in section 4.3.2. Therefore, the minimum price that the direct marketer has to receive to participate in the flexibility market is as derived in equation (4).

[8] This assumption is not uncontroversial. For a discussion see for example Bundesnetzagentur (2018).

[9] §15 EEG defines that the compensation shall cover 95% of lost revenues plus additional expenses minus saved expenses. If lost revenue exceeds 1 percent of the annual revenues, the amount will be fully compensated from that date. For simplicity, we will calculate with 100% and it is assumed that v-RES plants have no saved expenses in case of curtailment.

$$R_{FM} = R_{EinsMan} \Leftrightarrow P_{FM} - P_{ID} + P_{DA} = P_{DA} - P_{BC} + MP + MMV \Leftrightarrow P_{FM} = P_{ID} - P_{BC} + EEG \quad (4)$$

Note, that the grid operator would compensate the asset owner rather than the direct marketer for lost revenues in the EinsMan. The direct marketer "inherits" this responsibility from the grid operator when it decides to curtail the infeed voluntarily, as explained in section 3.3.3.

The formula above highlights that the direct marketer must cover the EEG remuneration that the asset owner would otherwise receive from the grid operator if the plant were curtailed through EinsMan. Additionally, less than the full EEG tariff might be accepted, because compared to EinsMan the direct marketer saves part of the costs for balancing energy ($P_{ID} - P_{BC}$) by balancing on the intraday market compared to the costs of buying balancing energy, which is the normal case in EinsMan.[10] Furthermore, the direct marketer might have some additional costs or savings when curtailing itself voluntarily as opposed to forcibly. In equation (5), we introduce a term C_{add} which is specific to each direct marketer and can be positive or negative to represent those additional costs or savings. In our further considerations we will simplify the equation by specifying that

$$P_{FM} = P_{ID} - P_{BC} + EEG + C_{add} \quad (5)$$

$$= EEG - X \quad (6)$$

The bid price is therefore the plant-specific EEG tariff corrected by an additional X-term, the value of which will be determined by the direct marketer (cf. eq. (6)).

5. Model implications

5.1. Distortions caused by EEG subsidies

As seen in section 4.3.3, the bid price for v-RES plants in the EEG scheme with direct marketing is based on the plant-specific EEG subsidies. In the following we show that part of these costs incurs regardless of whether the v-RES plant produces normally, is curtailed by EinsMan, or voluntarily reduces output because of its commitments in the flexibility market.

While producing normally, cash flows (*CF*) from the grid operator to the v-RES plant and direct marketer per MWh are

$$CF_{GO,\ normal} = \boldsymbol{MP}. \quad (7)$$

If curtailed by EinsMan, cash flows from the grid operator to the v-RES plant and direct marketer per MWh are

$$CF_{GO,\ EinsMan} = \boldsymbol{MP} + MMV. \quad (8)$$

In the flexibility market the bid price per MWh of the direct marketer (expecting EinsMan as the alternative) would be

$$P_{FM} = P_{ID} - P_{BC} + \boldsymbol{MP} + MMV + C_{add} \quad (9)$$

and this is the basis for the cash flow from the grid operator to the direct marketer if the bid is selected. It is apparent from equations (7)–(9) that the market premium is a fixed part of cash flows from the grid operator to the v-RES plant/direct marketer: Either

1. the v-RES plant is not running, and the grid operator must pay the market premium as part of congestion management (as compensation in case of curtailment, or implicitly as part of the bid price in case of voluntary shutdowns) or
2. the v-RES plant is running, and the market premium must be paid as EEG remuneration.

From an economic perspective, the market premium is paid either way and thus should not be included in the calculus when the optimal congestion management flexibilities are chosen. Consequently, an alternative flexibility option will only decrease overall costs when it is cheaper than renewable flexibilities excluding the market premium.

The merit order of flexibility options is biased if v-RES plants with market premia of different amount and/or other flexibility options without market premia bid in the same market. These uncorrected merit orders are sources of bias because they are including factors which should not be relevant for the decision about which flexibility options to contract.

Three effects can be distinguished:

1. Shifts in the merit order within the group of v-RES plants with the same energy source, because of different market premia. This leads to an inefficient selection of v-RES plants to contract because the market premium will be paid either way and should not play a role when choosing flexibility options.
2. Shifts in the merit order between v-RES plants with the same energy source and other flexibility options, which make v-RES plants look more expensive than they are compared to alternatives.
3. The grid operator will not be able to judge how much it should be willing to pay for flexibility in lieu of EinsMan curtailment because it does not know the real (effective) cost when choosing flexibility options.

These distortions can be avoided by paying the market premium independently of the bidding decisions of the direct marketer – then it is not part of the opportunity cost considerations of the direct marketer which consequently does not include it in its flexibility offer. Adjusting the merit order of the flexibility market can hence ensure that the grid operator will find the overall economic optimum when selecting flexibilities based on the merit order. We suggest the implementation of side payments, which exclude market premia from the flexibility market as described in section 5.2.

5.2. Side payments resolve distortions

As detailed in section 4.3.3, the reason for direct marketers to include the market premium in the bid price on the flexibility market is that it "inherits" the responsibility to remunerate the asset owner if it voluntarily curtails the v-RES plant, because the grid operator will only compensate the asset owner if the v-RES plant is running or "forcibly" curtailed by EinsMan.

There are therefore two ways to enable the direct marketer to not include the market premium in the bid price:

1. Pay the asset owner regardless of whether the v-RES plant is voluntarily curtailed or not.
2. Add an additional payment equaling the market premium to the direct marketer when contracting voluntary curtailment on the flexibility market.

As there are other reasons to voluntarily curtail the v-RES plant (e.g., scheduled or unscheduled maintenance) for which the grid operator should not reimburse the asset owner with the market premium, payment should be coupled to a contract on the flexibility market. We will refer to such payments as *side payments*. They will be paid retroactively to qualifying EEG subsidized power plants taking part in the flexibility market for each unit of flexibility sold. Because of the introduction of guaranteed side payments, which cover parts of the opportunity costs explained in section 4.3.3, renewable flexibility providers will decrease their bids in competitive markets by the same amount. Consequently,

[10] This is under the assumption that the direct marketer does not try to game the system by profiting of balancing energy.

the market premium will be externalized from the flexibility market and an undistorted flexibility merit order is obtained. All bids then represent true additional costs incurred by the grid operator.

The next section will illustrate the effect side payments (and a lack thereof) have on the merit order of flexibility options.

6. Application and results

Based on the considerations in sections 3-5, we construct a simple fictitious example to illustrate how flexibility markets can contribute to congestion management. Because we focus on the distortions caused by subsidies for v-RES, we keep the example as simple as possible while still preserving the distorting effects. We compare the results of congestion management without a flexibility market, which is basically EinsMan, i. e. renewable curtailment, to the market results of a flexibility market with and without side payments. In this comparison we focus on overall system costs and the energy infeed of renewables.

6.1. Focus on wind energy

In principle, our model applies to all kind of v-RES plants that are in direct marketing, remotely controllable, and subsidized in the EEG system. In this regard, the most important technologies are wind and solar based renewable energy sources. However, in Germany, only 25 percent of installed solar plants are in direct marketing, which in turn is approximately a fifth of installed wind onshore capacity that is marketed directly.[11] On this account, we focus on onshore wind power plants to provide generation-side flexibility in our example. This is in line with the reality of congestion management, where curtailment of solar power plants only accounts for three percent of overall curtailed volumes. Curtailment of onshore wind power in contrast, represents more than 80 percent of overall volumes as part of EinsMan.[12] For a discussion of this assumption see section 7.3.

6.2. Setup of the example

Given that our focus is on congestion management, we disregard all cost and revenue streams related to energy spot markets, as these stay constant, and consequently system costs only include costs for congestion management plus the EEG remuneration of renewables since the latter may be affected by the congestion management. In doing so, one must be careful when adding up these different cost components, since the costs of congestion management are financed by network charges whereas the costs for supporting EEG plants are paid out of the EEG account. In this section both cost components will be treated as part of system costs, but they are analyzed more in detail from a grid operator's perspective in section 7.1.2.

Table 1 details a situation where the grid operator predicts a congestion at a transformer of the distribution grid that it can manage using five different flexibility providers in a market area.

In the example, three wind farms may offer to curtail power infeed voluntarily on the market. The wind farms differ in the years they were commissioned which lead to specific remunerations defined by the respective EEG (Netztransparenz, 2019). In addition, two other flexibilities are available on the market, a shiftable load and a power-to-gas plant. Opportunity costs for the shiftable load are calculated as the mean spread of quarterly intraday prices in 2018 as explained in section 4.1.

Table 1
Available flexibility options. Total needed flexibility is 3 MW, available flexibility per plant is 1 MW.

Name	Type	Commissioning Month/Year	EEG[a]	MP	MMV + - X	Other
RES 1	Wind Onshore	01/2018	71	36	35	
RES 2	Wind Onshore	01/2017	80	45	30	
RES 3	Wind Onshore	07/2011	90	55	20	
Other 1	Shiftable Load					14
Other 2	Power-to-Gas					60

[a] The remunerations are rounded to integer values.

The opportunity costs of the power-to-gas plant depend on levies and taxes for additional electricity consumption as well as the efficiency of the plant and the value of gas produced. Given that, the opportunity costs of the power-to-gas plant are assumed as approx. 60 €/MWh.[13]

For simplicity, all actors can offer a flexibility of 1 MW and the grid operator's flexibility demand is 3 MW. The monthly market value for wind farms, which is the same for all wind onshore plants, is set by 35 €/MWh (as it was in January 2018). Another column details the monthly market value offset by a value "X" which represents any markup or a markdown from the total feed in tariff, such that

$$P_{FM} = EEG - X = MP + MMV - X. \tag{10}$$

The "X" from equation (10) sets wind power plants apart from each other in the flexibility market and may represent risk premia, expected cost savings or any other factors.[14]

6.3. Curtailment via EinsMan

Currently, no market-based instrument allows the distribution grid operator to make use of shiftable or additional loads to manage congestions. That is why the grid operator would fall back on EinsMan and curtail the wind infeed as illustrated below.

On the left side of Fig. 4, the merit order of the EinsMan measures is shown. Efficient EinsMan would select the power plant with the lowest EEG tariff first for congestion management, followed by the second and third cheapest flexibilities. To successfully avert the grid congestion, all wind power plants must be curtailed resulting in EinsMan costs of 241 €. This amount corresponds to the overall costs since no market premia for operating wind power plants must be paid. At the same time, it is the benchmark to assess the market result of the flexibility platform.

6.4. Flexibility market without side payments

The introduction of a flexibility market enables the grid operator to contract flexible loads for congestion management. Due to the inclusion of new and cheaper flexibility options, the merit order is changed as shown in Fig. 5. In this case, revenues from the flexibility market must cover opportunity costs of renewables as described in section 4.3.3. Thus, the bids must correspond approximately to the foregone EEG remuneration. Non-renewable flexibilities in contrast are assumed to bid at marginal costs.[15] If non-renewables can bid at lower cost than EEG

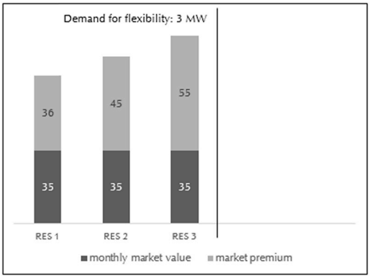
[11] Based on own calculations.

[12] Bundesnetzagentur and Bundeskartellamt (2018).

[13] We assume that the power-to-gas plant is exempted from network charges, EEG and KWK levies, so that levies and taxes sum up to approximately 26 €/MWh. Given the average intraday wholesale prices of electricity and gas in 2018 as well as an efficiency factor of 60% the opportunity costs were calculated.

[14] See section 4.3.3.

[15] For a discussion of strategic bidding see section 7.2.1.



Costs for flexibility

Name	Type	MMV	MP	other
RES 1	Wind Onshore	35	36	
RES 2	Wind Onshore	35	45	
RES 3	Wind Onshore	35	55	

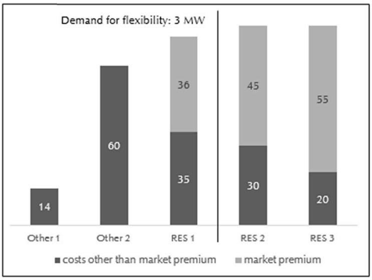
Costs for running RES

Name	Type	MMV	MP	other

Total relevant system costs

Type	Value
Congestion management	241
Running plants	0
Total	**241**

Fig. 4. Merit order of flexibility options and costs in the EinsMan curtailment.



Costs for flexibility

Name	Type	MMV +- X	MP	other
RES 1	Wind Onshore	35	36	
Other 1	non RES			14
Other 2	non RES			60

Costs for running RES

Name	Type	MMV	MP	other
RES 1	Wind Onshore			
RES 3	Wind Onshore		55	
RES 2	Wind Onshore		45	

Total relevant system costs

Type	Value
Congestion management	145
Running plants	100
Total	**245**

Fig. 5. Merit order of flexibility options and associated costs on a flexibility market.

remuneration, EEG-power plants move to the right in the flexibility merit order.

In this case, Other 1 as cheapest flexibility option would be selected first, followed by Other 2. To solve the congestion completely, the grid operator can limit curtailment to only the cheapest of the wind power plants (RES 1). This results in congestion management costs summing up to 145 €.[16] However, since the power plants RES 2 and RES 3 generate power and are compensated in accordance with §19 EEG in combination with §20 EEG, the market premium is paid to these units and has to be considered in a system perspective. This finally adds up to overall costs of 245 € which leads to an increase of system costs compared to the reference case of EinsMan despite decreasing congestion management costs.

6.5. Flexibility market with side payments

Contrary to the previous case, the wind power plants RES 1, RES 2 and RES 3 receive their individual market premium as a side payment if they successfully bid in the flexibility market. Thus, their bids do not have to fully cover the EEG tariff plus mark-up but only the *MMV* adjusted for X as shown in equation (11).

$$bid_{flex}^{wind} = MMV \cancel{+MP} - X = MMV - X \qquad (11)$$

Consequently, the merit order changes significantly (see Fig. 6). Now, the wind flexibilities are cheaper compared to Other 2 and the grid operator would select Other 1, RES 2 and RES 3 to manage the congestion. This results in congestion management costs of 164 €. These costs consist of the payments for accepted bids on the market (Other 1 and the two wind power flexibilities) and the side payments for RES 2 and RES 3, which the grid operator will pay out after determination of the market premium.

Given this scenario, RES 1 feeds in power without restrictions and its market premium must be paid additionally, leading to overall system costs of 200 € - a new minimum.

7. Discussion

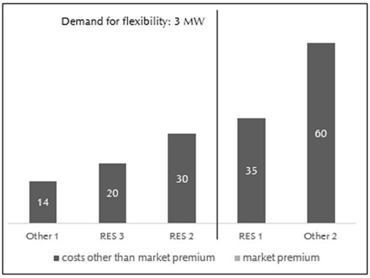
7.1. Discussion of the results

Table 2 provides an overview of the considered cases. Several aspects about the results after introducing side payments are worth noting and will serve as basis for further discussions:

1. Overall system costs decrease significantly, since all arising cost components are reflected adequately in the decision of flexibility selection. It should be noted that, although overall system costs decrease, costs of congestion management increase slightly.
2. Flexibilities provided by renewables improve their position in the merit order. This leads to more curtailment of renewable power.
3. The position of the renewable plants within their "group" changes. While RES 1 was the cheapest option before, it is now the most expensive because its "X" is the smallest. This illustrates that once the individual market premia stop distorting the bids, individual cost factors of the plants become relevant.

Based on these results, there are several aspects that need to be discussed in more detail. Firstly, there is a trade-off between the minimization of congestion management and overall system costs when

[16] Note that we assume that the market clears on a pay-as-bid basis. For a discussion of market clearing rules see section 7.2.1.



Costs for flexibility

Name	Type	MMV +- X	SP	other
Other 1	non RES			14
RES 2	Wind Onshore	30	45	
RES 3	Wind Onshore	20	55	

Costs for running RES

Name	Type	MMV	MP	other
RES 1	Wind Onshore		36	
RES 2	Wind Onshore			
RES 3	Wind Onshore			

Total relevant system costs

Type	Value
Congestion management	164
Running plants	36
Total	**200**

Fig. 6. Merit order of flexibility options and associated costs on a flexibility market utilizing side payments.

Table 2
Summary of associated costs in different cases of congestion management.

Type	Funding source	EinsMan	Market w/o SP	Market w/ SP
Congestion management	Grid tariff	241	145	164
Operating EEG plants	EEG account	0	100	36
Total Costs		**241**	**245**	**200**
RES infeed in MW		0	2	1

choosing the best market design. Additionally, the introduction of side payments to achieve the best results in terms of overall costs leads to lower renewables infeed compared to a flexibility market without side payments. Lastly, the interplay between market clearing rules and bidding strategies of flexibility providers needs to be investigated in more detail.

7.1.1. Congestion management vs overall system costs

The introduction of a flexibility market decreases congestion management costs compared to EinsMan in both cases, both with and without side payments (−77 €, respectively −96 €). This is logical because system operators can apply new and cheaper flexibilities to manage congestions. It is also clear that congestion management costs in a market without side payments are lowest because the grid operator can select flexibilities with the lowest absolute costs, which the grid operator must compensate itself.

However, in a market without side payments, the market premium has a great effect on the selection of flexibilities, although the market premium has to be paid anyway. Therefore, a market without side payments, even though it leads to the lowest costs of congestion management, may induce higher overall costs than curtailment via EinsMan when considering congestion management and EEG cost as shown in previous sections (EinsMan: 241 €, Market without SP: 245 €). In order to prevent the grid operators from making market decisions based on non-influenceable cost components, side payments are introduced. In this case, the overall costs are minimized because the cheapest flexibilities that have an actual impact on the overall system costs are selected.

7.1.2. Grid operator's perspective

Since grid operators are acting for their own account, they are likely to prefer a market outcome which minimizes their own cost, namely congestion management costs, rather than overall system costs. Minimizing congestion management costs in a market without side payments would lead to lower network fees and relieve the grid operator's customers. Given that the minimization of congestion management costs can entail higher overall system costs, this comes along with a transfer of costs to all customers via a higher renewable levy. From an overall system perspective, higher EEG-levy payments for all consumers would overcompensate the lower network charges for a few customers of a certain grid operator – at least in an electricity system like the German one with multiple distribution grid operators.

Here, incentives for grid operators should be aligned so that they are incentivized to choose the solution with the lowest overall cost. This is also true for the more general issue of incentives to minimize congestion management costs. If those are considered as pure pass-through item as in the current German regulation, the incentives to implement flexibility markets are very limited. Yet a detailed discussion of regulatory approaches in that field is beyond the scope of this paper.

7.1.3. Renewables infeed

Since flexibility providers ask for compensation of the whole feed-in tariff as opportunity costs, the bids in the market without side payments are quite high and renewable flexibilities are expensive compared to other flexibility options. That is why without side payments, the wind power flexibilities would be selected last by the grid operator which leads to the highest renewable infeed possible (2 MW). With the implementation of side payments, the market prices for flexibility from renewables are lower. That is why renewable flexibilities would be selected more frequently and the infeed of renewables would decrease, in our numerical example to 1 MW.

However, compared to currently used EinsMan, the infeed of electricity generated by renewables increases in both variants of flexibility markets. This is a consequence of new demand-side flexibility options being made available for congestion management through the introduction of the flexibility markets.

In the end, it is hence a political choice, which flexibility market design is preferred. The choice then reflects a tradeoff between overall costs and the maximization of renewable infeed, which should be made consciously and not just as an unintended side effect.

7.2. Discussion of the assumptions

7.2.1. Market clearing rules and bidding strategies

So far, the discussion has focused on the impacts of the different flexibility market designs. We have assumed a naïve pay-as-bid approach, where flexibilities (somewhat naively, see below) only bid their opportunity costs and are paid according to their bids. While this approach allowed us to isolate the effects of distorting subsidies, one may wonder about the pricing implications of different market designs and this requires a consideration of the market rules and related agent behavior.

Basically, two alternatives may be considered for the pricing rule: uniform pricing (also called pay-as-cleared) or pay-as-bid (or discriminatory pricing). Whereas auction-based energy spot markets typically use uniform pricing, pay-as-bid has been prevalent in the German and other reserve power markets. At first sight, pay-as-bid has the advantage that the buyer (in our case the grid operator) does not pay more than requested by the sellers. Yet this reasoning does not consider that sellers will adapt their bidding strategy to the clearing rules (and will not naively continue to only bid opportunity costs, as we assumed in the example). Academic literature indicates that uniform pricing rules provide more incentives for cost-based bids whereas pay-as-bid leads to "guess-the-price" type bidding behavior (e.g. Wolfram, 1999; Newberry and McDaniel, 2002; Cramton, 2017). Under perfect competition and complete information, it may even be proven that the two auction formats lead to the same result (Müsgens et al., 2014). Hence bidding at marginal costs is no valid assumption in a pay-as-bid market design and, independently of the auction design in flexibility markets, the grid operator as buyer will pay more than the cost of the bidders. This is visualized in Fig. 7 for the case of uniform pricing – which is at identical marginal bids the upper bound for prices under pay-as-bid.

If there is moreover only a limited number of bidders, those may bid strategically and exert market power. In this setting, it is important to ensure that the system costs are not higher than under the traditional command and control strategy of EinsMan and that at the same time there are sufficient incentives for market entry. This calls for a market design combining the following three elements:

- Uniform pricing
- Clear price caps to avoid costs in excess of EinsMan
- Side payments

The two main arguments for uniform pricing are that bidding is simpler for small market participants since they can bid their marginal costs instead of having to guess the price and that the incentives for putting strategic bids are lower (since they only occur for marginal units). The price cap is required to ensure that the market provides economic benefits, and the side payments make setting this price cap both easier and more efficient. The discussion in section 5.1 has shown that the market premia *MP* must be paid independently of the market design as well as under EinsMan. Under EinsMan, the additional payment to all curtailed EEG units is the monthly market value *MMV*. Therefore, this is the price cap to be used in the flexibility market with side payments.[17] This ensures that the system costs in this market design do not exceed the cost of the conventional approach.[18] Furthermore, congestion management cost will not exceed the corresponding cost under EinsMan. If a mix of v-RES is curtailed in normal EinsMan procedures, there is no single monthly market value on which to base the price cap. In this case, a prudent solution would be to set price caps to the monthly market value of the dominant v-RES in the market area, but there is no guarantee that costs for congestion management will remain at or below corresponding EinsMan costs. This is also true for the flexibility market without side payments, where setting the price cap is not as easy since it will depend on the mix of EEG units curtailed under EinsMan and the corresponding *MP* – even if all units have the same monthly market value. If this is established correctly, the congestion management cost may be kept under the EinsMan cost, yet this is not true for the system cost as established previously. Fig. 7 even illustrates that the operating margins earned in a market without side-payments may be substantially higher than with side payments – i.e. the distributional effects are stronger, with the flexibility providers as beneficiaries and the consumers (via grid fees and EEG levy) as the losers. On the other hand, the limited operation margins in flexibility markets with side payments still provide incentives for market entry.

[17] This obviously implies that EinsMan is used as recourse action if the number of bids on the flexibility market below the price cap is insufficient.

[18] Here obviously the costs of setting up and operating the market are disregarded (as are the cost of the conventional EinsMan) and it is assumed throughout that all bids have the same effectiveness with respect to congestion relief – although a generalization is possible to bids with different sensitivities.

7.2.2. Coping with uncertainties

Side payments are intended to counter distortions in the market by removing the market premium from the bid price of renewable flexibilities. This is complicated by the fact that the values of the individual market premia are unknown at the point of bidding, because they are calculated as the difference between the monthly market value of the given renewable power source and the individual power plants EEG feed-in tariff.

Consequently, the direct marketers as well as the grid operators have to predict the monthly market value. The direct marketer has to predict the *MMV* because it provides the base line for its bid price. The grid operator has to predict the *MMV* to determine the price cap in order not to exceed EinsMan costs.

This raises the question, what effect uncertainties with respect to the *MMV* have on the market. If both the direct marketers as well as the grid operator are risk averse, they may place uplifts on their bids and discounts on the price cap, respectively. This reduces the leeway for market clearing since even if both the grid operator and the direct marketers predict the same *MMV,* bids may be placed above the price cap.

One possible solution could be for the grid operator to announce a "benchmark *MMV*" before the flexibility auctions. Rather than calculating the side payment based on the actual (later) established *MMV*, the side payment would be calculated using the benchmark *MMV*. The direct marketers could be sure that they will be paid the complete EEG-tariff if they bid the benchmark *MMV*. Of course, this approach also informs other market participants about the price cap set by the grid operator and they may adjust their bids accordingly – yet under uniform pricing the inframarginal bids do not have an incentive to do so.

Further research should be done on the benefits and possible drawbacks of such signaling as well as on alternative solutions to cope with uncertainties.

Another type of uncertainty concerns the volumes on offer from the sellers and the volumes needed by the grid operators to relieve congestions. We have assumed that both the sellers as well as the grid operators have somehow reached a decision on the volume to sell and buy respectively. This assumption is valid, because uncertainty about the volume should not change the price of the bid from the seller or the willingness to pay from the grid operator. This type of uncertainty would push both providers of flexibilities as well as grid operators to wait for the uncertainty to reduce. Other factors, such as the desire to ensure that enough flexibility can be contracted, counter this effect. The trade-off between securing flexibility early and waiting for more exact forecasts is not within the scope of this paper but is examined in Bellenbaum et al. (2019).

7.3. Generalizing the results

Even though reality is obviously more complex than the simple example chosen to illustrate the mechanism, the underlying principles remain true. There is no reason to assume that the distorting effects would decrease in a more complex setting. Taking into account meshed grids, where different flexibilities have different sensitivities on congestions, or assuming bigger markets, where more units are in competition with each other, does not change the fact that the basic calculations performed by direct marketers to form their bids consider factors that, in the end, distort the merit order of flexibility options. Further research should quantify the real world impact of the implementation of side payments – or the impact of the distorting effects if they are not implemented.

Additionally, we have focused on wind power as the dominant technology for v-RES in Germany. If multiple v-RES coexist in the flexibility market (as they often will), each should receive side payments according to the energy source specific monthly market values. This complicates the process of setting the price cap as detailed in section 7.2.1.

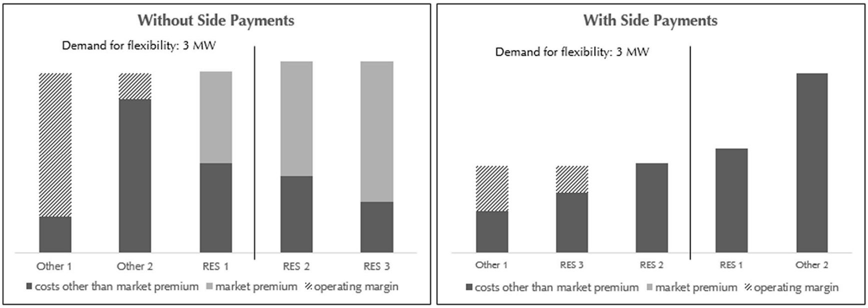


Fig. 7. Uniform pricing in markets without and with side payments and resulting.

8. Conclusion and policy implications

This paper analyses key design elements for flexibility markets based on theoretical considerations. Compared to conventional congestion management methods like the curtailment of renewables (e.g. EinsMan), flexibility markets can provide the opportunity to include new flexibility options and thereby reduce costs for congestion management. However, our analysis also shows that renewable flexibilities bidding at opportunity costs can lead to an inefficient and costly market outcome under the current regulatory framework. This market distortion is related to the payments made to renewable units if they are curtailed. These potential revenues are then included into the opportunity cost on which bids into the flexibility market are based.

To avoid such distortion, we recommend taking the following three actions labelled A1 to A3. Additionally, our work points at the key political choices D1 and D2 that are not directly addressed in our analysis but nonetheless consequential for the implementation of flexibility markets.

A1. Implementation of side payments. The size of the side payments should correspond to the so-called market premium paid to the renewable unit, since this payment is due anyway – either as renewable support payment if power infeed is unlimited, or as part of the compensation if the renewable unit was curtailed due to a congestion. These side payments are triggered whenever renewable units submit bids to the flexibility market, and these are selected by the grid operator. In this case, the grid operator's optimal flexibility selection on the market results in an overall system cost minimum.

A2. Payments of market premia (and side payments) from the EEG account in case of congestion management. To make such a market design attractive for grid operators who aim to minimize their congestion management costs, we suggest paying the market premia from the EEG account not only for operating units but also in case of congestion management. This would externalize the market premia as cost components from the grid operators' calculation and thus incentivize it to choose the solution with the lowest overall cost. Finally, the funding of market premia from the EEG account would prevent unwanted distributional effects between consumer groups that would occur if a single grid operator minimized its congestion management costs, and thus the network charges for its customers, at the expense of rising EEG levies for all consumers.

A3. Implementation of uniform pricing and a price cap to limit strategic behavior in flexibility markets. The flexibility market should be based on uniform pricing and include a price cap in order to limit incentives for strategic bidding. Setting the price cap to the level of the monthly market value can ensure that the market outcomes are beneficial compared to a conventional command and control approach.

The key issues that are consequential for the implementation of flexibility markets are:

D1. Tradeoff between system costs and renewable infeed. The introduction of a flexibility market with and without side payments decreases the curtailment of renewables by acquiring new flexibility options. While the flexibility market without side payments decreases renewable curtailment the most, it brings along high overall costs (in some cases even higher than in case of EinsMan). The renewable infeed in case of a flexibility market with side payments may be lower than in case of no side payments, yet it achieves a market outcome at minimum overall system costs.

D2. Regulatory treatment of flexibility costs. How to treat flexibility costs as part of congestion management is still an open question. If – as in the current German regulation – congestion management costs such as EinsMan reimbursements are considered as pure pass-through items, the incentives to implement flexibility markets are very limited.

Flexibility markets can contribute to the efforts to implement market-based congestion management measures as demanded by the European Commission. The actions we have specified and the necessary political decisions we have identified contribute to the efficient design of such flexibility markets.

Declaration of competing interest

None.

SINTEG
SMART ENERGY SHOWCASES

Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

Acknowledgements

We drafted this paper within the scope of the research project "enera" (project code 03SIN330) as part of the SINTEG funding program of Federal Ministry for Economic Affairs and Energy of Germany.

References

Agsten, M., Bauknecht, D., Becker, A., Brinker, W., Conrads, R., Diebels, V., Erge, T., Feuerhahn, S., Heinemann, C., Hermsmeier, J., Hollinger, R., Klose, T., Koch, M., Mayer, C., Pistoor, G., Rosinger, C., Rüttinger, H., Schmedes, T., Stadler, M., 2013. Abschlussbericht eTelligence: Neue Energien Brauchen Neues Denken. Oldenburg. http://www.etelligence.de/feldtest/file/eTelligence%20Projektabschlussbericht%20-%20November%202012.pdf.

Apel, R., Berg, V., Fey, B., Geschermann, K., Glaunsinger, W., Scheven, A. von, Stötzer, M., Wanzek, S., 2014. Regionale Flexibilitätsmärkte: marktbasierte Nutzung von regionalen Flexibilitätsoptionen als Baustein zur erfolgreichen Integration von erneuerbaren Energien in die Verteilnetze. Studie der Energietechnischen Gesellschaft im VDE. https://shop.vde.com/de/vde-studie-regionale-flexibilit%C3%A4tsm%C3%A4rkte.

BDEW, 2013. BDEW-Roadmap: Realistische Schritte zur Umsetzung von Smart Grids in Deutschland. bdew, Berlin, 71 pp. https://www.bdew.de/media/documents/Pub_20130211_Roadmap-Smart-Grids.pdf. (Accessed 12 December 2018).

BDEW, 2018. Redispatch in Deutschland: Auswertung der Transparenzdaten. bdew, Berlin, 27 pp. https://www.bdew.de/media/documents/Awh_20180212_Bericht_Redispatch_Stand_Februar-2018.pdf. (Accessed 16 November 2018).

Bellenbaum, J., Höckner, J., Weber, C., 2019. Designing Flexibility Procurement Markets for Congestion Management – Investigating Two Stage Procurement Auctions. HEMF Working Paper.

Bjørndal, E., Bjørndal, M., Cai, H., Panos, E., 2018. Hybrid pricing in a coupled European power market with more wind power. Eur. J. Oper. Res. 264 (3), 919–931. https://doi.org/10.1016/j.ejor.2017.06.048.

Bjørndal, M., Jørnsten, K., 2001. Zonal pricing in a deregulated electricity market. Energy J. 22 (1), 51–73.

Bjørndal, M., Jørnsten, K., 2007. Benefits from coordinating congestion management: the Nordic power market. Energy Policy 35 (3), 1978–1991. https://doi.org/10.1016/j.enpol.2006.06.014.

Bjørndal, M., Jørnsten, K., Pignon, V., 2003. Congestion management in the nordic power market: counter purchases and zonal pricing. Compet. Regul. Netw. Ind. 4 (3), 271–292. https://doi.org/10.1177/178359170300400302.

BMWi, 2016. Information on the Funding Programme Entitled 'Smart Energy Showcases – Digital Agenda for the Energy Transition' (SINTEG). https://www.bmwi.de/Redaktion/EN/Downloads/bmwi-papier-sinteg-kernbotschaften.pdf?_blob=publicationFile&v=3. (Accessed 9 January 2019).

Bundesnetzagentur, 2011. Smart Grid" und "Smart Market": Eckpunktepapier der Bundesnetzagentur zu den Aspekten des sich verändernden Energieversorgungssystems. Bonn, 50 pp. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/NetzzugangUndMesswesen/SmartGridEckpunktepapier/SmartGridPapierpdf.pdf?_blob=publicationFile. (Accessed 15 November 2018).

Bundesnetzagentur, 2018. Ergänzender Hinweis zur Entschädigung bei direktvermarkteten EE-Anlagen nach dem Einspeisemanagement-Leitfaden 3.0. Bonn. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/ErneuerbareEnergien/Einspeisemanagement/Leitfaden3_0_E/Leitfaden3.0final.pdf. (Accessed 14 January 2019).

Bundesnetzagentur, 2019. Quartalsbericht zu Netz- und Systemsicherheitsmaßnahmen: Gesamtjahr und Viertes Quartal 2018. Bundesnetzagentur, Bonn, p. 73.

Bundesnetzagentur, Bundeskartellamt, 2018. Monitoringbericht 2018, Bonn, 517 pp. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Bundesnetzagentur/Publikationen/Berichte/2018/Monitoringbericht_Energie2018.pdf?_blob=publicationFile&v=3.

Cramton, P., 2017. Electricity market design. Oxf. Rev. Econ. Policy 33 (4), 589–612. https://doi.org/10.1093/oxrep/grx041.

Ecofys, Fraunhofer, I.W.E.S., 2017. Smart-Market-Design in deutschen Verteilnetzen: entwicklung und Bewertung von Smart Markets und Ableitung einer Regulatory Roadmap. Studie im Auftrag von Agora Energiewende, 156 pp. https://www.agora-energiewende.de/fileadmin/Projekte/2016/Smart_Markets/Agora_Smart-Market-Design_WEB.pdf. (Accessed 21 March 2017).

Ehrenmann, A., Smeers, Y., 2005. Inefficiencies in European congestion management proposals. Util. Policy 13 (2), 135–152. https://doi.org/10.1016/j.jup.2004.12.007.

European Commission, 2016. Proposal for a REGULATION of the EUROPEAN PARLIAMENT and of the COUNCIL on the Internal Market for Electricity (Brussels).

Fotouhi Ghazvini, M.A., Lipari, G., Pau, M., Ponci, F., Monti, A., Soares, J., Castro, R., Vale, Z., 2019. Congestion management in active distribution networks through demand response implementation. Sustain. Energy, Grids. Netw. 17, 100185. https://doi.org/10.1016/j.segan.2018.100185.

Heussen, K., Bondy, D.E.M., Hu, J., Gehrke, O., Hansen, L.H., 2013. A clearinghouse concept for distribution-level flexibility services. In: 2013 4th PES Innovative Smart Grid Technologies Europe (ISGT Europe 2013): Conference] ; Lyngby [near Copenhagen], Denmark, 6 - 9 October 2013. 2013 4th IEEE/PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), Lyngby, Denmark. 10/6/2013 - 10/9/2013. IEEE, Piscataway, NJ, pp. 1–5.

Heydt, G.T., Chowdhury, B.H., Crow, M.L., Haughton, D., Kiefer, B.D., Meng, F., Sathyanarayana, B.R., 2012. Pricing and control in the next generation power distribution system. IEEE Trans. Smart Grid 3 (2), 907–914. https://doi.org/10.1109/TSG.2012.2192298.

Hirth, L., Schlecht, I., 2018. Market-based redispatch in zonal electricity markets. USAEE Working Paper. https://doi.org/10.2139/ssrn.3286798 (18-369).

Hogan, W.W., 1992. Contract networks for electric power transmission. J. Regul. Econ. 4 (3), 211–242. https://doi.org/10.1007/BF00133621.

Holmberg, P., Lazarczyk, E., 2015. Comparison of congestion management techniques: nodal, zonal and discriminatory pricing. Energy J. 36 (2) https://doi.org/10.5547/01956574.36.2.7.

Huang, S., Wu, Q., Oren, S.S., Li, R., Liu, Z., 2015. Distribution locational marginal pricing through quadratic programming for congestion management in distribution networks. IEEE Trans. Power Syst. 30 (4), 2170–2178. https://doi.org/10.1109/TPWRS.2014.2359977.

Karg, L., Kleine-Hegermann, K., Wedler, M., Jahn, C., 2014. E-Energy Abschlussbericht: Ergebnisse und Erkenntnisse aus der Evaluation der sechs Leuchtturmprojekte. http://www.e-energie.pt-dlr.de/images/ERGEBNISBERICHT.pdf. (Accessed 21 March 2017).

Koch, M., Schmedes, T., Stadler, M., Klose, T., Hollinger, R., Rüttinger, H., Rosinger, C., 2012. Integratives Smart Market Konzept zur Systemintegration dezentraler Erzeuger und als Handelsplattform für Netzbetreiber. In: Fechner, R. (Ed.), Smart Grid: Intelligente Energieversorgung der Zukunft ; Kongressbeitäge ; 5. - 6. November 2012, Internationales Congresscenter Stuttgart (ICS). VDE-Verl., Berlin.

Liu, W., Wu, Q., Wen, F., Ostergaard, J., 2014. Day-ahead congestion management in distribution systems through household demand response and distribution congestion prices. IEEE Trans. Smart Grid 5 (6), 2739–2747. https://doi.org/10.1109/TSG.2014.2336093.

Müsgens, F., Ockenfels, A., Peek, M., 2014. Economics and design of balancing power markets in Germany. Int. J. Electr. Power Energy Syst. 55, 392–401. https://doi.org/10.1016/j.ijepes.2013.09.020.

Netztransparenz, 2019. EEG-Vergütungskategorientabelle mit allen Kategorien bis Inbetriebnahmejahr 2019. https://www.netztransparenz.de/portals/1/EEG-Verguetungskategorien_EEG_2019_20190430.xls. (Accessed 27 August 2019).

Neuhoff, K., Barquin, J., Bialek, J.W., Boyd, R., Dent, C.J., Echavarren, F., Grau, T., Hirschhausen, C. von, Hobbs, B.F., Kunz, F., Nabe, C., Papaefthymiou, G., Weber, C., Weigt, H., 2013. Renewable electric energy integration: quantifying the value of design of markets for international transmission capacity. Energy Econ. 40, 760–772. https://doi.org/10.1016/j.eneco.2013.09.004.

Newberry, D., McDaniel, T., 2002. Auctions and trading in energy markets: an economic analysis. Cambridge working papers in economics 0233. Faculty of Economics, University of Cambridge. https://ideas.repec.org/p/cam/camdae/0233.html. Accessed 1 April 2019.

Oggioni, G., Smeers, Y., 2013. Market failures of Market Coupling and counter-trading in Europe: an illustrative model based discussion. Energy Econ. 35, 74–87. https://doi.org/10.1016/j.eneco.2011.11.018.

Sarfati, M., Hesamzadeh, M.R., Holmberg, P., 2019. Production efficiency of nodal and zonal pricing in imperfectly competitive electricity markets. Energy Strat. Rev. 24, 193–206. https://doi.org/10.1016/j.esr.2019.02.004.

Schweppe, F.C., Caramanis, M.C., Tabors, R.D., Bohn, R.E., 1988. Spot Pricing of Electricity. Springer US, Boston, MA.

Sotkiewicz, P.M., Vignolo, J.M., 2006. Nodal pricing for distribution networks: efficient pricing for efficiency enhancing DG. IEEE Trans. Power Syst. 21 (2), 1013–1014. https://doi.org/10.1109/TPWRS.2006.873006.

Steurer, M., 2017. Analyse von Demand Side Integration im Hinblick auf eine effiziente und umweltfreundliche Energieversorgung. Dissertation. Stuttgart, p. 268.
Torbaghan, S.S., Blaauwbroek, N., Nguyen, P., Gibescu, M., 2016. Local market framework for exploiting flexibility from the end users. In: 2016 13th International Conference on the European Energy Market (EEM): 6-9 June 2016, Porto, Portugal. 2016 13th International Conference on the European Energy Market (EEM), Porto, Portugal. 6/6/2016 - 6/9/2016. IEEE, Piscataway, NJ, pp. 1–6.
Villar, J., Bessa, R., Matos, M., 2018. Flexibility products and markets: literature review. Electr. Power Syst. Res. 154, 329–340. https://doi.org/10.1016/j.epsr.2017.09.005.
Vries, L.J. de, Hakvoort, R.A., 2002. An economic assessment of congestion management methods for electricity transmission networks. J. Netw. Ind. 3 (4), 425–466. https://doi.org/10.1177/178359170200300403.
Weibelzahl, M., 2017. Nodal, zonal, or uniform electricity pricing: how to deal with network congestion. Front. Energy 11 (2), 210–232. https://doi.org/10.1007/s11708-017-0460-z.
Weibelzahl, M., Märtz, A., 2018. On the effects of storage facilities on optimal zonal pricing in electricity markets. Energy Policy 113, 778–794. https://doi.org/10.1016/j.enpol.2017.11.018.
Wiedemann, T., 2017. Proaktives Verteilnetz - Proactive Distribution Grid. ETIP SNET Workshop. Aachen. https://www.etip-snet.eu/regional-workshops/central-region/aachen-workshop-presentations/. (Accessed 14 January 2019).
Wolfram, C.D., 1999. Electricity markets: should the rest of the world adopt the United Kingdom's reforms?: a formula for inefficient production and higher prices. Regulation 22 (4), 48–53.
Zhang, C., Ding, Y., Ostergaard, J., Bindner, H.W., Nordentoft, N.C., Hansen, L.H., Brath, P., Cajar, P.D., 2013. A flex-market design for flexibility services through DERs. In: 2013 4th PES Innovative Smart Grid Technologies Europe (ISGT Europe 2013): Conference] ; Lyngby [near Copenhagen], Denmark, 6 - 9 October 2013. 2013 4th IEEE/PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), Lyngby, Denmark. 10/6/2013 - 10/9/2013. IEEE, Piscataway, NJ, pp. 1–5.

Chapter 3

Flow-Based Market Coupling – What Drives Welfare in Europe's Electricity Market Design?

by Simon Voswinkel, Björn Felten, Tim Felling, and Christoph Weber
HEMF Working Paper No. 08/2019

Abstract

Over the last decades, two basic design alternatives for spot markets for electricity have been established: The zonal approach as used, for example, in Europe and the nodal approach as applied notably in US markets. Since 2015, Flow-Based Market Coupling is used as an advanced market coupling approach to facilitate the exchange of electricity between the zonally organized markets in Central Western Europe. But how good is this state-of-the-art zonal approach compared to its nodal benchmark? And can enhancements to its relatively new rules and procedures improve the performance of Flow-Based Market Coupling?

We develop a model framework to analyze and quantify the welfare of Flow-Based Market Coupling relative to benchmarks like nodal designs in a real-world setting. We find that under ideal circumstances, where price zones are well-configured, Flow-Based Market Coupling approaches the efficiency of nodal pricing – realizing 87 % of the possible gains in comparison to a scenario with unlimited trade as the lower benchmark. We also find it to be relatively robust in the presence of forecast errors. When taking the current European price zone configuration into account, however, the efficiency of Flow-Based Market Coupling decreases significantly. Only 59 % of the efficiency of the nodal market design can be attained, creating societal losses of more than 500 million Euros each year. Moreover, we find the measures envisaged by European regulators to do further harm in terms of welfare. These measures are designed to enhance trade but, to a certain extent, ignore the physical reality of the transmission network. This entails significant increases in redispatch quantities, and operational system costs further rise by about 100 million Euros per year.

Keywords: Flow-based market coupling; Zonal pricing; Generation shift keys; Electricity market modeling; Electricity market design; Congestion management.

JEL-Classification: L94 (Electric Utilities), Q40 (Energy – General), Q41 (Energy – Demand and Supply; Prices), Q43 (Energy and the Macroeconomy)

Highlights

- Flow-Based Market Coupling quite efficient without internal bottlenecks
- Flow-Based Market Coupling seems robust against forecast errors
- Internal bottlenecks decrease the efficiency of the mechanism substantially
- Regulatory changes (minRAM) significantly increase redispatch quantities and may increase costs
- Improvements to GSK methods mostly relevant for zones with few internal congestions

Contents

Abbreviations

CWE	Central Western Europe
Entso-E	European Network of Transmission System Operators
FBMC	Flow-Based Market Coupling
FRM	Flow reliability margin
GSK	Generation shift key
IVA	Individual validation adjustment
OPF	Optimal power flow
PTDF	Power transfer distribution factor
RAM	Remaining available margin
TSO	Transmission system operator
vRES	Variable renewable energy sources

Nomenclature

$a_{f,i}$	nodal power transfer distribution factor of line f corresponding to node i
$\tilde{a}_{f,z}$	zonal power transfer distribution factor of line f corresponding to zone z
β^+, β^-	cost factors for the redispatch model
γ	penalty term for the redispatch model
C_f	transmission line capacity of line f
c_i	marginal cost term for generation at node i
$\Delta F_f^{\mathrm{ref,(e)}}$	expected difference in reference flows on line f
$\Delta g_i^{+/-}$	change of generation at node i in the redispatching process
d_i	vertical load at node i
Δq_i	shift in nodal net export at node i
$\Delta \tilde{q}_z$	shift in zonal net export of zone z
Δt	duration of time step
$f \in \mathcal{F}$	index / set of transmission lines

F_f^{adj}	individual validation adjustment of line f
$\mathcal{F}^{\text{CNE}}$	set of critical network elements
g_i	electricity generation at node i
g_i^{max}	installed generation capacity at node i
$g_i^{*\text{z}}$	scheduled generation at node i (scheduled according to zonal market clearing)
$i \in I_{(z)}$	index / set of nodes (if with subscript z: set of nodes within zone z)
q_i	nodal net export at node i
$q_i^{(\text{e})}$	expected nodal export at node i
$\lambda_{z,i}^{(\text{p})}$	predetermined weight for zonal PTDF calculation at node i in zone z (= GSK)
M_f	flow reliability margin of f
N	number of nodes in the system
N_z	number of nodes in zone z
R_f^{nsfd}	RAM in non-standard flow direction
R_f^{sfd}	RAM in standard flow direction
$\tilde{q}_z$	zonal net exports of price zone z
$z \in Z$	index / set of price zones

1 Introduction

The rapid deployment of non-dispatchable renewable energy sources has placed system operators and market designers in a challenging situation. Now more than ever, the transmission of low-cost, potentially renewable electricity from its place of generation to the place of consumption is of paramount importance. Electricity transmission is governed by the laws of physics, and the quantities to be transmitted depend on where power plants are dispatched. To allow for as much low-cost generation as possible and, at the same time, ensure safe and reliable grid operation, it is therefore important to organize this power dispatch. For this purpose, two basic design alternatives for spot markets for electricity have been established over the last decades, which largely align with the two basic "power market architectures" discussed by Wilson (2002).

The first approach ("unbundled" according to Wilson's terminology) is based on zonally organized markets, where the same wholesale price of electricity prevails in each zone (usually country). This *zonal market design* is one cornerstone of the European target model for electricity market design (cf. Keay 2013). Since 2015, Flow-Based Market Coupling has been the method for implicit market coupling facilitating the exchange of electricity between markets of a total volume of around 1400 TWh (Austria, Belgium, Germany, France, Luxembourg, and the Netherlands). Its extension to Eastern Europe is planned for the end of 2020.

In contrast to the European approach, other big electricity markets, predominantly in the US, have opted for "integrated" markets long ago and accordingly made the switch to *nodal pricing*, where an individual price applies to each grid node.

Researchers have dedicated significant efforts to the question of advantages and disadvantages of zonal vs. nodal markets (Schweppe et al. 1988; Hogan 1992; Green 1997; Hogan 1999; Bjørndal and Jørnsten 2001; Ehrenmann and Smeers 2005) and, theoretically, the advantages of nodal pricing seem to be clear. In simple words, nodal pricing takes into account the physical realities of the electricity network most accurately and, thus, always yields the first-best solution in terms of operational system costs. In contrast, zonal pricing ensures a uniform price across each zone by abstracting from the physical reality of the grid. But how well does the specific implementation of the European market coupling actually perform compared to nodal pricing? On the one hand, the answer to this question is of great importance as an improved design of markets of this size can imply savings of several hundreds of millions of Euros per year. On the other hand, the absolute savings that the implementation of nodal pricing would entail serve as a reference value that needs to be balanced with the commonly asserted advantages of "unbundled" zonally organized markets such as stronger competitive incentives, reduced market power, and higher liquidity (cf. Wilson 2002; Consentec 2015; ACER 2018).

Apart from the efficiency of Flow-Based Market Coupling relative to nodal designs, the second issue of concern is the enhanced integration of European electricity markets. This is the proclaimed goal of the European Union (cf. ACER 2018) and, consequently, both improvements to the rules of Flow-Based Market Coupling as well as regulatory changes have been proposed or codified (cf. Van den Bergh and Delarue 2016; Dierstein 2017; Finck et al. 2018; Schönheit and Sikora 2018; Sebestyén et al. 2018; ACER 2019a; EU 2019; Schönheit et al. 2020). But what are the levers available to transmission system operators to effectively improve Flow-Based Market Coupling? And what are the impacts of planned regulatory changes on welfare in the European electricity market?

Others have shown inefficiencies of zonal market coupling approaches (Bjørndal and Jørnsten 2001; Ehrenmann and Smeers 2005) and analyzed different aspects of Flow-Based Market Coupling (Van den Bergh and Delarue 2016; Dierstein 2017; Finck et al. 2018; Schönheit and Sikora 2018; Sebestyén et al. 2018; Matthes et al. 2019; Wyrwoll et al. 2019; Schönheit et al. 2020). The first group of these papers uses stylized models to assess zonal market coupling procedures. These models are suitable for revealing defects of such a market design. However, in terms of quantification of welfare, they have two crucial shortcomings. First, they contain a large degree of simplification such as few power plants, few time steps, few load situations, etc. Evidently, this goes along with a significant loss of accuracy and representativeness. Second, the limited number of grid nodes considered in these studies makes the modeled power flows overly sensitive to dispatch changes, which tends to overstate inefficiencies of the market coupling processes (cf. Felten et al. 2019). The other group of papers considers somewhat larger data sets. Yet, they only look at the market clearing itself. But to accurately judge the efficiency of market coupling approaches, just looking at the market results is not enough. Redispatch, where transmission system operators adjust the dispatch of generators to ensure safe and reliable operation of the grid, has to be considered as well. To underline this statement, Germany's redispatch costs alone (including countertrading and curtailment of renewables) exceeded 1 billion Euros in 2018 (BNetzA 2019). Moreover, none of the existing papers that deal with Flow-Based Market Coupling in a realistic setting contrasts the market results with the first-best benchmark constituted by the nodal clearing results.

We have built a model framework to analyze the entire chain – from performing the capacity allocation process of Flow-Based Market Coupling, over determining the outcomes of the Central Western European electricity market clearing, to assessing the resulting need for redispatch (and the associated costs). We have designed the model framework to use identically structured optimization problems for zonal and nodal market clearings. By pre-processing grid-related input data, we can quantify the welfare of Flow-Based Market Coupling relative to benchmarks like nodal designs in a consistent manner. In order to assess real-world gains and losses, we use a realistic data set of the Central Western European transmission network. By conditioning the input data, different scenarios like idealized price zone configurations are assessed. In the same way, we scrutinize the effect of technical and regulatory changes to Flow-Based Market Coupling on the welfare in European electricity markets.

The remainder of this paper is structured as follows: Section 2 introduces the basic concepts of Flow-Based Market Coupling and nodal pricing. We also outline the relevant particularities of

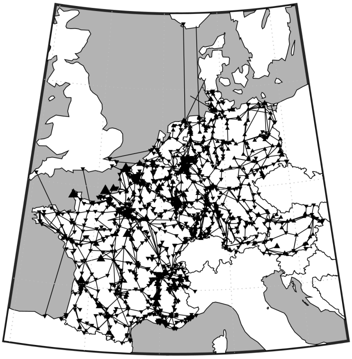


Figure 1: Illustration of information and power flow constraints that could be used in a nodal pricing regime.

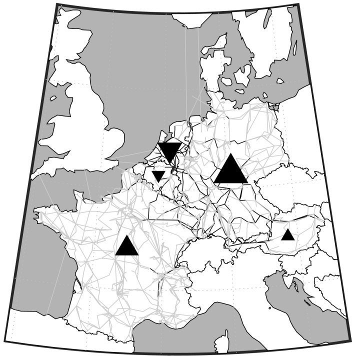


Figure 2: Illustration of information and power flow constraints used in a Flow-Based-Market-Coupling-style zonal pricing regime.

Flow-Based Market Coupling and our corresponding modeling approaches. Section 3 presents the numerical assessments which address the above research questions. Finally, we conclude with a discussion of the results and draw the main conclusions in Section 4.

2 Modeling Flow-Based Market Coupling

This section first presents the basic concepts behind nodal and zonal market designs in general and Flow-Based Market Coupling in particular. After the general description in Section 2.1, we present the model assumptions in Section 2.2 and derive the model equations in Section 2.3. Lastly, Section 2.4 details our evaluation approach.

2.1 Flow-Based Market Coupling and nodal designs

The European zonal design and nodal pricing regimes differ mainly in two aspects: The spatial granularity of dispatch information and the processes that complement the day-ahead market clearing.

The differences in granularity of dispatch information are visualized in Figure 1 and 2.

In Figure 1, triangles symbolize nodal net exports (upward facing triangles) and nodal net imports (downward-facing triangles).[1] A nodal net export is simply the balance of electricity generation of power plants and load of all consumers at a given grid node. Denoting the export at node i by q_i, the power flow on a transmission line f can be expressed by $\sum_{i\in I} a_{f,i}q_i$, where $a_{f,i}$ is the so-called Power Transfer Distribution Factor.[2] $a_{f,i}$ expresses how much more electricity is transmitted through line f, if 1 additional MW is injected at node i (and withdrawn at an arbitrarily-chosen reference node). Power flows may not exceed the corresponding transmission line's capacity C_f, which is a constraint explicitly considered in nodal pricing regimes.

In contrast, Flow-Based Market Coupling only considers net exports at the zonal level (cf. Figure 2). These zonal net exports $\tilde{q}_z$ are simply the sum of the nodal exports within one price zone ($\tilde{q}_z = \sum_{i\in I_z} q_i$, where I_z denotes the set of nodes within zone z). As only the zonally aggregated information is available to the market clearing entity, the power flow modeling is not as precise. It requires certain approximations and assumptions which are explained in Section 2.3.

Before explaining these approximations, we turn to the second main difference between the two market designs – the complementary processes. The sequential flow-based market coupling process starts with the capacity allocation process two days before delivery ("D-2"). This process is laid out in ACER (2019b) and determines the parts of the transmission lines' capacities that are available for trade. To this end, the responsible transmission system operators forecast the grid topology, generation, load, and exchange programs for the day of delivery with a common grid model (Entso-E 2016). In essence, the function of this process is to condition the feasible region of the market clearing problem around an operating point, which we will refer to as the base case (cf. Felten et al. 2019). Thereby, power flows due to expected intra-zonal exchanges can be considered, and line capacities that are already in use due to intra-zonal power flows and loop flows can be approximated. Doing so, Flow-Based Market Coupling features an approach to tackle some of the inherent zonal inaccuracies. More details on this process are presented in Section 2.3.3 and different implementation choices are evaluated in Section 3.

Subsequently, one day ahead of delivery ("D-1"), the market clearing is performed, using the outcomes of the capacity allocation process at the D-2 stage. Markets are cleared with the objective of welfare maximization respecting line capacity constraints imposed on zonal net exports (cf. Nemo-C 2019).[3] A further source of inefficiency compared to nodal settings is the reliance on information determined at D-2, which may have changed in between. The impact of such process-induced uncertainties will also be investigated in Section 3.

On the day of delivery ("D"), the power plants are dispatched according to the schedules determined by the market clearing. However, there is one major exception to this procedure. If the market result is physically infeasible, the dispatch has to be adjusted by redispatch actions in order

[1] For brevity, we only refer to net exports henceforth, implying that negative net exports are understood as net imports.
[2] For this formulation, some simplifying assumptions are used, which are explained in Section 2.2.
[3] Note that there is also some trading ongoing during the day of delivery, called Intraday trading. We omit this trading subsequently as it helps in processing information updates after D-1, notably on renewable forecasts, but does not change the handling of grid constraints.

to prevent congestion or to sustain n-1 security. As illustrated in Section 1, redispatch has become a crucial process in the European zonal market design and is an essential part of the assessments in Section 3.

In contrast to zonal pricing, the adjacent processes described above are not relevant for nodal pricing. As the grid is accurately represented in the market clearing, no prior capacity allocation has to occur, and redispatch is unnecessary since the market result is always physically feasible.

2.2 Assumptions

Throughout the entire paper, we assume effective competition, implying that generators bid at their marginal costs and do not bid strategically. Furthermore, we assume inelastic demand. That is to say that the market clearing process, which has the goal of welfare maximization, can also be expressed as a cost minimization.

In terms of power plant dispatch, we assume that all power plants are available for dispatch and redispatch. Moreover, we neither consider minimum generation, intertemporal constraints (e.g., reservoir filling levels, minimum downtimes / operation times, etc.) nor must-run restrictions (e.g., minimum operation from combined heat and power plants or the like). For hydro power plants having the flexibility of reservoirs, we consider shadow prices in daily resolution, which makes their dispatch rationale comparable to the one of thermal power plants. Neglecting intertemporal constraints entails some inaccuracies. However, it allows us to establish ceteris paribus comparisons of the Flow-Based Market Coupling (FBMC) zonal market design to contrast the zonal solution with the first-best solution (i.e., the solution under a nodal market design). A further assumption is that non-dispatchable renewables-based generation (from solar and wind) and generation of units connected to lower voltage levels are directly considered in the vertical load d_i. Hence, this generation is completely fed into the grid as long as technically possible.[4]

Concerning the calculation of the flow-based parameters, we assume that the modeled state of the grid is already optimized and that there are no more non-costly remedial actions available. Additionally, we assume that there are no long-term allocated capacities. The corresponding intermediate steps in the official guidelines (ACER (2019b)) are therefore skipped. We further assume that there are no coordinated validation adjustments.

In terms of power flow modeling, we use the DC-lossless assumption. Additionally, we do not consider topology changes of the grid. Thus, the power flows can be modeled as a set of linear equations using power transfer distribution factors (PTDF / $a_{f,i}$).[5] Grid security considerations are approximated by reducing the thermal capacity of transmission lines to 85 % of their nominal values. In terms of commercial exchanges (cross-zonal transactions), we do not consider long-term

[4] Curtailment is only allowed at relatively high penalties. Here, we set costs to 100 EUR / MWh to account for compensation of lost subsidies.

[5] For an explanation of these PTDFs, see Section 2.3.

nominations. Thus, the entire exchange results from the market clearing (and potential adjustments through redispatch). We simplify the stages of the market clearing by considering the day-ahead and intraday trading stages as one common step. That is to say, only the effect of foresight deviations for formulating grid constraints is considered. In terms of dispatch optimization (given the predetermined grid constraints), foresight is always assumed to be perfect.

2.3 Deriving the electricity market clearing problems

We subsequently first discuss the objective function for both zonal and nodal market clearing followed by the discussion of corresponding constraints. Thereafter, the modeling of the base case and the redispatch approach are explained.

2.3.1 Common terms in the nodal and zonal market clearing problems

The objective function in both cases is given by Equation (1). It consists of the minimization of the system costs (i.e., the sum of the marginal generation costs c_i multiplied by the relevant electricity output g_i at node i for all nodes $i \in I$). The marginal costs c_i are a constant term for generation g_i at each node.[6] Δt denotes the duration of the considered time step.[7]

Equation (2) simply expresses the convention that the surplus of electricity output (i.e., generation g_i minus vertical load d_i at node i) constitutes the nodal net export q_i. Equation (3) assures that the overall generation meets demand. Equation (4) expresses the upper and lower generation limits at all nodes. Thus, the limiting value $g_i^{\max}$ is the aggregate generation capacity of the generation unit at node i.

$$\min_{g_i} \sum_{i \in I} c_i g_i \Delta t \tag{1}$$

$$\text{s.t.} \quad q_i = g_i - d_i \qquad \forall i \in I \tag{2}$$

$$\sum_{i \in I} q_i = 0 \tag{3}$$

$$0 \leq g_i \leq g_i^{\max} \qquad \forall i \in I \tag{4}$$

As stated above, Equation (1) to (4) apply no matter whether market clearing is implemented on a nodal or zonal basis.

[6]We assume at most one generator per node. Per topology change, a network with multiple generators per node may be converted to an equivalent one that fits this restriction.

[7]Throughout this paper, we consider hourly time steps.

2.3.2 Constraints in the electricity market clearing problems

The most important processes in Flow-Based Market Coupling take place at the D-1 and D-2 stages. The parameters needed to perform Flow-Based Market Coupling (also called *flow-based parameters*) are determined at the D-2 stage while the market clearing itself happens at the D-1 stage. In this section, we will show how the flow-based parameters are approximated from the physically accurate nodal constraints. The nodal power flow constraints can be written as:

$$-C_f \leq \sum_{i \in I} a_{f,i} q_i \leq C_f \;\; \forall f \in \mathcal{F} \tag{5}$$

Herein, C_f denotes the transmission capacity of line f being element of the set of all transmission lines $\mathcal{F}$. The inner term describes the actual power flow on line f, with the power transfer distribution factor (PTDF) $a_{f,i}$ and the nodal net positions q_i. The power flow constraint in Flow-Based Market Coupling takes a similar form:

$$R_f^{\text{nsfd}} \leq \sum_{z \in Z} \tilde{a}_{f,z} \tilde{q}_z \leq R_f^{\text{sfd}} \quad \forall f \in \mathcal{F}^{\text{CNE}} \tag{6}$$

As explained in Section 2.1, the market clearing under Flow-Based Market Coupling only considers zonal net exports $\tilde{q}_z$. The sensitivity of power flows to these zonal net exports is approximated by zonal power transfer distribution factors $\tilde{a}_{f,z}$. Basically, these zonal power transfer distribution factors are weighted averages of the nodal factors $a_{f,i}$:

$$\tilde{a}_{f,z} = \sum_{i \in I_z} \lambda_{z,i}^{(\text{p})} a_{f,i} \tag{7}$$

To arrive at $\tilde{a}_{f,z}$, each $a_{f,i}$ of a node i in zone z is weighted with a so-called *generation shift key* $\lambda_{z,i}^{(\text{p})}$ (GSK). For deriving such generation shift keys, various methods exist (cf. Amprion et al. 2014; Van den Bergh and Delarue 2016; Dierstein 2017; Entso-E 2017), three of which we assess in this paper:

- *by capacity*: The $a_{f,i}$s are weighted with the share of dispatchable power plant capacities of each node in each zone.
- *by N*: The $a_{f,i}$s are weighted in inverse proportion to the number of nodes in one zone.
- *by NEX*: The $a_{f,i}$s are weighted with the nodal shares of zonal net exports expected in the base case.

Table 1: Modeled GSK determination procedures. In the equations, it is implied that $i \in I_z$.

weighting method	formal description	abbreviation
by installed capacity	$\lambda_{z,i}^{(\mathrm{p})} = \frac{g_i^{\max}}{\sum_{i \in I_z} g_i^{\max}}$	by capacity
by number of nodes	$\lambda_{z,i}^{(\mathrm{p})} = \frac{1}{N_z}$	by N
by net exports	$\lambda_{z,i}^{(\mathrm{p})} = \frac{q_i^{(\mathrm{e})}}{\sum_{i \in I_z} q_i^{(\mathrm{e})}}$	by NEX

Table 1 provides a formal description of the modeled GSK methods. Here, I_z denotes the set of nodes i inside of zone z, superscript (p) indicates that the GSKs are prior to market clearing. Furthermore, $g_i^{\max}$denotes the installed electric capacity on node i and $q_i^{(\mathrm{e})}$ is the expected net export of node i. N_z is the number of nodes within zone z.

The lower and upper limits imposed on the line flows in Equation (6) are the so-called *remaining available margins* (RAM) in standard flow direction R_f^{sfd} and non-standard flow direction R_f^{nsfd}. These are composed of several elements, summarized here for R_f^{sfd} as follows:

R_f^{sfd} = Line capacity
− power flow according to the base case expectation
+ power flow calculated zonally using the base case zonal net exports
− validation adjustments
− flow reliability margin

Consequently, RAMs can be calculated as follows (cf. ACER 2019b):

$$R_f^{\mathrm{sfd}} = C_f - \Delta F_f^{\mathrm{ref,(e)}} - F_f^{\mathrm{adj}} - M_f \tag{8}$$

$$R_f^{\mathrm{nsfd}} = -C_f - \Delta F_f^{\mathrm{ref,(e)}} + F_f^{\mathrm{adj}} + M_f \tag{9}$$

Besides the capacity C_f of the lines (which is already part of the nodal market clearing), RAMs also reflect the difference in reference flows on the corresponding line ($\Delta F_f^{\mathrm{ref,(e)}}$), an individual validation adjustment (IVA / F_f^{adj}) and a so-called flow reliability margin (FRM / M_f). Given the assumptions detailed in Section 2.2, the difference in reference flows $\Delta F_f^{\mathrm{ref,(e)}}$ is the power flow at expected market outcome reduced by the flow that the zonal power flow approximation implies for the same expected market outcomes (cf. ACER 2019b).

$$\Delta F_f^{\mathrm{ref,(e)}} = \sum_{i \in I} a_{f,i} q_i^{(\mathrm{e})} - \sum_{z \in Z} \left(\tilde{a}_{f,z} \sum_{i \in I_z} q_i^{(\mathrm{e})} \right) \tag{10}$$

Here, the superscript (e) indicates that $q_i^{(e)}$ is an expected quantity, i.e., it is determined in accordance with the expectations at the D-2 stage (cf. Section 2.1).[8] The IVA is a term that can be used to reduce remaining available margins if operational security limits may be exceeded.[9] FRMs take into account the inherent uncertainties of the zonal FBMC process (external exchanges, approximations of the FBMC procedures and differences between forecast and realized programs, (cf. ACER 2019b)).

Recently, regulators have modified the procedures to be more restrictive in the way the flow-based parameters are calculated (cf. ACER 2019a). In principle, experts and regulators can influence the power flow constraints in two basic ways:

1. Using individual validation adjustments and flow reliability margins: Increasing these values tends to reduce redispatch.

2. Equation (6) applies to a subset of transmission lines $\mathcal{F}^{\text{CNE}} \subseteq \mathcal{F}$, which are called critical network elements. For the longest time, the so-called 5 %-rule has constituted the procedure of choice to determine this subset. That is, if the absolute value of the subtraction of any two zonal PTDFs of a line exceeds 5 %, the line is considered to be critical. However, especially whether to include intra-zonal lines in this set or not is a frequently discussed question. Regulators tend towards exclusion, making inclusion in the future subject to special approval (ACER 2019b).

The RAMs are calculated according to Equation (8) and (9). However, to avoid convergence issues we limit the RAM to zero, such that $R_f^{\text{nsfd}} \leq 0 \leq R_f^{\text{sfd}}$.

2.3.3 Modeling the base case

In Section 2.1, we have outlined the stage-wise FBMC process. The FBMC parameters (RAMs and PTDFs) are determined at the D-2 stage with a common grid model, which we refer to as the base case. The base case comprises the best estimate of the situation on the day of delivery, which includes forecasts for variable renewable energy sources (vRES) infeed, electricity exchanges, load, and generation. Usually, this is done based on historical data, e.g. reference situations or reference days. All FBMC parameters that are introduced in the following subsections result from this base case estimation. Because in the real-world process, the base case definition is based on best estimates of TSOs, a reproduction of the exact procedures is hardly possible. Thus, we apply a methodology that first derives an estimated state of the electricity system and, on the basis of that state, calculates the FBMC parameters (cf. Figure 3). The system state estimation consists of three steps. It starts with executing an optimal power flow (OPF) calculation. Based on the OPF solution, RAMs and GSKs are calculated and a zonal clearing is simulated. The results of this zonal

[8] We distinguish quantities that are expected ((e)) and those which are predetermined according to heuristic procedures ((p)).
[9] Until recently, the IVAs were called final adjustment values (FAV).

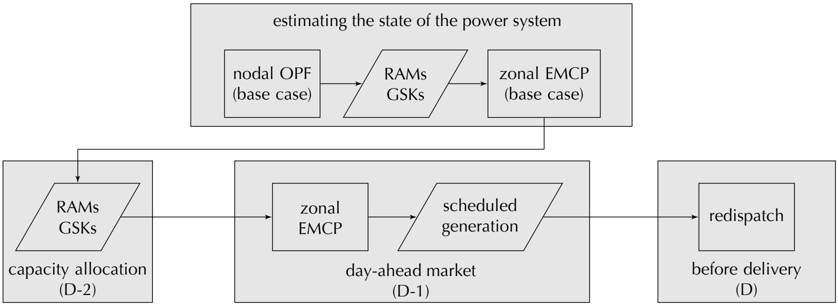


Figure 3: Flow chart for complete simulation run

clearing are used as the *base case*. As the RAMs and, in some cases, the GSKs of the second step are different from those of the final zonal clearing, the base case will diverge from the final market clearing result to some extent. However, a similar divergence will be observed for the real-world estimation procedure. The procedure is visualized in Figure 3.

While we do model imperfect foresight in the base case in some of the scenarios, we always assume perfect foresight from the point of day-ahead market clearing.[10] This means that the day-ahead market results contain no uncertainties apart from those potentially introduced by the capacity allocation process (i.e., contained in the RAMs). Consequently, the use of intraday markets to balance forecast errors is not relevant here and subsequently disregarded. It also follows that the results from the nodal optimization are directly comparable with the results from the zonal optimization including redispatch.

2.3.4 Redispatch

The aggregation in the zonal market design[11] leads to inaccuracies concerning the physical state of the grid. Because of these inaccuracies, feasible solutions of the zonal market clearing will – in many cases – not be in line with the physical realities of the electricity grid. In these cases, *redispatching* of generators will be required. Redispatching entails increasing the electricity generation of a power plant on one side of a congested element and decreasing the electricity generation of a different power plant on the other side, thereby changing the power flow across the congested network element. The costs of the additional generation by the power plant increasing its load will partly be offset by the cost savings of the other power plant decreasing its generation. However, the power plant with the increase in generation will usually be more expensive than the power plant with decreasing generation. Otherwise, it would have generated from the beginning.

[10] This holds for both the zonal as well as for the nodal market clearing, which we will use as a benchmark.
[11] With this statement, we refer to the zonal power flow approximation and associated processes.

The starting point for formulating the objective function of the redispatch problem and its constraints is the nodal problem as described in Equation (1) - (5). However, instead of optimizing the system from zero, the approach is to fix the generation as calculated in the zonal optimization and introducing the new variables Δg_i^+ and Δg_i^- to denote the change of generation in the redispatch process in both positive and negative direction.[12] The division into separate variables for positive and negative changes is done to enable differing costs for positive and negative changes while preserving the linear nature of the objective function. Fixing the generation of the zonal optimization is accomplished by subtracting the scheduled generation g_i^{*z} from the demand on each node d_i.[13] The factors β^+ and β^- describe relative differences in the costs used during the redispatch compared to the true costs. This reflects inefficiencies in the non-market-based redispatch process compared to truthful bidding in competitive markets. γ is added to every cost term for positive changes in generation and subtracted from every cost term for negative changes in generation. This puts a penalty on any redispatch quantity which is reflective of the main focus on minimizing redispatch quantities instead of costs in the process. When considering the actual values for system costs, γ is excluded to reflect only actual costs. The factors β^+ and β^- represent actual inefficiencies of this non-market-based approach and are therefore not excluded from cost considerations.[14]

The redispatch problem thus becomes:

$$\min_{\Delta g_i^+, \Delta g_i^-} \sum_{i \in I} \left((\beta^+ c_i + \gamma) \Delta g_i^+ - (\beta^- c_i - \gamma) \Delta g_i^- \right) \Delta t \tag{11}$$

$$\text{s.t.} \quad q_i = \left(\Delta g_i^+ - \Delta g_i^-\right) - \left(d_i - g_i^{*z}\right) \qquad \forall i \in I \tag{12}$$

$$\sum_{i \in I} q_i = 0 \tag{13}$$

$$g_i^{\min} - g_i^{*z} \leq \left(\Delta g_i^+ - \Delta g_i^-\right) \leq g_i^{\max} - g_i^{*z} \qquad \forall i \in I \tag{14}$$

The power flow constraints in Equation (5) stay unchanged.

It should be noted that without including penalty factors the result of the zonal optimization followed by redispatch always leads exactly to the nodal solution in costs.[15] This is because the nodal solution is the most cost-efficient way to satisfy all constraints.

[12] As in the original problem, we assume at most one generator per node (cf. Section 2.3.1).

[13] The scheduled generation is the result of the zonal market clearing (indicated by $^{*}z$).

[14] For the numerical assessment, we set the proportional factors to $\beta^+ = 1.3$ and $\beta^- = 0.8$. This means that increasing production of a power plant costs more than decreasing production of the same power plant returns. The blanket penalty term γ is set to 300 EUR / MW. Results for a sensitivity analysis with the blanket penalty term, but with proportional factors of $\beta^+ = \beta^- = 1$, i.e. without a cost penalty, are shown in Appendix A.3 and are referred to in the main text where relevant.

[15] In terms of quantities, it is possible that different dispatches achieve the same minimal costs.

2.4 Assessing the features and proposed adjustments of Flow-Based Market Coupling

In this section, we describe the methodology by which we assess the performance of Flow-Based Market Coupling and the influence of the flow-based parameters. Note that we use the same price zone configuration throughout the entire paper, which corresponds to the current configuration in the extended CWE region.[16]

2.4.1 Assessment runs

The numerical analyses contain multiple assessments. First, we consider idealized price zones without internal congestion. Second, we consider imperfect price zones, using a realistic grid model with internal congestions. We assess the overall performance of both approaches by comparing the performance of Flow-Based Market Coupling with the reference GSK-method to nodal pricing and the unlimited trade scenario (see Section 3.1.1). In addition to the overall assessment, we present the effects of using different methods for calculating GSKs and the effects of uncertainties introduced by the process of calculating the flow-based parameters (see Section 2.4.3).

Finally, we assess the impacts of two different ways to deal with internal congestions: TSOs have been accused of reducing RAMs as a way of limiting internal congestions in the past (cf. ACER 2017). On the other hand, new regulation will be implemented to address this behavior. Both types of adjustments to the capacity allocation process are detailed in Section 2.4.4.

2.4.2 Idealized and imperfect price zones

In the first sequence of numerical assessments (cf. Section 3.1.1 to 3.1.2), we choose to analyze FBMC in a setting without intra-zonal bottlenecks. Each price zone then corresponds to a "copperplate". This is done by setting the thermal capacities of intra-zonal lines to a sufficiently large value. Then, only constraints of cross-border lines can become binding. Evidently, this is far from reality. However, it provides insights into the performance of FBMC in an ideal setting. In contrast to this ideal setting, we also analyze a realistic setting in which all transmission lines are modeled with their thermal capacities expected for the year 2020. In terms of defining the set of critical network elements, we apply the 5 % rule described in Section 2.3. This allows us to evaluate the effect of intra-zonal bottlenecks both on the overall performance of FBMC and on the leverage of the different FBMC varieties. Therefore, we calculate all of the sensitivities introduced in Section 2.4.3 with both the idealized and the realistic grid setting.

[16]This comprises all countries where Flow-Based Market Coupling is currently in effect – namely Germany, Austria, the Netherlands, Belgium, France – and additionally Switzerland. Even though, in practice, Switzerland does not participate in FBMC, in our model, we assume FBMC to be applicable in all modeled countries.

2.4.3 Procedural sensitivities

Generation shift keys: In Section 2.3, we have highlighted that various options for GSK calculation exist. We assess the effects of using different GSK methods by executing the mathematical operations from Table 1 and undertaking the model sequence in Figure 3.

Process-induced uncertainties: The base case, as introduced in Section 2.3.3, is usually based on historical data enhanced by the expertise of TSOs. This naturally entails forecast errors. The FBMC parameters that are probably influenced most by forecast errors are the RAMs. Their calculation requires forecasts for the calculation of the term $\Delta F_f^{\text{ref},(e)}$, which involves the expected nodal net exports $q_i^{(e)}$. Main drivers of uncertainties in $q_i^{(e)}$ are the infeed from vRES, load variations and unexpected power plant outages. Thus, if realized net exports deviate from their expected values, this entails inadequacies in power flow constraints. In addition, forecasting procedures of these values, making use of historical data, may also be a source of uncertainty.

Other values that are subject to inaccuracies are the zonal line load sensitivities (i.e., $\tilde{a}_{f,z}$). GSKs can be expected to be one major source of inadequacy. As the considered GSK determination methods are rather heuristic-based than forecast-based procedures, we assess forecast errors and GSK procedures separately.

To show the effects of process-induced uncertainties, we introduce forecast errors in the base case calculation.[17] In particular, we use imperfect onshore and offshore wind forecasts for calculating the estimate of the state of the power system (cf. Figure 3). We approximate the error in wind forecasting by using actual forecast errors published on the Entso-E transparency platform (Entso-E 2018). We calculate the errors relative to the installed capacity per country and apply the relative errors to each node according to the node's installed wind capacity. The absolute error is limited by the installed capacity at each node, so that forecast wind generation at each node cannot exceed the installed capacity nor drop below zero.

The forecast errors affect the calculation of the flow-based parameters twofold:

1. The calculation of GSKs using the *by NEX* method is affected, because this method relies on the nodal net positions and these net positions depend on the vertical load at each node. The nodal vertical load is affected at each node where wind forecast errors are present. The other GSK methods are not affected, as they are static and therefore do not change with different base cases.
2. RAMs can be affected as well, as the calculation of $\Delta F_f^{\text{ref},(e)}$ depends on the nodal net positions (cf. Equation (10)). However, in the case of the by NEX GSK method, RAMs will not be affected, as $\Delta F_f^{\text{ref},(e)}$ is always zero when using the by NEX method.

[17]The forecast errors are added to the nodal as well as the first zonal market clearing, which acts as the base case for the subsequent zonal clearing.

This means that depending on the GSK method either RAMs or GSKs will be affected by the uncertainty but never both.

2.4.4 Adjustments to the capacity allocation process

As described in Section 2.3, regulators have recently taken a stricter stance on modifications to the remaining available margins. However, this was not always the case (cf. ACER 2017; Amprion 2017; CREG 2017). Section 2.3 has highlighted that capacity adjustments of both intra-zonal and inter-zonal lines can be used to either foster trade between market participants (increase RAMs) or decrease redispatch (decrease RAMs). Thus, we assess the four following sensitivities to evaluate the effects of capacity adjustments:

1. To simulate the effect of reducing the RAMs on critical network elements to decrease redispatch, we reduce the RAMs of the 10 most overloaded transmission lines by 25 % of the thermal line capacities. We label that case ″individual contingency margins (overall)″, since it includes a discretionary choice of RAM modifications.

2. As a slight variation of the previous sensitivity, we reduce the RAMs of the 10 most overloaded cross-border lines by 25 % of the thermal line capacities. This case is labeled ″individual contingency margins (cross-border)″.

3. In line with the new regulation in ACER (2019a), we exclude all intra-zonal lines from the set of critical network elements $\mathcal{F}^{\text{CNE}}$ (cf. Equation 6). Yet, line capacities remain unchanged from their realistic values, and these lines are therefore relevant in the nodal market clearing and in the redispatch stage of the FBMC process.

4. To model the new electricity regulation of the European Union to guarantee a minimum size of of RAMs (cf. ACER 2019a; EU 2019), we assess a sensitivity that ensures at least 70 % free capacity on all critical network elements.

5. Finally, we assess the combination of sensitivities 3 and 4.

Because intra-zonal congestions are the main reason for these adjustments, we restrict this analysis to the realistic grid scenario. The results of these investigations are presented in Section 3.3.

2.4.5 Statistical Analysis

When observing differences (e.g., in annual market clearing costs, redispatch costs, and amounts) between different cases, the question arises whether these are significant or not. This is especially the case where the total overall cost difference is small and the direction of the difference for individual time steps varies. Whether the observed differences are significant can be tested using statistical methods – if we interpret the 8760 hourly simulation results as outcomes of a stochastic multivariate process. In fact, input time series of the optimization problem like demand and vRES infeed can be viewed as auto-correlated time series with time-varying mean and variance. And the optimization results are transforms (″derivatives″ in finance language) of these data and hence stochastic processes themselves. The same holds for the differences between results obtained under different settings. Hence, we can test for statistical significance using Newey-West adjusted standard errors to account for the autocorrelation and heteroscedasticity present in time series data.[18] An overview of the corresponding results can be found in Figure 8 and Figure 9.

3 Numerical assessments for Central Western Europe

The subsequent sections present the numerical assessments. As discussed in the previous section, we consider two different grid scenarios. We start with an idealized setting in Section 3.1 where we assume national "copperplates" (i.e., price zones without internal transmission bottlenecks). Thereafter, we consider a realistic grid scenario where intra-zonal congestion is relevant (cf. Section 3.2). For each of the scenarios, we evaluate the overall performance of flow-based market coupling compared to nodal pricing. Additionally, results for different methods of determining generation shift keys and the effects of uncertainty in the base case are presented as sensitivities. Finally, we consider adjustments to the capacity allocation, either due to regulatory measures or efforts to reduce redispatch undertaken by transmission system operators (cf. Section 3.3).

As stated in the introduction, comprehensive consideration of both market and redispatch costs is indispensable. A more constrained FBMC parametrization might cause market clearing costs to increase while costs for redispatch might decrease. Only the sum of both costs gives an indication regarding the overall performance of different FBMC configurations. Thus, we always present both the overall system costs and the subdivision in market clearing and redispatch costs. Aside from costs we also analyze redispatch quantities. Details on these results are given in Appendix A.1. Furthermore, results for a sensitivity analysis with a different parametrization of redispatch costs are presented in Appendix A.3. Where relevant, these results will also be referred to in the main text.

[18] The lag is set by rounding up $L = T^{\frac{1}{4}}$ to the nearest integer, with T being the sample size (8760 for a full year with no non-converging hours) (see Greene (2012, p. 960)).

3.1 Idealized price zones

3.1.1 Overall performance of Flow-Based Market Coupling

To classify the performance of Flow-Based Market Coupling, we contrast it to two extremes.[19]

On one hand, we compare system costs to those of a nodal pricing set-up as the theoretical optimum. On the other hand, we compare results to the "unlimited trade" scenario. This market clearing configuration may be seen as an extreme version of the "unbundled" market architecture discussed by Wilson (2002). It assumes no constraints for trade on the day-ahead market – which means that congestions are not managed at all in the market clearing – and completely relies on redispatch for the relief of congestions. This scenario serves as the lower benchmark in terms of welfare. In combination with the nodal solution, this provides a range in which the performance of FBMC can be assessed on a relative scale. 0 % efficiency corresponds to the solution of the unlimited trade scenario, 100 % efficiency is per definition equivalent to the nodal solution. As mentioned above, for all assessments in this section, we presume the idealized grid setting without intra-zonal bottlenecks.

Figure 4 provides a depiction of the system costs for nodal pricing, Flow-Based Market Coupling and unlimited trade. These system costs are composed of market clearing costs, being the operational costs at the D-1 stage, and redispatch costs. For the unlimited trade and Flow-Based Market Coupling scenarios, redispatch leads to additional costs due to inefficiencies of the non-market based approach (cf. Section 2.3.4). By contrast, nodal pricing does not incur additional redispatch costs, as the grid is accurately represented in the market clearing.

On the one side, "extremely unbundled" market clearing costs are lowest for the unlimited trade scenario because market coupling is unconstrained. Yet, there is a trade-off with redispatch costs, which amount to over 1 billion Euros. On the other side, "integrated" nodal clearing costs exceed the clearing costs in the unlimited case by about 200 million Euros, but the clearing is efficient. It does not necessitate redispatch, and overall system costs are lowest.

The total cost difference between unlimited trade and nodal pricing amounts to 825 million Euros. Clearing costs for Flow-Based Market Coupling are higher than costs for nodal pricing, while redispatch costs are quite low (38 million Euros) compared to unlimited trade. Generally, zonal clearing costs are expected to be lower than clearing costs in nodal pricing, because the grid does not constrain the clearing algorithm as much. However, it is not guaranteed that all feasible solutions within nodal pricing are also part of the flow-based domain: the approximations based on the expected market result may constrain the algorithm where nodal pricing could lead to a more fine-grained solution. Consequently, in this scenario, where the level of congestion is very

[19] As the reference case for Flow-Based Market Coupling, we choose a setting which uses the *by capacity* GSK method, is based on perfect foresight and uses the 5 % rule as the selection criterion for identifying critical network elements (where applicable).

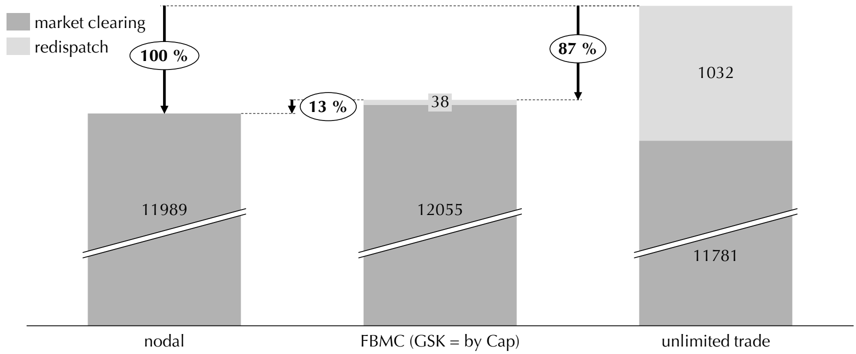


Figure 4: Overall comparison of system costs for the idealized setting. Costs in million EUR.

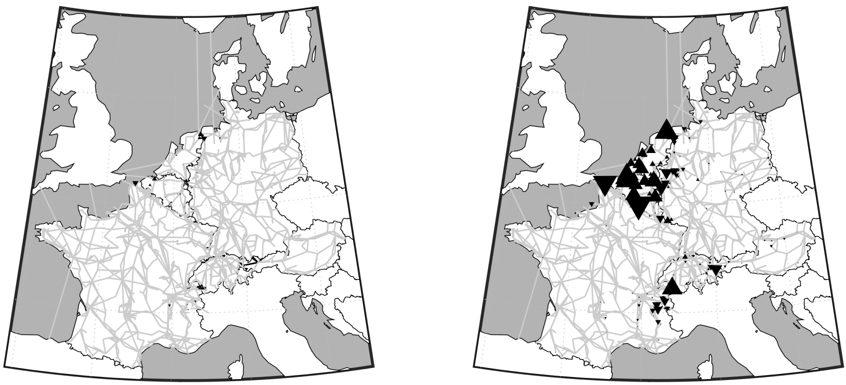


Figure 5: Sum of redispatch per node. Left: Flow-Based Market Coupling with *by capacity* generation shift keys. Right: Unlimited trade. Critical network elements in black.

low, these approximations actually lead to overall higher market clearing costs for Flow-based Market Coupling compared to nodal pricing.

The difference in redispatch costs is related to redispatched quantities. A comparison of these quantities at nodal level is given in Figure 5. Therein, up- and downward pointing triangles indicate positive and negative redispatch of power plants, respectively, while their sizes correspond to the sum of yearly quantities.

The necessity for redispatch is much higher in the unlimited trade scenario, which results in the aforementioned significantly higher redispatch costs. The total cost difference between nodal pricing and Flow-Based Market Coupling is 105 million Euros, which means that Flow-Based Market Coupling saves 720 million Euros compared to unlimited trade.

Table 2: Results with idealized price zones. Costs in million EUR.

no.	description	total cost			cost difference clearing + RD		
		clearing	RD	total	vs. nodal	vs. ref	vs. perfect foresight
1	nodal	11 989	0	11 989	0	-	-
2	unlimited trade	11 781	1032	12 813	825	719.6	-
perfect foresight:							
3	by capacity (= ref)	12 055	38	12 094	105	0.0	-
4	by NEX	12 018	196	12 214	225	120.4	-
5	by N	12 065	45	12 110	121	16.3	-
process-induced uncertainty:							
6	by capacity	12 057	42	12 098	110	4.7	4.7
7	by NEX	12 024	193	12 217	229	123.7	3.4
8	by N	12 065	49	12 115	126	20.9	4.6

Treating the 825 million Euros that nodal pricing saves compared to the unlimited trade scenario as the benchmark, the savings of 720 million Euros represent 87 % of what is theoretically achievable.

3.1.2 Procedural sensitivities

In Section 2, we have introduced the main characteristics of Flow-Based Market Coupling and the flow-based parameters. One of these parameters are the generation shift keys. Additionally, the stage-wise process of Flow-Based Market Coupling poses the question of whether uncertainties that are introduced in the capacity allocation process are relevant to its performance. Table 2 shows the results for three different generation shift key methods with perfect and imperfect foresight, the latter of which we refer to as process-induced uncertainty. The results from the overall evaluation in Section 3.1.1 are provided as a reference. Besides the total costs for market clearings and for redispatch, the table also shows the cost difference between the summed costs of the zonal clearing with redispatch and the nodal clearing. To facilitate comparisons between the different scenarios, one column always shows the total cost difference to the reference scenario (with *by capacity* generation shift keys).

Regarding overall costs, there are significant differences between the generation shift key methods. The *by N* method gets within 16 million Euros of the reference case. However, the *by NEX* method produces lower clearing costs than both other methods, but redispatch costs are higher by over 150 million Euros, resulting in a 120 million Euros difference in total costs compared to the reference. By having a closer look at the implications of the *by NEX* method, we obtain the explanation: The generation shift keys are calculated as the share of nodal exports from zonal exports. This means that non-zero generation shift keys are assigned even to nodes without generation units participating in the market clearing. Additionally, positive generation in zones with negative exports leads to negative generation shift keys being assigned to the nodes of these generation units.

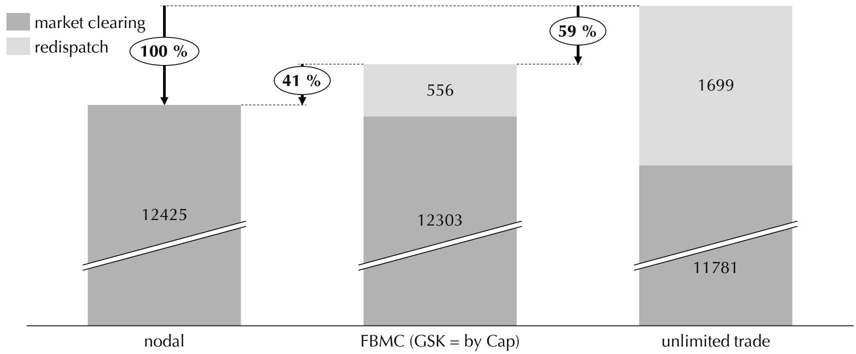


Figure 6: Overall comparison for imperfect bidding zones. Costs in million EUR.

Consequently, Flow-Based Market Coupling considers these generation units to contribute to line loadings in the opposite direction compared to the physical reality. Both effects distort the line loading approximations made by Flow-Based Market Coupling, ultimately entailing much higher overall costs.

In contrast to the variation of generation shift keys, the consideration of process-induced uncertainties through the introduction of forecast errors as described in Section 2.4.3 shows less significant results (cf. Table 2, lower part). Costs increase by 3.4 to 4.7 million Euros depending on the considered generation shift key method. The smallest increase in costs is observed for the *by NEX* method, however, its overall costs are still significantly higher than the costs of the other sensitivities.

3.2 Imperfect price zones

So far, we have investigated all scenarios assuming national copperplates. Now, we depart from this idealizing assumption. In order to assess the influence of imperfect price zones (i.e., zones with internal transmission bottlenecks), we now consider the actual thermal capacities of intra-zonal branches. The grid state is thus aligned with the reality in Central Western Europe for the reference year 2020.

3.2.1 Overall performance of Flow-Based Market Coupling

We start by recalculating our initial assessment of Section 3.1.1, consisting of the nodal set-up, the reference case using Flow-Based Market Coupling with *by capacity* generation shift keys and the unlimited trade scenario with the new grid setting. Figure 6 and Figure 7 show the results.

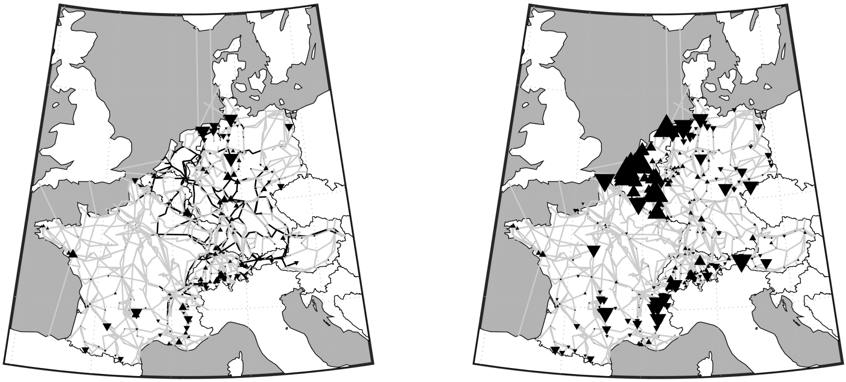


Figure 7: Sum of redispatch per node with realistic thermal capacities. Left: Flow-Based Market Coupling with *by capacity* generation shift keys. Right: Unlimited trade. Critical network elements in black.

As expected, system costs in both the nodal pricing as well as the unlimited trade scenarios increase because of the additional congestion. However, while nodal costs increase by 436 million Euros, costs of the unlimited trade scenario increase by 667 million Euros – a difference of 231 million Euros. Because the market clearing results in the unlimited trade scenario remain unchanged, this increase is directly related to higher redispatch costs. For Flow-Based Market Coupling, costs increase by 766 million Euros.

The larger increase in total costs in the Flow-Based Market Coupling approach compared to both the nodal as well as the unlimited trade scenario means that the relative cost-effectiveness of Flow-Based Market Coupling decreases. Nodal pricing achieves savings of 1055 million Euros compared to unlimited trade, whereas Flow-Based Market Coupling achieves 620 million Euros. Consequently, while Flow-Based Market Coupling applied with price zones *without* internal congestions achieved a relative cost-effectiveness of 87 % in reference to the cost difference between nodal and unlimited trade (cf. Section 3.1.1), this value drops to 59 % for the investigated internally congested grid.

3.2.2 Procedural sensitivities

The results shown above point to a large negative impact that imperfect price zones have on the performance of Flow-Based Market Coupling in terms of welfare. Additional impacts of imperfect price zones become apparent from regarding the further results. As Table 3 shows, cost differences between the different generation shift key methods decrease significantly. While the *by NEX* method constituted by far the most expensive method in Section 3.1.2 with an increase of 120 million Euros in reference to the *by capacity* method, the cost increase is reduced to

Table 3: Results with imperfect price zones (reference grid expansion). Costs in million EUR. RD = redispatch

no.	description	total cost			cost difference clearing + RD		
		clearing	RD	total	vs. nodal	vs. ref	vs. perfect foresight
11	nodal	12 425	0	12 425	0	-	-
12	unlimited trade	11 781	1699	13 480	1055	620.4	-
perfect foresight:							
13	by capacity (= ref)	12 303	556	12 859	435	0.0	-
14	by NEX	12 178	684	12 862	437	2.4	-
15	by N	12 316	560	12 875	450	15.8	-
process-induced uncertainty:							
16	by capacity	12 301	561	12 862	438	3.1	3.1
17	by NEX	12 181	683	12 865	440	5.2	2.8
18	by N	12 312	565	12 876	452	17.0	1.2

2 million Euros for the *by NEX* method, which is statistically nonsignificant (cf. Appendix Figure 9). The cost increase of the *by N* method compared to the reference stays almost the same in absolute terms (16 million Euros), but much higher total costs than without internal congestions imply lower relative differences between the methods.

In contrast, the effect of process-induced uncertainty remains rather small. Costs increase by between 1.2 million Euros for the *by N* method and 3.1 million Euros for the *by capacity* method. The cost increases resulting from changing from perfect to imperfect foresight (while using the same generation shift key method) are found to be statistically significant for the *by capacity* and *by NEX* method (cf. test procedure described in Appendix 2.4.5). The corresponding increase for the *by N* method is not significant.

3.3 Adjustments to the capacity allocation process

3.3.1 Overview

In 2019, the European Union adopted the *Clean energy for all Europeans package* (EU 2019). It contains a new electricity regulation, that, among others, sets out new rules about the selection of critical network elements and the sizing of remaining available margins (i.e., affecting the set $\mathcal{F}^{\text{CNE}}$ and the values of $R_f^{\text{nsfd/sfd}}$ in Equation 6). Together with the decision of the Agency for the Cooperation of Electricity Regulators (ACER) about the capacity calculation methodology of the CORE Region (which Central Western Europe is a part of), it specifies that remaining available margins must gradually increase to comply with a minimum number of 70 % of the thermal line capacity (ACER 2019a; ACER 2019b). Additionally, the inclusion of internal critical network elements into the capacity calculation will only be allowed in special cases. In this section, we

Table 4: Results with adjustments to the allocation process and imperfect bidding zones (reference grid expansion). Costs in million EUR. RAM = Remaining Available Margin. CNE = Critical Network Element. RD = redispatch.

no.	description	total cost			RD (TWh)
		clearing	RD	vs. ref	
13	by capacity (= ref)	12 303	556	0.0	9.50
adjustments to the capacity allocation process (all with generation shift key = by capacity):					
19	individual contingency margins (overall)	12 318	550	8.2	9.29
20	individual contingency margins (cross-border)	12 442	520	103.1	8.77
21	no internal CNEs	12 057	839	36.2	15.50
22	minimum RAM 70 %	12 064	806	10.3	14.06
23	no internal CNEs + minimum RAM 70 %	12 038	888	67.0	16.97

look at the effects of these measures: *Minimum remaining available margins*, no *internal critical network elements* and the combination of both.

Additionally, we consider the actions that may have contributed to this new legislation: In the past, there have been allegations that grid operators had reduced remaining available margins on cross-border lines to solve problems within their own bidding zones (cf. ACER 2017). This is modeled by introducing *individual contingency margins* of 25 % that are applied to the remaining available margins on transmission lines – either to the top ten most often congested lines *overall* or to the top ten most often congested *cross-border* transmission lines.

3.3.2 Results

Table 4 shows the results for the sensitivity regarding the adjustments to the capacity allocation process. As in Section 3.2, we use the grid model with intra-zonal transmission bottlenecks.

In our calculations, the top 10 overloaded lines overall are all intra-zonal lines. It follows that there is no overlap between the two sets of lines considered for reduction of remaining available margins described by cases 19 and 20 in Table 4. Placing individual contingency margins and thus reducing remaining available margins on the top 10 overall overloaded lines (case 19) and the top 10 overloaded cross-border lines (case 20) increases costs in both cases but by substantially different amounts. In case 19, costs rise by 8 million Euros, while the measures of case 20 increase costs by 103 million Euros. Somewhat reduced redispatch volumes (−0.73 TWh per direction compared to the reference case) and therefore reduced redispatch costs are traded for much higher clearing costs.

Omitting internal critical network elements (case 21), forcing remaining available margins to at least 70 % of the lines' maximum thermal capacities (case 22) and the combination of both (case 23) lead to higher costs as well. However, in contrast to case 19 and case 20, the increase is due to higher costs for redispatch, which cannot be offset by decreasing market clearing costs.

In particular, omitting internal critical network elements (case 21) lowers the zonal clearing costs but increases redispatch costs by a much larger amount, leading to an increase in total costs of 36 million Euros. Even though internal constraints are not adequately represented in a zonal setting, completely omitting them to facilitate cross-zonal trade leads to higher costs overall. For case 22, the observed increase of 10 million Euros compared to the reference case is not as big and statistically not significant. For both cases, lower clearing costs are paid by an increase in redispatch volumes per direction of 4.55 TWh and 6.00 TWh, respectively. The simultaneous application of both measures (case 23), which corresponds to the actually implemented policy, leads to a superadditive increase in costs of 67 million Euros while the combined effect for redispatched quantities is 7.47 TWh and hence subadditive.

Results from the redispatch-sensitivity analysis (see Appendix A.3), where marginal costs for redispatch are assumed to be the same as in the zonal clearing stage, are different for the cases without internal CNEs and with minRAM. The redispatch volume remains comparable because the volume penalty still applies. However, due to the decreased marginal costs compared to penalized redispatch, the redispatch costs are much lower. This results in substantially decreased costs of −42.4 million Euros to −49.9 million Euros compared to the reference case. The overall effects of minRAM and the omission of internal CNEs, including their direction, are therefore heavily dependent on the parametrization of redispatch costs. Because the assumption of equal costs between redispatch and zonal clearing is not realistic, these values may be interpreted as the upper bound for the savings that can be achieved by artificially increasing the capacity available for trade. Overall, no definitive conclusions may be drawn regarding the cost effects of minRAM and the omission of internal CNEs.

4 Discussion and conclusion

The paper at hand presents an in-depth analysis of Flow-Based Market Coupling – one of the cornerstones of the European target model for electricity market design. Subsequently, we summarize the key findings of our analyses and link them to current (political) discussions and publications.

Our analyses have revealed that, given a suitable (or idealized) setting for Flow-Based Market Coupling, the overall performance of this approach comes close to that of the theoretical first-best solution (i.e., the nodal pricing solution). Using the *unlimited trade* setting as a second reference, Flow-Based Market Coupling realizes around 87 % of the welfare gains that are possibly achievable through integrated nodal market clearing. In absolute terms, inefficiencies in terms of welfare of around 100 million Euros (0.9% of nodal generation costs) remain. In a broader sense, this might be considered as the cost of maintaining zonally organized electricity markets with their associated benefits (such as reduced vulnerability to market power, non-existence of redistributive effects due to heterogenous prices within countries, highly liquid markets, and, in the case of stable zones, no transitional costs).

This picture changes substantially once conditions are not ideal. In a realistic setting, where intra-zonal bottlenecks are commonplace, Flow-Based Market Coupling only attains about 59 % of the theoretically possible welfare gains. This implies welfare losses of around 440 million Euros (3.5% of nodal generation costs) compared to nodal pricing. Such a degree of inefficiency may outweigh the benefits of zonal pricing.

Contrasting both situations we conclude: Flow-Based Market Coupling generally has the capability to perform reasonably well. Its problems arise from intra-zonal bottlenecks. As shown by Felten et al. (2019), zonal pricing is ineffective in managing intra-zonal congestion, because the main cause of this congestion – intra-zonal trade – is not controlled by the market clearing process. Solution approaches to this problem include the acceleration of the grid expansion process or the enhancement of price zone configurations. Either one would help to improve the efficiency of Flow-Based Market Coupling.

The results of the sensitivity calculation, where we consider different generation shift key methods, show an overall cost difference between the *by capacity* and the *by NEX* methods of more than 120 million Euros in the idealized setting. Moreover, the welfare changes due to the use of different methods are all statistically significant. However, when considering imperfect price zones with intra-zonal bottlenecks, the impacts of different methods for calculating generation shift keys become very small. Even the quite naive *by NEX* method achieves results close to the reference method. In several cases, the cost differences are not even statistically significant. While this result may surprise at first sight, it is in line with results presented in Dierstein (2017) and Finck et al. (2018). Again, the explanation lies in the existence of intra-zonal bottlenecks. The ineffectiveness in managing intra-zonal congestion is more dominant than the change in generation shift keys.[20]

The main takeaway from these analyses is that the improvement of generation shift key procedures is likely to become relevant in the future or for individual price zones with few intra-zonal bottlenecks. Once grid expansion proceeds or price zones are reconfigured, the choice of method will have relevant leverage on overall welfare. For the current situation, however, these changes hardly make a difference.

Besides different methods for determining generation shift keys, we have also analyzed the role of forecast errors. We show that these errors do not have a major effect on welfare. FBMC therefore seems to be quite robust against forecast errors.

Finally, we have analyzed the impact of regulatory measures and adjustments by transmission system operators to the capacity allocation process. Under the national *copperplate* setting, redispatch amounts and costs are fairly low. Redispatch quantities in the *copperplate* setting are below 1 TWh; redispatch costs are below 50 million Euros. Such a situation would probably not trigger

[20]Similar to our realistic setting, Dierstein (2017) and Finck et al. (2018) investigate price zones with a relevant number of intra-zonal bottlenecks. And Dierstein (2017) shows that the difference of the *by capacity* and *by N* method is comparably small. The six different generation shift keys of Finck et al. (2018) do not show major effects on overall generation shifts either. Mainly with regard to generation shifts in Czech Republic (where intra-zonal bottlenecks may not be as dominant), Finck et al. (2018) observe bigger generation shifts.

regulatory measures or compel transmission system operators to introduce *individual contingency margins*. However, redispatch quantities rise to around 10 TWh in a realistic setting. Curing this by reducing the remaining available margins then turns out to have an appreciable effect: Redispatch amounts are reduced by 8 % when contingency margins are applied just to the top 10 most overloaded cross-border lines. But this redispatch reduction comes at the expense of overall welfare, with more than 100 million Euros additional system costs.

In contrast, regulatory measures that increase remaining available margins or exclude intra-zonal line constraints from the market clearing problem have a major impact – especially on redispatch amounts. The heavy increases by 50 % or more would certainly pose an operational challenge to Europe's transmission system operators and increase non-market-based payments for redispatch by about 300 million Euros. Trade is fostered by increasing or neglecting remaining available margins with a decrease in market clearing costs in line with results from similar scenarios reported in Wyrwoll et al. (2018) and Matthes et al. (2019). The effect on overall system costs remains unclear, as the direction of the effect varies between different parametrizations of redispatch costs. Because the operational challenges presented by the large increase in redispatch volume remain consistent between parametrizations and the effect on welfare is not clearly positive, the adoption of minRAM and the removal of internal CNEs cannot be recommended.

In conclusion, our investigations have systematically analyzed the performance and levers of Flow-Based Market Coupling. Most importantly, (i) the relevance of intra-zonal bottlenecks on the performance of Flow-Based Market Coupling calls for reconfiguration of price zones or accelerated grid expansion, (ii) the regulatory changes introduced in EU (2019) and ACER (2019a) carry the risk – or rather high probability – of severe increases in redispatch amounts and costs, and (iii) other debated features of flow-based market coupling (e.g., generation shift keys, forecast errors) may not be as important as commonly believed – at least as long as intra-zonal transmission line capacities are scarce.

Putting the results into the broader debate on unbundled vs. integrated power market architectures, our results clearly point to the limits of unbundled architectures. In the middle of a shifting generation landscape, the current unbundled European approach will only remain manageable and viable, if the discrepancy to the outcomes of the integrated approach does not increase further. This calls for either rapid extensions of the intra-zonal network, especially in Germany, or an adjustment of the zonal boundaries.

References

ACER – Agency for the Cooperation of Energy Regulators (2017). *ACER / CEER – Annual Report on the Results of Monitoring the Internal Electricity and Natural Gas Markets in 2016*. URL: https://acer.europa.eu/Official_documents/Acts_of_the_Agency/Publication/ACER%20Market%20Monitoring%20Report%202016%20-%20ELECTRICITY.pdf (visited on 07/30/2019).

ACER – Agency for the Cooperation of Energy Regulators (2018). *ACER / CEER – Annual Report on the Results of Monitoring the Internal Electricity and Natural Gas Markets in 2017*. URL: https://www.acer.europa.eu/en/Electricity/Market%20monitoring/Pages/Current-edition.aspx (visited on 08/30/2019).

ACER – Agency for the Cooperation of Energy Regulators (2019a). *ACER Decision 02-2019 on the Core CCR TSOs' proposals for the regional design of the day-ahead and intraday common capacity calculation methodologies*. URL: https://acer.europa.eu/Official_documents/Acts_of_the_Agency/Individual%20decisions/ACER%20Decision%2002-2019%20on%20CORE%20CCM.pdf (visited on 09/29/2019).

ACER – Agency for the Cooperation of Energy Regulators (2019b). *Day-ahead capacity calculation methodology of the Core capacity calculation region*. Annex 1 to ACER Decision 02-2019 on Core CCM. URL: https://acer.europa.eu/Official_documents/Acts_of_the_Agency/ANNEXESTODECISIONOFTHEAGENCYNo022019/Annex%20I%20-%20ACER%20Decision%20on%20Core%20CCM.pdf (visited on 09/29/2019).

Amprion (2017). *Flow Based Market Coupling - Development of the Market and Grid Situation 2015 - 2017*. URL: https://www.amprion.net/Dokumente/Dialog/Downloads/Studien/CWE/CWE-Studie_englisch.pdf (visited on 07/30/2019).

Amprion, APX, Belpex, Creos, Elia, Epex Spot, RTE, TenneT, and TransnetBW (2014). *Documentation of the CWE FB MC solution as basis for the formal approval-request*. URL: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/NetzzugangUndMesswesen/Marktkopllung/140530%20%20CWE%20FB%20MC%20Approval%20document.pdf (visited on 07/24/2019).

Bjørndal, M. and K. Jørnsten (2001). "Zonal Pricing in a Deregulated Electricity Market". In: *The Energy Journal* 22.1, pp. 51–73. DOI: 10.5547/ISSN0195-6574-EJ-Vol22-No1-3.

BNetzA – Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen (2019). *Quartalsbericht zu Netz- und Systemsicherheitsmaßnahmen*. Gesamtjahr und Viertes Quartal 2018. Bonn. URL: https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/Versorgungssicherheit/Netz_Systemsicherheit/Netz_Systemsicherheit_node.html (visited on 08/08/2019).

CREG – Commission de Régulation de l'Électricité et du Gaz (2017). *Study - functioning and design of the Central Western European day-ahead flow-based market coupling for electricity: Impact of TSOs discretionary actions*. URL: https:

//www.creg.be/sites/default/files/assets/Publications/Studies/F1687EN.pdf (visited on 07/30/2019).

Consentec (2015). *Economic efficiency analysis of introducing Economics efficiency analysis of introducing smaller bidding zones*. URL: https://www.eex.com/blob/7412/97dfe4307af0ded860ba2c0e3ffb1e99/20150213-consentec-eex-bidding-zones-data.pdf (visited on 10/13/2016).

Dierstein, C. (2017). "Impact of Generation Shift Key determination on flow based market coupling". In: *14th International Conference on the European Energy Market (EEM), Dresden*. DOI: 10.1109/EEM.2017.7981901.

Ehrenmann, A. and Y. Smeers (2005). "Inefficiencies in European congestion management proposals". In: *Utilities policy* 13.2, pp. 135–152. DOI: 10.1016/j.jup.2004.12.007.

Entso-E – European Network of Transmission System Operators (2016). *All TSOs' proposal for a common grid model methodology in accordance with Article 17 of Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management*. URL: https://www.netztransparenz.de/portals/1/Content/EU-Network-Codes/CACM/CGMM/CGMM-EN-final.pdf (visited on 01/20/2020).

Entso-E – European Network of Transmission System Operators (2017). *Generation and load shift key implementation guide*. URL: https://docstore.entsoe.eu/Documents/EDI/Library/cim_based/07_Generation_and_Load%20Shift_Key_Implementation_Guide_v2r1.pdf (visited on 07/30/2019).

Entso-E – European Network of Transmission System Operators (2018). *Entso-E Transparency Platform*. URL: https://transparency.entsoe.eu/ (visited on 10/09/2018).

EU – European Parliament and European Council (2019). *Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.158.01.0054.01.ENG (visited on 08/08/2019).

Felten, B., T. Felling, P. Osinski, and C. Weber (2019). *Flow-Based Market Coupling Revised - Part I: Analyses of Small- and Large-Scale Systems*. HEMF Working Paper No. 06/2019. URL: https://ssrn.com/abstract=3404044 (visited on 07/04/2019).

Finck, R., A. Ardone, and W. Fichtner (2018). "Impact of Flow-Based Market Coupling on Generator Dispatch in CEE Region". In: *2018 15th International Conference on the European Energy Market (EEM)*. IEEE, Piscataway, New Jersey, pp. 1–5. DOI: 10.1109/EEM.2018.8469927.

Green, R. (1997). "Electricity transmission pricing: an international comparison". In: *Utilities Policy* 6.3. Transmission Pricing, pp. 177–184. DOI: 10.1016/S0957-1787(97)00022-2.

Greene, W. H. (2012). *Econometric analysis*. eng. 7. ed., internat. ed. Boston: Pearson.

Hogan, W. W. (1992). "Contract networks for electric power transmission". In: *Journal of Regulatory Economics* 4.3, pp. 211–242. DOI: 10.1007/BF00133621.

Hogan, W. W. (1999). "Transmission Congestion: The Nodal-Zonal Debate Revisited". In: *Harvard University, John F. Kennedy School of Government, Center for Business and Government* 29.4.

Keay, M. (2013). *The EU "target model" for electricity markets: fit for purpose?* Oxford Energy Comment. URL: `https://www.oxfordenergy.org/?s=target+model&post_type=publications` (visited on 12/02/2018).

Matthes, B., C. Spieker, D. Klein, and C. Rehtanz (2019). "Impact of a Minimum Remaining Available Margin Adjustment in Flow-Based Market Coupling". In: *2019 IEEE Milan PowerTech*, pp. 1–6. DOI: `10.1109/PTC.2019.8810504`.

Nemo-C – Nemo Committee (2019). *Euphemia Public Description*. Single Price Coupling Algorithm. URL: `https://www.epexspot.com/en/market-coupling/pcr` (visited on 08/31/2019).

Schönheit, D. and R. Sikora (06/2018). "A Statistical Approach to Generation Shift Keys". In: *2018 15th International Conference on the European Energy Market (EEM)*, pp. 1–6. DOI: `10.1109/EEM.2018.8469900`.

Schönheit, D., R. Weinhold, and C. Dierstein (2020). "The impact of different strategies for generation shift keys (GSKs) on the flow-based market coupling domain: A model-based analysis of Central Western Europe". In: *Applied Energy* 258, p. 114067. DOI: `https://doi.org/10.1016/j.apenergy.2019.114067`.

Schweppe, F. C., M. C. Caramanis, R. D. Tabors, and R. E. Bohn (1988). *Spot pricing of electricity*. eng. Kluwer Acad. Publ.

Sebestyén, M., D. Divenyi, and P. Sörés (2018). "An Enhanced Calculation Method of Generation Shift Keys in Flow Based Market Coupling". In: *2018 15th International Conference on the European Energy Market (EEM)*, pp. 1–5. DOI: `10.1109/EEM.2018.8469923`.

Van den Bergh, K. and E. Delarue (2016). "An improved method to calculate injection shift keys". In: *Electric Power Systems Research* 134, pp. 197–204. DOI: `10.1016/j.epsr.2016.01.020`.

Wilson, R. (2002). "Architecture of Power Markets". In: *Econometrica* 70.4, pp. 1299–1340. DOI: `10.1111/1468-0262.00334`.

Wyrwoll, L., A. Blank, C. Müller, and R. Puffer (2019). "Determination of Preloading of Transmission Lines for Flow-Based Market Coupling". In: *2019 16th International Conference on the European Energy Market(EEM)*, pp. 1–6.

Wyrwoll, L., K. Kollenda, C. Müller, and A. Schnettler (2018). "Impact of Flow-Based Market Coupling Parameters on European Electricity Markets". In: *2018 53rd International Universities Power Engineering Conference (UPEC)*, pp. 1–6. DOI: `10.1109/UPEC.2018.8541904`.

Appendix

A Further detailed results

A.1 Redispatch indicators

Table 5: Redispatch indicators for all cases.

no	description	annual redispatch				peak hour	
		∅		total		total	
		viol. constr.	measures	all	cross-border	all	measures
		#	#	TWh	TWh	MWh	#
idealized price zones:							
2	unlimited trade	5.1	23.1	25.9	25.9	19 357	61
perfect foresight:							
3	by capacity (= ref)	1.0	2.5	0.6	0.6	837	5
4	by NEX	1.9	6.2	3.3	3.3	2779	13
5	by N	1.0	2.4	0.7	0.7	892	8
process-induced uncertainty:							
6	by capacity	1.0	2.6	0.7	0.7	886	8
7	by NEX	1.9	6.2	3.3	3.3	2805	13
8	by N	1.0	2.6	0.8	0.8	1091	11
imperfect price zones:							
perfect foresight:							
12	unlimited trade	38.6	53.9	41.1	37.0	20 116	72
13	by capacity (= ref)	11.6	20.2	9.5	3.3	11 564	91
14	by NEX	11.8	22.2	11.2	5.5	10 526	135
15	by N	11.5	20.4	10.0	3.4	10 798	92
process-induced uncertainty:							
16	by capacity	11.7	20.4	9.6	3.4	11 715	93
17	by NEX	11.8	22.2	11.2	5.5	10 568	104
18	by N	11.5	20.6	10.0	3.5	10 843	91
adjustments to the capacity allocation process (all with GSK = by capacity):							
19	individual contingency margins	11.3	19.8	9.3	3.2	11 564	91
20	indiv. cont. margins (cross-border)	9.7	18.4	8.8	2.9	11 564	91
21	no internal CBs	16.6	27.6	15.5	8.7	12 263	160
22	minimum RAM 70 %	15.1	25.2	14.1	7.6	11 661	95
23	no internal CBs + min. RAM 70 %	18.2	30.2	17.0	10.3	11 699	97

A.2 Significance tests

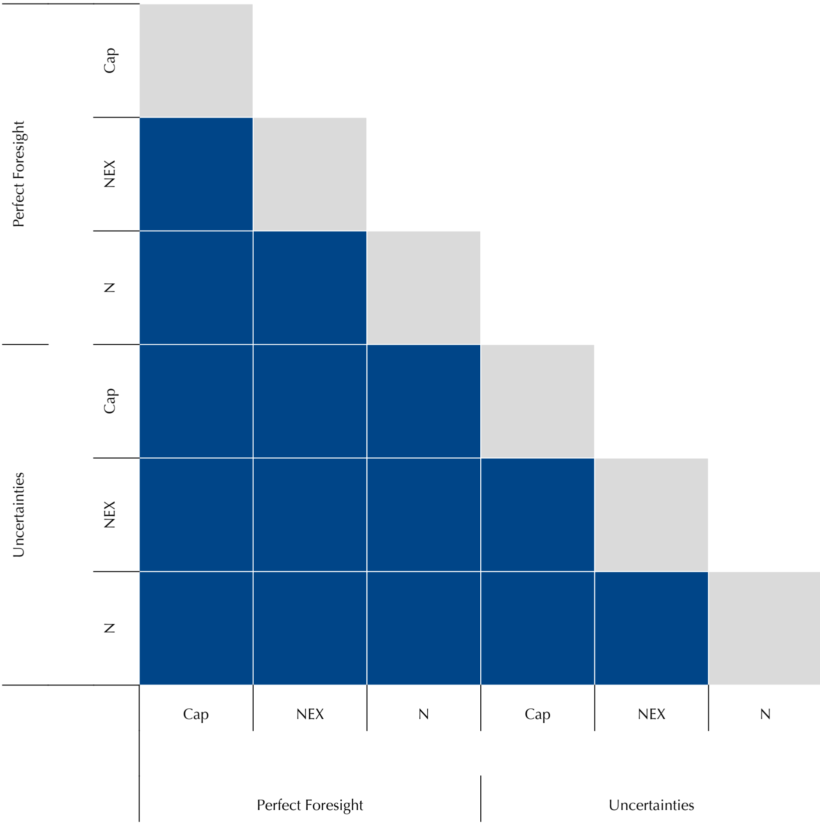


Figure 8: Significant differences between total cost (zonal + redispatch) time series with copper plate assumption using Newey-West adjusted standard errors. Significant differences (at $p = 0.05$) are highlighted in blue.

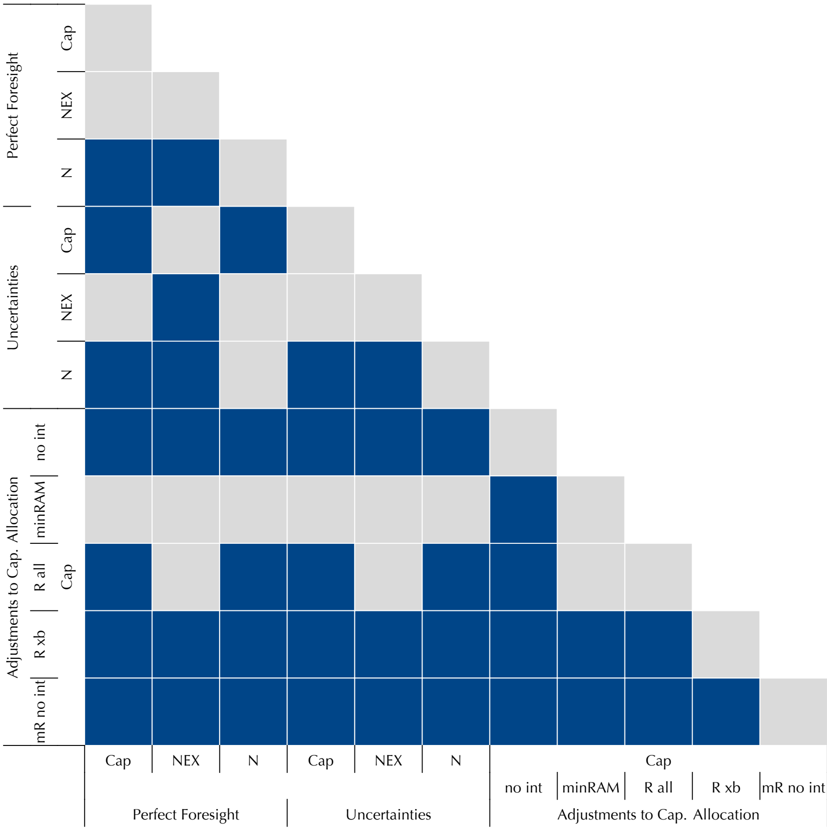


Figure 9: Significant differences between total cost (zonal + redispatch) time series with realistic grid (internal congestions) using Newey-West adjusted standard errors. Significant differences (at $p = 0.05$) are highlighted in blue. Groupings: GSK method, type of adjustment to allocation process. No int: No internal critical network elements. R all: Individual contingency margins. R xb: Individual contingency margins (only cross-border). minRAM: minimum RAM of 70 %. mR no int: minRAM + no int.

A.3 Sensitivity analysis without redispatch penalty factors

The following tables show the results for the redispatch sensitivity analysis, where the proportional factors (see Section 2.3.4) are set to $\beta^+ = \beta^- = 1$. Consequently, marginal costs for redispatch are equal to the costs used in the zonal clearing stage. The volume penalty of γ remains unchanged at 300 EUR / MW, as it reflects the aim to only adjust the market results where it is required for operational security.

This parametrization is unrealistic: It assumes that no additional costs (including opportunity costs) arise from the out-of-market order by the TSOs. However, this parametrization reveals the lower bound for the extra costs incurred by redispatch and, consequently, shows the upper bound for the benefits of measures such as minRAM, which significantly increase redispatch volumes (see Section 3.3).

Table 6 shows the results for the idealized price zone configuration (see Section 3.1). FBMC achieves 79.5 % (compared to 87 % with the original parametrization) of the savings that nodal pricing achieves compared to the unlimited trade scenario. Regarding the procedural sensitivities of different GSK methods and process-induced uncertainty, the difference between the *by capacity* method (3) and the *by NEX* method (4) decreases from 120.4 million Euros to 90.4 million Euros, while the difference to the *by N* method stays the same. The rather large change in the *by NEX* method can be explained by the higher redispatch volumes compared to the other GSK methods, with the associated costs decreased because of the decreased marginal costs for redispatch. Adding process-induced uncertainty shows a similar impact as in the original parametrization and the cost changes are similar to the perfect foresight cases. Overall, the results are largely unaffected by the changed redispatch costs.

Table 7 shows the results for the imperfect price zone configuration. In this case, FBMC achieves only 29 % (compared to 59 % before) of the savings that can be achieved with nodal pricing compared to the unlimited trade scenario. This is an effect of drastically reduced redispatch costs in the unlimited trade scenario, which decrease the cost difference to nodal pricing from 1055 million Euros to 511 million Euros. Conversely, the cost difference from FBMC to nodal pricing decreases only from 435 million Euros to 361 million Euros. The lower marginal costs for redispatch also disproportionately affect the redispatch costs of the *by NEX* method (14), whose costs are now lower than those of the *by capacity* method (13). However, the cost differences were very small in the original case, and are still not comparable to the difference with perfect price zones. Consequently, the overall conclusion does not change: Different GSK methods show little effect because the ineffectiveness of managing intra-zonal congestion is more dominant.

Table 8 shows the results for the sensitivity with additional adjustments to the capacity calculation process. For individual contingency margins (19 and 20), there is only little change in the cost difference to the reference case. However, the exclusion of internal CNEs (21), minRAM (22) and the combination of both (23) shows substantially different results compared to the original case. These measures are affected substantially by decreased marginal costs for redispatch because they significantly increase redispatch volumes (see Section 3.3). Without the proportional factors, the cost differences to the reference case are negative, i.e. the measures become financially beneficial. While, as described above, the assumption of equal marginal costs for the zonal clearing and redispatch is unrealistic, this result nevertheless disallows drawing clear conclusions in any direction from a cost perspective. However, the redispatch volumes are similar to the original case. Therefore, concerns regarding the effect on system stability remain.

Table 9 shows results for the redispatch volume indicators. Overall, the volumes are similar to the main results. This shows that, as intended, the volume penalty γ leads to a minimization of the redispatch volumes, which is independent of the parametrization of the marginal costs for redispatch.

Table 6: Results with idealized price zones. Costs in million EUR. Without penalty factors, with quantity penalty.

no.	description	total cost			cost difference clearing + RD		
		clearing	RD	total	vs. nodal	vs. ref	vs. perfect foresight
1	nodal	11 989	0	11 989	0	-	-
2	unlimited trade	11 781	685	12 466	478	379.3	-
perfect foresight:							
3	by capacity (= ref)	12 055	32	12 087	98	0.0	-
4	by NEX	12 018	159	12 177	189	90.4	-
5	by N	12 065	38	12 103	115	16.3	-
process-induced uncertainty:							
6	by capacity	12 057	33	12 090	101	3.0	3.0
7	by NEX	12 024	155	12 180	191	92.9	2.5
8	by N	12 065	41	12 106	117	19.0	2.7

Table 7: Results with imperfect price zones (reference grid expansion). Costs in million EUR. RD = redispatch. Without penalty factors, with quantity penalty.

no.	description	total cost			cost difference clearing + RD		
		clearing	RD	total	vs. nodal	vs. ref	vs. perfect foresight
11	nodal	12 425	0	12 425	0	-	-
12	unlimited trade	11 781	1155	12 936	511	150.3	-
perfect foresight:							
13	*by capacity* (= ref)	12 303	483	12 786	361	0.0	-
14	by NEX	12 178	592	12 770	345	−15.7	-
15	by N	12 316	476	12 792	367	5.7	-
process-induced uncertainty:							
16	by capacity	12 301	486	12 787	363	1.5	1.5
17	by NEX	12 181	592	12 773	348	−13.3	2.4
18	by N	12 312	480	12 792	367	6.0	0.3

Table 8: Results with adjustments to the allocation process and imperfect bidding zones (reference grid expansion). Costs in million EUR. RAM = Remaining Available Margin. CNE = Critical Network Element. RD = redispatch. Without penalty factors, with quantity penalty.

no.	description	total cost			RD (TWh)
		clearing	RD	vs. ref	
13	*by capacity* (= ref)	12 303	483	0.0	9.50
adjustments to the capacity allocation process (all with generation shift key = by capacity):					
19	individual contingency margins (overall)	12 318	480	11.9	9.28
20	individual contingency margins (cross-border)	12 442	459	115.2	8.76
21	no internal CNEs	12 057	680	−49.0	15.50
22	minimum RAM 70 %	12 064	672	−49.9	14.06
23	no internal CNEs + minimum RAM 70 %	12 038	705	−42.4	16.97

Table 9: Redispatch indicators for all cases. Without penalty factors, with quantity penalty.

no	description	annual redispatch				peak hour	
		∅		total		total	
		viol. constr.	measures	all	cross-border	all	measures
		#	#	TWh	TWh	MWh	#
idealized price zones:							
2	unlimited trade	5.1	24.0	25.9	25.9	19 366	65
perfect foresight:							
3	by capacity (= ref)	1.0	2.5	0.6	0.6	837	5
4	by NEX	1.9	6.2	3.3	3.3	2864	13
5	by N	1.0	2.4	0.7	0.7	892	8
process-induced uncertainty:							
6	by capacity	1.0	2.5	0.7	0.7	886	8
7	by NEX	1.9	6.3	3.3	3.3	3015	14
8	by N	1.0	2.5	0.8	0.8	1091	11
imperfect price zones:							
perfect foresight:							
12	unlimited trade	38.6	55.2	41.1	37.0	20 094	80
13	by capacity (= ref)	11.6	20.4	9.5	3.4	11 557	94
14	by NEX	11.8	22.6	11.2	5.6	10 531	139
15	by N	11.5	20.7	10.0	3.5	10 791	93
process-induced uncertainty:							
16	by capacity	11.7	20.6	9.6	3.5	11 706	103
17	by NEX	11.8	22.6	11.2	5.6	10 582	105
18	by N	11.5	20.8	10.0	3.6	10 836	94
adjustments to the capacity allocation process (all with GSK = by capacity):							
19	individual contingency margins	11.3	20.0	9.3	3.3	11 557	94
20	indiv. cont. margins (cross-border)	9.7	18.6	8.8	2.9	11 557	94
21	no internal CBs	16.6	28.0	15.5	8.7	12 238	165
22	minimum RAM 70 %	15.1	25.5	14.1	7.6	11 642	110
23	no internal CBs + min. RAM 70 %	18.2	30.6	17.0	10.3	11 697	102

Chapter 4

Improving flow-based market coupling by integrating redispatch potential – Evidence from a large-scale model

by Michael Bucksteeg, Simon Voswinkel, and Gerald Blumberg
submitted to Energy Policy in May of 2023

Improving flow-based market coupling by integrating redispatch potential—Evidence from a large-scale model by Michael Bucksteeg, Simon Voswinkel and Gerald Blumberg

Abstract

Power markets have been gradually integrated to achieve the target of a single European market. A major step was the introduction of the flow-based market coupling (FBMC) in Central Western and Eastern Europe (Core region). FBMC reflects the physical constraints of the underlying transmission grid in detail. However, the European Commission and regulators imposed minimum margins to increase cross-border trade and to foster price convergence between the different bidding zones, neglecting physical constraints and increasing redispatch volumes. Integrating redispatch potentials into FBMC allows for moving closer to physical reality while maintaining a high level of cross-border trade. In this study, we develop a multi-stage model covering capacity calculation, market coupling, and redispatch stages. This study is the first to evaluate different options for integrating FBMC and redispatch potentials based on a large-scale numerical analysis of Central Europe. The results reveal that minimum margins effectively increase cross-border trade. However, this comes at a high cost due to additional redispatch needs, which reduce overall welfare. Integrating redispatch potentials in the market-clearing stage leads to a more efficient increase in cross-border capacities and elevates welfare. In the case of combining both approaches, the analysis indicates improved welfare of roughly 80 M€ per year.

Keywords: Flow-based market coupling, European electricity market, cross-border trade, congestion management, redispatch

JEL-Classification: Q4

MICHAEL BUCKSTEEG

(CORRESPONDING AUTHOR)

University of Hagen, Germany

Universitätsstraße 41, 58097 Hagen

Michael.Bucksteeg@fernuni-hagen.de

SIMON VOSWINKEL

House of Energy Markets and Finance

University of Duisburg-Essen, Germany

Simon.Voswinkel@uni-due.de

GERALD BLUMBERG

House of Energy Markets and Finance

University of Duisburg-Essen, Germany

Gerald.Blumberg@uni-due.de

Content

1 Introduction

In European electricity markets, the allocation of transmission capacity is based on a zonal pricing approach with redispatch (Bjørndal and Jornsten 2001). The zonal markets have been gradually integrated to achieve the target of a single European market. A major step toward the target model was the introduction of the so-called flow-based market coupling (FBMC) (ACER 2019; European Parliament 2019). Compared to the former bilateral net transfer capacity (NTC) method, the FBMC approach entails an improved representation of physical limitations in the transmission grid. It allows the allocation of transmission capacity to the most efficient trades between market zones. Since its launch in Central Western Europe in 2015, FBMC has increased cross-zonal trading capacities in the region (Schönheit et al. 2021b). Consequently, the integration of zonal markets and welfare could be enhanced even though welfare gains remained below expectations (Kristiansen 2020; Lang et al. 2020).

Although some studies have addressed the benefits of nodal pricing in Europe (Neuhoff et al. 2013; Bjørndal et al. 2018; Bjørndal et al. 2014), its introduction has failed mainly due to political reservations (Antonopoulos et al. 2020). Consequently, the debate focuses on further developing the zonal market design. In zonal markets, the delimitation of market zones is essential for efficiency and effectiveness in managing grid congestion. A large body of research deals with adequate bidding zone configurations and related impacts on electricity markets and congestion management (Trepper et al. 2015; Egerer et al. 2016; Felling and Weber 2018; Deilen et al. 2019; Felling 2019; Felling et al. 2023). Further contributions extend this by studying the long-term effects and risks of changing zonal configurations (Grimm et al. 2016; Bertsch et al. 2017; Deilen et al. 2019). Zonal markets with imperfect bidding zone configurations drive the need for redispatch measures to maintain stable grid operation. Moreover, the extension of fluctuating renewable energy sources, such as wind and solar energy, has led to a considerable increase in redispatch volumes and costs (ACER and CEER 2021). This development has stimulated further debate about optimizing redispatch procedures (Kunz and Zerrahn 2015; Zerrahn and Kunz 2016) and market-based redispatch (Hirth and Schlecht 2020; Grimm et al. 2018; Bjørndal et al. 2017; Grimm et al. 2022; Martin et al. 2022).

Despite the advancement of zonal market coupling through the FBMC approach, regulators had concerns about the level of commercial cross-border exchanges, which remained below expectations (CREG 2017; ACER and CEER 2021). While combining a zonal approach with an improved representation of physical constraints is (on average) beneficial in terms of welfare and system security, it implies that not only cross-zonal but also internal transmission lines may limit cross-border trade. In this context, the FBMC method includes several specifications and parameters, such as the selection of critical network elements (CNEs), generation shift keys (GSKs), reliability margins, and adjustment values, leading to imperfections and affecting the capacities allocated to the market (Marien et al. 2013; van den Bergh et al. 2016; Schönheit et al. 2020; Felten et al. 2021).

Consequently, minimum margins (or trading capacities) were introduced to increase cross-border exchanges (Henneaux et al. 2021; Schönheit et al. 2021a). Accordingly, trading capacities were raised to a minimum, possibly violating underlying physical transmission constraints. At the same time, transmission system operators (TSOs) have contemplated integrating costly remedial actions (i.e., redispatch) into the market-clearing algorithm to increase the capacity domain given to the market (Elia Group 2019). The basic idea is that if a network element limited cross-border exchange, redispatch measures or bids would be considered to increase the available margin on this line during the market clearing. Combining FBMC and redispatch, recent studies have analyzed different options to increase cross-border exchanges (Elia Group 2019; Poplavskaya et al. 2020; Schlecht and Hirth 2021). Hirth and Schlecht (2020), Schlecht and Hirth (2021), and Ehrhart et al. (2022) noted that these market-based redispatch mechanisms may be subject to strategic bidding behavior—that is, inc-dec gaming—which remains the main argument against this market design option. However, Schlecht and Hirth (2021) also discussed an approach in which the TSO is responsible for determining the available redispatch potential and providing the corresponding parameters to the market-clearing algorithm (i.e., the mandatory redispatch potential based on costs).[1]

While this TSO-based approach avoids the inc-dec gaming issue and could be implemented at short notice, an in-depth analysis of related design options and their effects on electricity markets and grid operation is still pending. This study fills this gap and contributes to the existing literature in four respects. First, an extended FBMC problem that includes the redispatch potential is formulated. Second, three potential design options for determining the available redispatch units (incorporated in the market-clearing problem) are proposed. Third, the effects of the minimum remaining available margins (minRAMs) and the integration of redispatch potentials on welfare and system security are studied using a large-scale model covering Central Western and Eastern Europe (i.e., Core capacity calculation region). Fourth, we discuss the implications of introducing the analyzed design options, which can guide policymakers and regulators.

The remainder of this paper is organized as follows. In section 2, the general approach of this paper is presented. This includes an overview of the procedure for modeling FBMC in the current capacity calculation and allocation and operational redispatch planning processes implemented in the Core region. Moreover, the general effects of integrating the redispatch potential are discussed. In section 3, the

[1] However, market-based redispatch mechanisms may still be necessary in the future to integrate load-based flexibilities (including storages) if costs cannot be derived appropriately. Yet, these flexibilities may contribute to an efficient redispatch regime and account for a potential lack of positive redispatch potential in the future. Different design options are currently under discussion (Blumberg et al. 2022; Heilmann et al. 2022).

methods are introduced. First, the modeling of FBMC, including capacity calculation and redispatch optimization, is described. Second, the procedures and design options for integrating redispatch are proposed. The results of the proposed methodology are discussed in section 4. Finally, we draw the main conclusions and discuss the implications for policymakers.

2 Flow-based market coupling and redispatch potential

In zonal electricity markets, sufficient cross-zonal trading capacity is essential for creating welfare and synergies regarding system security. FBMC involves several actors and processes that will be briefly introduced in the following. Based on this introduction, the integration of redispatch potentials into the market-clearing algorithm and the underlying intentions are described. The setup of this study is based on the sequence of the day-ahead capacity calculation process for the Core region, the subsequent market coupling process, and the operational procedures for securing the grid (ACER 2019).

TSOs are responsible for the capacity calculation starting two days ahead of delivery. As shown in Figure 1, the capacity calculation is based on a grid model and forecasts regarding generation and load. The outputs of the capacity calculation are the trading capacities. In the case of FBMC, these commercial transaction constraints are represented by the remaining available margins (RAMs) of selected critical network elements (CNEs). The set of CNEs should contain the grid elements most affected by cross-zonal trade, helping to monitor the flows and sending more accurate congestion signals (Kristiansen 2020). Moreover, generation shift keys and zonal sensitivities (i.e. power transfer distribution factors, PTDFs) are determined that translate changes in the net position (i.e., net import or export) of a market zone into a shift in the power flow on a CNE. In other words, the physical constraints of the transmission grid limiting power flows between market zones are translated into commercial transaction constraints given to the market operators responsible for the subsequent market coupling (ACER 2019).

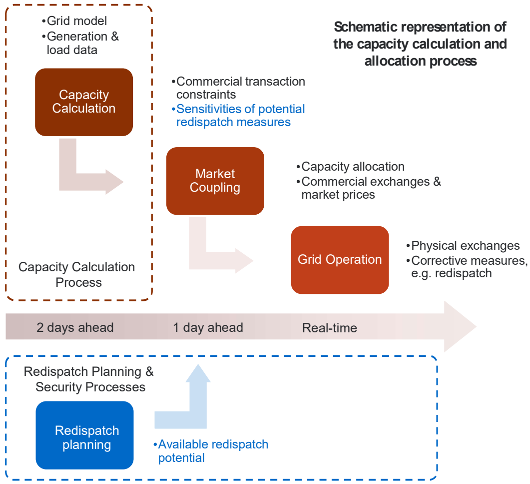


Figure 1: Schematic representation of the flow-based capacity calculation and allocation process

The market-clearing algorithm is executed by the market operators using the constraints submitted by the TSOs and based on the market participants' bids one day before delivery. With the objective of maximizing welfare, the scarce transmission capacity (of the considered CNEs) is allocated to the most efficient trades between all relevant market zones. The market-clearing algorithm delivers commercial exchanges (i.e., net positions) and market-clearing prices for each market zone. For the subsequent grid operation, TSOs must translate the commercial exchanges back into physical ones. Due to the zonal approximation of the FBMC, the market results may violate physical constraints requiring corrective measures, such as redispatch, to maintain stable grid operation. This involves (cross-border) redispatch planning and security processes to guarantee the real-time availability of redispatch units, which are carried out in parallel with the capacity calculation and allocation process (see also Figure 1) (ACER 2019).

Figure 2 depicts a schematic representation of a flow-based domain for a two-zone example and reveals the main motivation for integrating the redispatch potential into the market-clearing algorithm. In the initial situation, the import to market zone B is limited by a CNE (see Net Import B1 and the dashed red line in situation 1). A limitation of CNEs results from technical parameters and the system state and is expressed via the RAM (and zonal PTDFs). For instance, a system state associated with a high wind infeed within a market zone may induce a power flow on the CNE, reducing its RAM and cross-zonal exchange capabilities remaining for the day-ahead market coupling. Accordingly, relaxing the power flow on the CNE increases its RAM and cross-zonal exchange capabilities. In Figure 2, this is indicated by the shift of the limiting CNE, allowing a higher import to zone B (see Net Import B2 and the solid red line in situation 2).

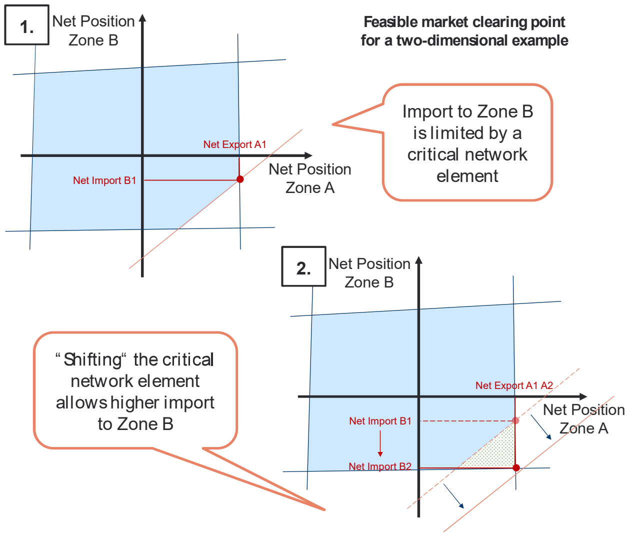


Figure 2: Schematic representation of a flow-based domain for a two-zone example

Suppose that the CNE is an internal grid element almost fully utilized during the capacity calculation. In this case, the internal CNE would limit cross-zonal trade, which should not occur according to the targets of Regulation (EU) 2019/943 (European Parliament 2019). In the long- and mid-term, grid expansion or a reconfiguration of bidding zones would be suitable measures to address this issue (Bertsch et al. 2017; Felling and Weber 2018; Felling et al. 2023). In the short term, a TSO may use remedial actions to reduce the utilization of the CNE during the capacity calculation. However, two problems arise related to costly remedial actions (i.e., redispatch). First, considering redispatch during the capacity calculation involves the instruction and activation of the affected redispatch units representing a market intervention by the TSO. Second, the TSO needs to know the market-clearing point, which is subject to uncertainty (e.g., forecast errors) one day before the market clearing, giving rise to inefficiencies.

Integrating the redispatch potential into the market-clearing algorithm helps overcome both issues. The EUPHEMIA algorithm was developed to solve the day-ahead market coupling problem (N-SIDE 2023). It delivers the market-clearing prices and net positions at the bidding zonal level. To include the redispatch potential, TSOs determine the available redispatch units (called the "redispatch potential") during the capacity calculation process and provide the market operators with the required information. These include the location, available capacity, and activation costs of redispatch units and their sensitivity regarding the power flow on CNEs. If a CNE limits the cross-zonal exchange, the adjusted market-clearing algorithm may utilize the redispatch potential to make available additional capacity for this purpose. Due to the simultaneous optimization of market bids, cross-zonal trading capacities, and the redispatch potential, redispatch will only be used when efficient. This applies if the additional costs for utilizing the redispatch potential are smaller than (or equal to) the welfare gain due to additional cross-zonal exchange.

Integrating redispatch into the market-clearing algorithm requires amendments to the existing processes, as marked in blue in Figure 1. However, the focus of this contribution is to extend the market-coupling problem and to determine the redispatch potential, as described in the following section.

3 Methods

3.1 Modeling flow-based market coupling

Comprehensive modeling of the European electricity market is essential to understanding the effects of integrating the redispatch potential into the market mechanism. The central pillar of European integrated electricity markets is FBMC, as outlined in section 1. The following sections describe the calculation of the flow-based parameters and the zonal market outcome (section 3.1.1), the extension of the market coupling problem by integrating the redispatch potential (section 3.1.2), and the redispatch problem that guarantees secure grid operation (section 3.1.3).

3.1.1 Basic model

We extend the model developed by Voswinkel et al. (2019) and include the determination of the flow-based parameters, zonal market clearing, and the optimization of redispatch. The upper part of *Figure 3* details the steps taken in the model.

The calculations are based on a grid model that covers most of the Core capacity calculation region (Core CCR)[2] and Switzerland. The model comprises around 2500 nodes and over 4000 extra-high-voltage (transmission) grid lines. Generation and demand assets may be located at each node.

The determination of flow-based parameters is based on the expected market outcome. This approximation is also called the base case. As the market result depends on flow-based parameters, we first performed an initial nodal optimal power flow and calculated provisional flow-based parameters using the market results of this nodal optimal power flow. These provisional flow-based parameters were then used to perform a zonal market clearing, which approximates the zonal market results mentioned above. The flow-based parameters were determined based on this approximation. Next, they were used to simulate the day-ahead market, another zonal calculation in which the capacity is allocated and the market is cleared. This resulted in the scheduled generation, commercial exchanges, and mar-

[2] Austria, Belgium, Czech Republic, France, Germany, Luxemburg, Netherlands, Poland, Slovakia, Slovenia,.

ket-clearing prices. The last step involved translating the commercial exchanges into physical exchanges via the known nodal positions of the scheduled generators and optimizing the redispatch to ensure a stable grid state without overloaded network elements.

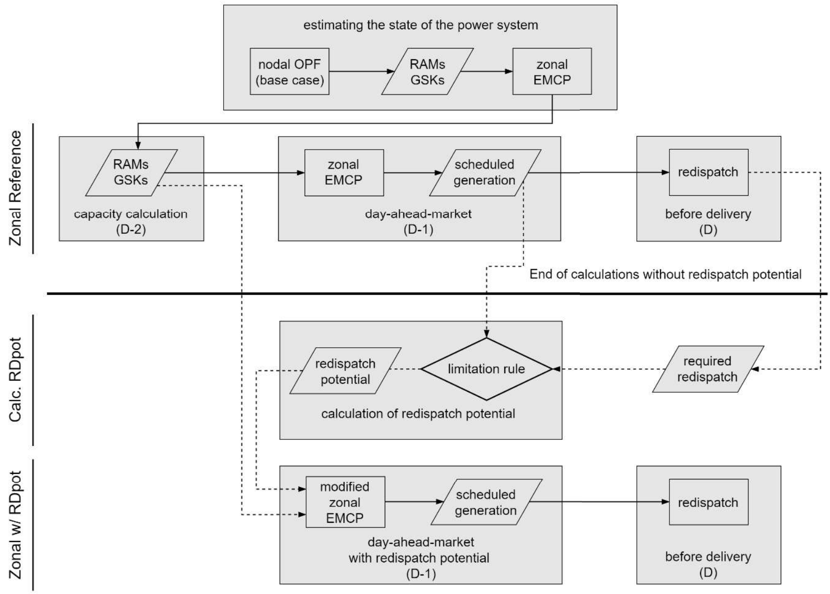


Figure 3: Model flow with a redispatch potential (based on Voswinkel et al. 2019)

The CNEs, the network elements for which the RAMs were calculated and which constrain the flow-based domain, were chosen based on the grid topology. All interconnectors—that is, network elements that cross zonal borders—were automatically considered CNEs. Additionally, internal network elements may qualify as critical if their sensitivity to zonal exchanges exceeds 5% (ACER 2019).

3.1.2 The market coupling problem with a redispatch potential

Integrating the redispatch potential requires further development and partial integration of the capacity calculation and redispatch planning processes. However, the determination of the capacity calculation inputs and the base case simulation remain unchanged in the proposed modeling framework. This also applies to the redispatch model, which uses the state variables of the market coupling model.

Consequently, the main difference is incorporating the redispatch potential and the corresponding decision variables into the day-ahead market-clearing problem, as presented in the following paragraphs.[3] Newly added variables and equations are bolded, and decision variables are in capital letters:

$$min \sum_{u \in U} c_u \cdot G_u + \boldsymbol{f^+ c_u \cdot RDpot_u^+} + \boldsymbol{f^- c_u \cdot RDpot_u^-} \tag{1}$$

s.t.

$$NEX_z = \sum_{u \in U_z} G_u - \sum_{i \in I_z} d_i \qquad \forall z \in Z \tag{2}$$

$$\sum_{z \in Z} NEX_z = 0 \tag{3}$$

$$\boldsymbol{\sum_{u \epsilon U} RDpot_u^+ + RDpot_u^- = 0} \tag{4}$$

$$g_u^{min} \leq G_u + \boldsymbol{RDpot_u^+} + \boldsymbol{RDpot_u^-} \leq g_u^{max} \qquad \forall u \in U \tag{5}$$

$$\boldsymbol{0 \leq RDpot_u^+ \leq rdpot_u^{max,+} \qquad \forall u \in U} \tag{6}$$

$$\boldsymbol{rdpot_u^{min,-} \leq RDpot_u^- \leq 0 \qquad \forall u \in U} \tag{7}$$

$$\begin{aligned} ram_f^{nsfd} \leq \sum_{z \in Z} ptdf_{f,z} \cdot NEX_z + \boldsymbol{\sum_{i \in I} \sum_{u \epsilon U^{RD-}} ptdf_{f,i} \cdot RDpot_u^-} \\ + \boldsymbol{\sum_{i \in I} \sum_{u \epsilon U^{RD+}} ptdf_{f,i} \cdot RDpot_u^+} \leq ram_f^{sfd} \qquad \forall f \in F^{CNE} \end{aligned} \tag{8}$$

In equation (1), the cost minimization problem is extended by adding the expected costs associated with the inclusion of the positive and negative redispatch potentials $RDpot_u^+$ and $RDpot_u^-$, respectively. Both potentials are multiplied by the associated costs c_u and redispatch penalty factors f^+/f^- (see section 3.1.3). This ensures that the redispatch potential is utilized only when the additional costs related to the (expected later activation of the) redispatch potential are smaller than the cost savings from the additional cross-zonal exchange. The use of the redispatch potential is limited by equation (8). Equation (2) represents the demand balance and ensures that the net position is positive (negative) when generation exceeds (falls below) the electricity demand of the respective market zone.

Equation (4) balances the redispatch potential included in each time step. In equation (5), the redispatch potential is added to the generator constraints, ensuring that the generation plus the redispatch potential stays within the generator's technical limits. Equations (6) and (7) specify that the utilized redispatch potential $RDpot_u^+$ and $RDpot_u^-$ for each generation unit *u* must stay within the external

[3] Please note that this contribution assumes the inclusion of the redispatch *potential only in the market clearing*, not the *activation* of redispatch units during the market-clearing stage, as considered in Poplavskaya et al. (2020).

bounds $rdpot_u^{max,+}$ and $rdpot_u^{max,-}$ set via the limitation rules (see section 3.2). Finally, the impact of the redispatch potential on line flows on critical network elements and the corresponding enlargement of the flow-based domain, as described in section 2, is modeled in equation (8). Here, the total redispatch potential in each direction per node i is multiplied by the nodal sensitivity $ptdf_{f,i}$ of the redispatch unit and may increase the cross-zonal exchange expressed via the zonal net position NEX_z, which is limited by the RAMs ram_f^{sfd} and ram_f^{nsfd} in the standard flow direction (sfd) and non-standard flow direction ($nsfd$) on each CNE f. If the minimum margins to increase cross-border trade are considered, ram_f^{sfd} and ram_f^{nsfd} are redefined as

$$ram_{f,minRAM}^{nsfd} = \min\left(ram_f^{nsfd}, \quad R_{amr} \cdot F_{max}^{nsfd}\right) \ \ \forall f \in F^{CNE} \tag{9}$$

$$ram_{f,minRAM}^{sfd} = \max\left(ram_f^{sfd}, \quad R_{amr} \cdot F_{max}^{sfd}\right) \ \ \forall f \in F^{CNE} \tag{10}$$

where R_{amr} is the minRAM factor (e.g., 70%) and $F_{max}^{sfd/nsfd}$ is the thermal capacity of the CNE in the standard and non-standard flow direction. Accordingly, the available margins were ensured to correspond to the minimum threshold of the thermal line capacity.

3.1.3 Redispatch problem

As explained in section 2, imperfections in FBMC necessitate corrective measures by the TSOs after the market clearing to maintain stable grid operations. Redispatch is carried out to avoid the violation of physical grid constraints by adjusting the output of generators.[4]

In modeling terms, this translates to an optimization problem in which the costs for generation adjustments are minimized, subject to the nodal transmission constraints, in accordance with a nodal DC load flow model. Due to regulations regarding cost compensation, which acknowledge the increased costs of generators when included in the redispatch process, penalty factors are introduced, adjusting the marginal costs of the generators (Higher Regional Court Dusseldorf, Decision of 4/28/2015). A general penalty is added, multiplied by the absolute amount of re-dispatched energy. This minimizes re-dispatched quantities and accounts for current practice, minimizing the number of interventions of TSOs in the market results.

The nodal balance q_i of the redispatch problem is given by the scheduled generation output g_u^* resulting from the market coupling optimization plus the activated positive and negative redispatch volumes $\mathrm{RD_u^+} + \mathrm{RD_u^-}$ of all generation units u minus the vertical load d_i connected to the node i:

[4] The redispatch problem is only briefly described in this section. For more details, including related equations, see Voswinkel et al. (2019).

$$q_i = \sum\nolimits_{u \in U_I} (g_u^* + RD_u^+ + RD_u^-) - d_i \quad \forall z \in ZZ \tag{11}$$

The redispatch amounts for each generator are constrained by their minimum and maximum power outputs. Additionally, the generation capacity already dispatched in the market-clearing problem is subtracted from the generation constraints—a generator fully utilized in the market-clearing problem is not available for positive redispatch. Still, the generator can reduce its output through negative redispatch.

$$g_u^{min} - g_u^* \leq RD_u^+ + RD_u^- \leq g_u^{\max} - g_u^* \tag{12}$$

3.2 Determination of the redispatch potential

In this contribution, we differentiate between the redispatch potential given to the market-clearing algorithm—that is, the generation units available for redispatch measures (see section 3.1.2)—and the activated redispatch measures optimized during the redispatch stage (see section 3.1.3). Determining the redispatch potential within the capacity calculation and redispatch planning process requires an estimation of the expected dispatch of generation units. Based on this estimation, the generation units available for redispatch can be identified.

From a market perspective, giving all the identified redispatch potential to the market-clearing algorithm would be preferable. However, as outlined in section 2, the generation units available for redispatch are estimated by TSOs two days before delivery. In addition to the uncertainty of the estimates, the effectiveness of a redispatch measure on congestion depends on the generator's location. Consequently, it would not be expedient to make all redispatch potential available to the market-clearing algorithm, as the uncertainty might lead to deviations between the market and redispatch stages and infeasible physical outcomes requiring further redispatch measures. In the following section, we propose three different limitation rules for determining and limiting the redispatch potential.

3.2.1 Available redispatch units

Available redispatch units are determined according to the expected dispatch of the standard zonal FBMC described in section 3.1.2 without considering the redispatch potential. This assumes perfect foresight of TSOs regarding generation and load forecasts and the system state. Based on their many years of experience with grid operation and redispatch, TSOs have a clear picture of the available redispatch units depending on the system state. Nevertheless, forecast deviations may lead to overloads of transmission lines during grid operation and deviations in the availability of redispatch units, which ultimately cause additional costs (Kloubert et al. 2015). However, forecast errors similarly affect the analyzed design options, but their detailed modeling is beyond the scope of this contribution.

The capacity of the generation units available for redispatch is determined as follows:

- Generation capacity available for **negative redispatch potential**: a reduction of the generation output $g_u^{zonal,ref}$ to provide negative redispatch requires an operation of the respective generation unit in the zonal reference. Otherwise, the negative redispatch potential $RDpot_u^-$ of generation unit *u* has to be zero. This is expressed by the assignment:

$$rdpot_u^{min,-} = -g_u^{zonal,ref}|u \notin U^{curtailment} \tag{13}$$

  The market-based curtailment of renewable energy is correspondingly excluded from the negative redispatch potential.
- Generation capacity available for **positive redispatch potential**: an increase in generation to provide positive redispatch requires an operation of the respective generation unit in the zonal reference case below the maximum power output g_u^{max}. Otherwise, the available positive redispatch potential $RDpot_u^+$ of generation unit *u* has to be zero. This implies the assignment:

$$rdpot_u^{max,+} = g_u^{max} - g_u^{zonal,ref} \tag{14}$$

The two parameters $rdpot_u^{min,-}$ and $rdpot_u^{max,+}$ serve as input for the extended market coupling problem and constrain the utilized redispatch potential during the market-clearing stage (see equations (6) and (7)).

3.2.2 Redispatch potential limitation rules

The limitation of the redispatch potential is motivated by the trade-off between increasing cross-zonal trade during the market-clearing stage and maintaining system security during the redispatch (and grid operation) stage. Integrating all available redispatch potential units into the market based on nodal sensitivities would come close to computing a nodal dispatch. This might give rise to the question why the zonal dispatch is maintained at all. More importantly, the limitation of the redispatch potential avoids a too extensive re-optimization of the zonal market-clearing solution and thus facilitates the TSO task of maintaining system security. Also, redispatch potential might otherwise reduce the objective function value (see equation [1]) without increasing cross-border trade.

The first limitation rule addresses the effectiveness of redispatch measures. Accordingly, interventions on market outcomes of generation units with low sensitivities on CNEs are avoided. In other words, ineffective redispatch units are excluded. Using the subsets for negative and positive redispatch potential based on equations (13) and (14), pairwise sensitivities between generation units available for $rdpot_u^{min,-}$ and $rdpot_u^{max,+}$ are determined. According to equations (17) and (18), the generators u

forming the top three highest pairwise sensitivities on a CNE for both directions are selected for the redispatch potential given to the market-clearing algorithm. One generator can be part of several pairs.

I. **RDpot_sens**: a limitation to generation units available for negative or positive redispatch potential and having sufficiently large pairwise sensitivities on the considered CNEs. The rule is best expressed mathematically by first constructing the set of sensitivities S_f for a given CNE f for all pairs of generators $(u, u') \in U^+ \times U^-$. Thereby U^+ and U^- indicate the set of generators with available redispatch potentials strictly different from zero, i.e.:

$$U^+ = \{u \in U \mid rdpot_u^{max,+} \neq 0\} \tag{15}$$
$$U^- = \{u \in U \mid rdpot_u^{min,-} \neq 0\}$$

The set of sensitivities S_f can then be written as:

$$S_f = \{(ptdf_{f,u} - ptdf_{f,u'}), u \in U^+, u' \in U^-\} \tag{16}$$

Introducing the notation $S_f^{(k+)}$ and $S_f^{(k-)}$ to indicate the subset of the top k positive (largest) and negative (lowest) values of set S_f, the available redispatch potentials under this rule can then be written as:

$$rdpot_{u'}^{min,-,sens} = \begin{cases} rdpot_{u'}^{min,-} & iff\ \exists f\ \exists u, (ptdf_{f,u} - ptdf_{f,u'}) \in S_f^{(3+)} \cup S_f^{(3-)} \\ 0 & otherwise \end{cases} \tag{17}$$

$$rdpot_u^{max,+,sens} = \begin{cases} rdpot_u^{max,+} & iff\ \exists f\ \exists u', (ptdf_{f,u} - ptdf_{f,u'}) \in S_f^{(3+)} \cup S_f^{(3-)} \\ 0 & otherwise \end{cases} \tag{18}$$

The second limitation rule emphasizes the maintenance of system security. Suppose that a CNE is congested in the zonal reference case. In this case, the utilization of the redispatch potential aims to increase the flow capability of this CNE (which increases the flow-based domain). Consequently, the zonal dispatch changes and cross-zonal trade increases. However, using the redispatch potential combined with the zonal approximation based on generation shift keys tends to increase the (physical) congestion on the CNE and the redispatch needed during grid operation. Finally, suppose there is a considerable need for redispatch already in the zonal reference case. In that case, considering the same generation units as the redispatch potential may lead to a conflict of use. Accordingly, the generation units activated for redispatch in the zonal reference case are excluded from the set of redispatch potentials determined according to the first limitation rule.

II. **RDpot_sens_red**: a limitation to generation units or generation capacity considered under the first limitation rule and, in addition, not activated during (or "scheduled" for) the redispatch stage:

$$rdpot_{u'}^{min,-,sens\_red} = \begin{cases} rdpot_{u'}^{min,-,sens} & iff\ \exists u, RD_u^{+,ref} = 0 \wedge RD_u^{-,ref} = 0 \\ 0 & otherwise \end{cases} \tag{19}$$

$$rdpot_{u}^{max,+,sens\_red} = \begin{cases} rdpot_{u}^{max,+,sens} & iff\ \exists u, RD_u^{+,ref} = 0 \wedge RD_u^{-,ref} = 0 \\ 0 & otherwise \end{cases} \tag{20}$$

The variables $RD_u^{+,ref}$ and $RD_u^{-,ref}$ in equations (19) and (20) result from the nodal balance constraint of the redispatch problem of the zonal reference case (see equation [9]).

4 Case study

4.1 Data and scenario framework

For the case study, a scenario was built considering the state of grid development and renewables expansion in 2022. Input prices, such as fuel and CO_2 prices, are based on pre-crisis expectations (before 2022) to suppress temporary distortions in electricity markets in our analysis (see also below).

The spatial coverage of the transmission grid comprises the Core region (i.e., Central Western Europe[5] plus Switzerland) and Central Eastern Europe.[6] The grid model is based on publicly available data, such as the static grid models of the TSOs and information provided in the network development plans (50Hertz et al. 2019; Entso-E 2021b; JAO 2021; OpenStreetMap 2021). Approximately 2,700 nodes and more than 5,000 branches were modeled, including lines and transformers. Additionally, interconnectors to all other European countries were incorporated. Accordingly, the market simulations were carried out for Europe considering a hybrid approach in which the FBMC method is applied for the Core region and the NTC approach for the remaining countries.

Information for generation capacities was taken from the generator database of the Chair of Energy Economics and Management Science at the University of Duisburg-Essen. This database is mainly based on the Platts power plant database and enhanced by plant-specific research, primarily relying on plant owners' web presence and press releases and publicly available databases (Platts 2018; Bundesnetzagentur 2020). Aggregate installed capacities for conventional power plants and renewable-

[5] Austria, Belgium, France, Germany, Luxembourg, the Netherlands, and Switzerland
[6] Poland, Czech Republic, Hungary, Slovenia and Slovakia

based generators were adjusted to match the status quo of 2022 (Entso-E 2021a). The nodal distributions of the installed capacity of renewable and conventional power plants are shown in Figure 4 and Figure 5.

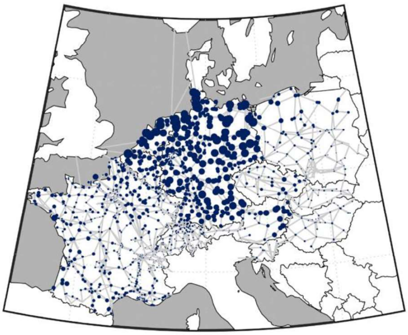


Figure 4: Distribution of renewable nodal capacity

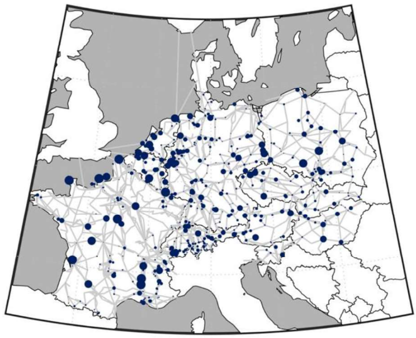


Figure 5: Distribution of conventional nodal capacity

Electricity demand data were generally taken from Entso-E (2021a). For Germany, load data from Baginski et al. (2018) were used to have more appropriate forecast data for the calculation year 2022. Renewables infeed values were modeled for each node based on a bottom-up approach, as described in Felten et al. (2019) and different input data sources (Baginski et al. 2018; Open Power System 2020). To derive a conservative estimation of the impact of the discussed methodological adaptions, we took the renewables infeed of the low value scenario for 2022 specified by Baginski et al. (2018) for Germany as one of the major redispatch demand–driving countries.

For the input prices, we used quotations for fuel and CO_2 futures from the European Electricity Exchange—that is, three-month averages of the 2025 futures price notations from the fourth quarter of 2021 (before the Ukraine crisis) for coal, natural gas, fuel oil, and light oil and CO_2 certificates (energate 2022). The price for CO_2 certificates amounts to 70.15 EUR/t CO_2. The prices for nuclear and lignite are based on values used in the German Grid Development Plan (50Hertz et al. 2019). The range of the marginal costs obtained for the corresponding generation technologies in the Core region is shown in Figure 6.

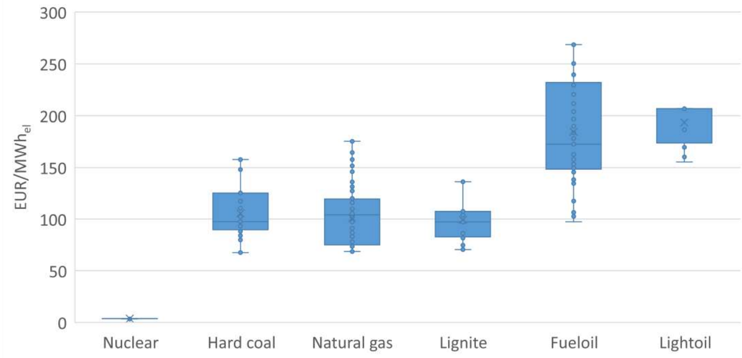


Figure 6: Marginal generation costs of conventional generation technologies

A threshold of 70% was considered for implementing minRAMs in the Core region. This assumption deviates from the current practices of Core TSOs, which partly foresee derogations from the 70% minRAM before 2025 (e.g., Germany, Poland, and the Netherlands) (ACER 2022). For modeling the integrated redispatch potential during the market-clearing stage and activating redispatch measures during the redispatch phase, penalty factors of 0.9 for downward and 1.2 for upward redispatch were assumed. As described in section 3.1.3, the penalty factors consider current compensation rules and support the efficient utilization of redispatch.

4.2 The obtained redispatch potential and its utilization

First, the results of determining the redispatch potential are presented. Subsequently, the use of the redispatch potential in the market-clearing model is illustrated before discussing the impacts of this potential in the next subsection. Theoretically, the model region's re-dispatchable maximum generation capacity amounts to 250.8 GW. The generation units available for positive and negative redispatch potentials vary depending on the respective generation and load situations (see section 3.2.1). Considering all available redispatch units in the set of redispatch potentials according to equations (13) and (14), the hourly positive full redispatch potential varies between 55.8 GW and 207.2 GW, averaging 135.3 GW. The range of negative redispatch potentials is similar, from −58.8 to −195.0 GW (−120.4 on average). For comparison, the redispatch potential without limitation is referred to as **RDpot_unlim** in the following paragraphs.

- **RDpot_sens** excludes ineffective redispatch units and considers only available redispatch units with a high positive impact (i.e., nodal sensitivity) on network elements at maximum capacity. Under this limitation rule, the hourly positive (negative) redispatch potential ranges from 36.0 to 102.4 GW (−48.6 to −126.8 GW). These are very high capacities that are fed into the market-clearing algorithm as redispatch potentials. This is because the figures include generation units that are part of the top three highest pairwise sensitivities on any CNE for both directions. However, the results reveal that despite the large redispatch potential, only small fractions (around $\pm$0.9 GW on average) are utilized in the market-clearing algorithm (see also Figure 7).

- **RDpot_red** emphasizes the maintenance of system security and considers a limitation to generation units or generation capacity not activated during (or "scheduled" for) the redispatch stage of the zonal reference case. The resulting redispatch potential is very close to the full redispatch potential, as only activated (or scheduled) generation units are excluded. Accordingly, the hourly positive (negative) redispatch potential ranges from 33.8 to 99.6 GW ($-$47.3 to $-$126.4 GW). Again, these very high values may be counterintuitive in the context of system security. Yet, this limitation rule addresses the use conflict described in section 3.2.2 and leaves the selection of efficient redispatch potentials to the market-clearing algorithm (limiting the possibility of discretionary choices by TSOs).

Figure 7 shows the hourly net effect of the utilized redispatch potential on welfare resulting from market clearing and redispatch. In most hours, using a redispatch potential increases welfare. However, there are situations in which the utilization of the redispatch potential results in an overall welfare loss—that is, when redispatch costs exceed the welfare gain in the market-clearing stage. The occurrence of such situations can be reduced when applying the limitation rules. For instance, under RDpot_sens_red, the number of hours with welfare loss are reduced from 2,475 to 1,838 hours per year. This reduction corresponds to a decrease in inefficiency from 22 to 13 M€ per year. Further details on the market impacts are discussed in the next section.

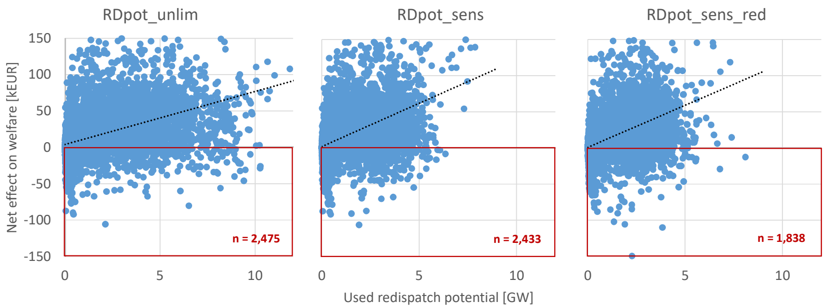


Figure 7: Utilized redispatch potential and its net effect on system costs for all 8,760 hours of one year

The following exemplary situation illustrates the utilization of redispatch potentials in the model. Figure 8 shows all modeled transmission lines in gray and CNEs always binding in the market coupling problem in black. The red lines indicate binding CNEs without the utilization of a redispatch potential, and the green lines show CNEs binding after the utilization of the redispatch potential. The utilized redispatch potential is characterized by red (downward) and blue (upward) triangles. Moreover, the black open triangles show the changes in the zonal net position after utilizing the redispatch potential.

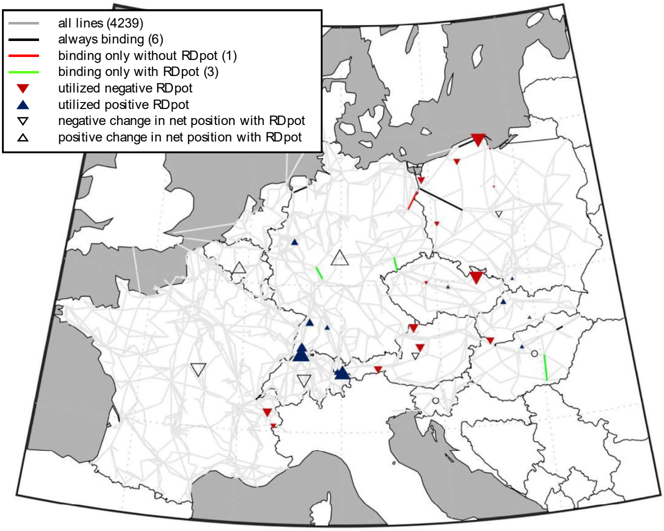


Figure 8: Utilization of the redispatch potential and binding CNEs for a situation with a high load and wind infeed

The exemplary situation with high electricity demand and a high infeed from wind energy (mainly at the northern periphery of the Core region) is associated with considerable north–south transit. Without a redispatch potential, Germany's export is limited by a CNE at the border to Poland. France has to counter the flows from the north and export more than 8.5 GW. With a redispatch potential, binding CNEs and net positions change. The utilization of a redispatch potential in southern Germany and Switzerland relaxes the available margins on internal CNEs in Germany, moving the "center of gravity" for constrained lines further south.

Consequently, Germany's net position increases by 6.7 GW during the market-clearing stage. At the same time, France and Switzerland reduce their net positions by 5 GW and 4 GW, respectively. Overall, cross-border trade can be increased by 1.4 GWh for the considered hour in the Core region. This case shows that including a redispatch potential allows for more efficient utilization of the existing infrastructure. The following section discusses the results at the aggregate level.

4.3 Impact of the redispatch potential

4.3.1 Market effects

Both concepts (i.e., minimum margins and redispatch potential) aim to improve cross-border trade and price convergence. Relaxing transmission constraints using minimum margins or redispatch potentials enlarges commercial capacities (i.e., flow-based domains) and increases cross-border trade.

Figure 9 shows the effects on **cross-border trade** for the *minRAM* and *RDpot* cases individually and combined. Among the individual measures, *minRAM* exhibits the largest increase in cross-border trade. The sweeping increase in commercial capacities enabled by *minRAM* allows for a considerable increase in cross-border trade of 19.3 TWh, corresponding to 14%. Including redispatch potentials induces higher cross-border trade by up to 9.5 TWh (8%). Depending on the limitation rule and the limitation of the redispatch potential, the benefits regarding cross-border trade are reduced to 6.1 TWh compared to the reference case. Combining *minRAM* and *RDpot* allows for a total increase in cross-border trade in the range of 20.7 to 22.0 TWh.

Regarding the effect of *minRAM* on individual countries, imports to Belgium are reduced (−38%), while imports to Poland (+142%) and Germany (+14%) are increased. Exports from Germany, France, the Netherlands, and Hungary increase substantially. At the same time, *RDpot* increases imports mainly to Central Eastern Europe and Switzerland, while exports from Germany and France increase. Overall, the effects are less extreme and somewhat balanced across the capacity calculation region under *RDpot*. Both approaches, *minRAM* and *RDpot*, result in reduced CO_2 emissions, mainly driven by increased exports of wind energy from Germany and nuclear energy from France, balancing decreased generation using coal and lignite in Germany and Poland (see Figure 16 in the Appendix).

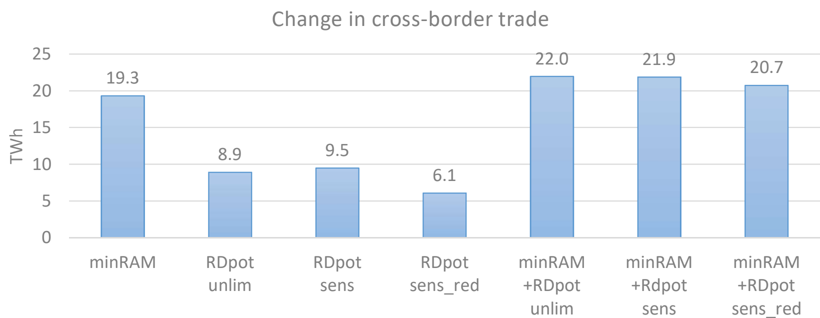


Figure 9: Change in cross-border trade (sum of exports) compared to the reference case

A decrease in price spreads is equivalent to an increase in the **convergence of electricity market prices** and a result of reduced (market) congestion between the bidding zones. The average price spread is defined as the average difference between the hourly minimum and maximum price in the Core region over all hours of the year.

The changes in the average price spread compared to the reference case are shown in Figure 10. The results follow the same pattern as the change in cross-border trade discussed above. However, the gap between the *minRAM* and *RDpot* solutions is smaller than the change in cross-border trade. Without combining the design options, the application of minimum margins decreases the average price spread to a greater extent (−6.1 €/MWh) than the inclusion of redispatch potentials (up to −4.9 €/MWh), whereby the more restrictive limitation rules induce smaller changes (−3.1 €/MWh for the most restrictive rule). Combining *minRAM* with *RDpot* allows for a considerable further decrease in price spreads. In this case, limiting the redispatch potential has no substantial effect.

The decrease in price spreads can be attributed to two effects. On the one hand, both concepts lead to a reduction in (very) high electricity prices. Based on a maximum price of 182.68 €/MWh under the reference case, applying minimum margins leads to a decrease of 61.65 €/MWh. When including redispatch potentials, the reduction of peak prices ranges from 40.48 to 50.87 €/MWh. Combining both approaches achieves a decrease in the range of 57.80 to 70.72 €/MWh. Consequently, *minRAM* and *RDpot* contribute to stable electricity prices in high-demand and scarcity situations. On the other hand, *minRAM* reduces the occurrence of negative prices resulting from increased cross-border exchange capabilities and the reduced market-based curtailment of wind generation.

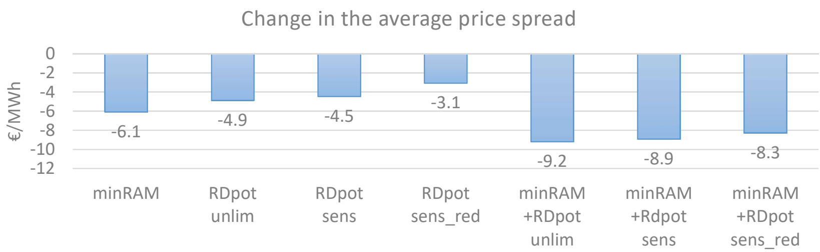


Figure 10: Change in the average price spread compared to the reference case

From a market perspective, increased cross-border trade and price convergence due to minimum margins and redispatch potential are expected to reduce market-clearing costs. However, the analysis needs to incorporate the redispatch stage to provide a complete picture of the overall (socio-economic) cost. Figure 11 shows the differences in the **market-clearing and redispatch costs** compared to the reference case. Moreover, the net welfare effect (i.e., the change in total costs) is depicted in the boxes.

For *minRAM*, the sweeping increase of commercial capacities to a minimum level reduces market-clearing costs while significantly increasing redispatch costs. The latter overcompensates the cost reduction, leading to an overall welfare loss of 107.3 M€ per year, confirming the results of Schönheit et al. (2021a). Nevertheless, this welfare loss must be regarded under the assumed minRAM of 70% and redispatch penalty factors, which may overestimate the actual redispatch costs. At the same time, the implicit assumption of a perfectly coordinated cross-border redispatch might underestimate redispatch costs.[7]

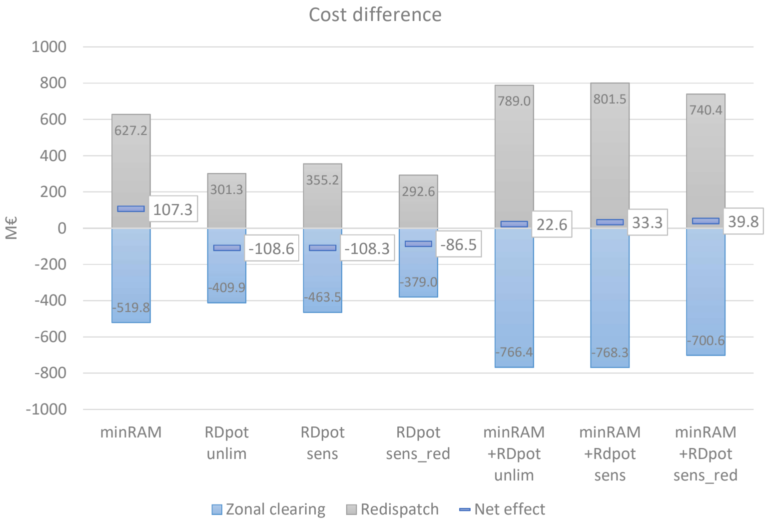


Figure 11: Change in the market-clearing and redispatch costs compared to the reference case

Using the redispatch potential approach allows for increasing commercial capacities whenever efficient—that is, when the costs of securing increased commercial capacities via (virtual) redispatch are lower than the achieved reduction of market-clearing costs. Compared with *minRAM*, considering redispatch potentials reduces the effects on market-clearing and redispatch costs, and an overall welfare gain of up to 108.6 M€ per year is obtained. Market-clearing costs in *RDpot_sens* are lower than in *RDpot_unlim*. Here, limiting the redispatch potential reduces the re-optimization of the zonal market clearing where it does not increase cross-border trade. Consequently, there is a shift from market

[7] For perspective, a sensitivity with a low minRAM of 31% leads to an overall cost decrease of 39.3 M€ per year, which is still less than the cost reductions achieved with an integrated redispatch potential. Moreover, excluding the redispatch penalty factors reduces redispatch costs for all scenarios, whereby the comparison of minRAM and the redispatch potential is not impacted.

clearing to (calculatory) redispatch potential costs (see Figure 17 in the Appendix), and cross-border trade is slightly higher under *RDpot_sens*.[8]

Combining *minRAM* with *RDpot* helps to partly compensate for the adverse effects of *minRAM*. However, an overall welfare loss is still observable, albeit reduced to about 40 M€ per year.

Both approaches, minimum margins and redispatch potentials, lead to an increase in cross-border trade. In light of the cost effects discussed above, Figure 12 puts the change in total costs in relation to the additional cross-border trade.

For *minRAM*, the additional costs associated to the additional cross-border trade of 19.3 TWh leads to an average cost of 5.6 €/$MWh_{addTrade}$. In contrast, in all three cases with redispatch potential alone, the additional cross-border trade leads to an overall cost reduction that may reach up to 14.2 €/$MWh_{addTrade}$. *RDpot_sens_red* achieves the best benefit-to-effort ratio.

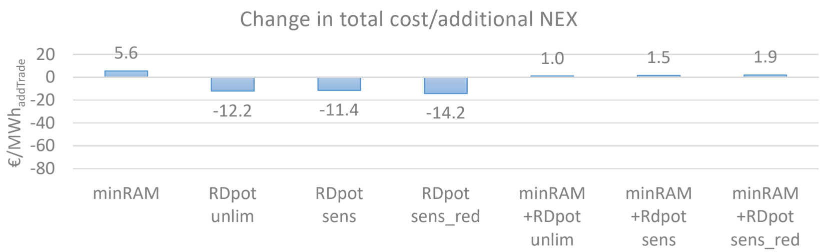


Figure 12: Change in total costs per unit of additional cross-border trade compared to the reference case

4.3.2 System operation effects

The analysis of the effects of introducing minimum margins or redispatch potentials revealed increased redispatch costs. This indicates an increase in grid congestion, which implies a trade-off between economic benefits and system security.

Corresponding to the increased redispatch costs, Figure 13 shows the change in the average number of **violated transmission constraints** implied by the market results before the utilization of congestion management measures (i.e., redispatch). The pattern corresponds to the results detailed above, with the redispatch potential leading to a third of the additional violated constraints in the *minRAM* case.

[8] The re-optimization of the zonal market clearing using the redispatch potential could be avoided by implementing an additional constraint suppressing the overall negative costs for the redispatch potential (see also equation [1]). However, this constraint may also suppress situations where pairs of negative and positive redispatch potentials have negative costs and increase cross-border trade. Consequently, we refrained from implementing such a constraint.

Combining *minRAM* and *RDpot* leads to the most violated constraints. In this regard, limiting the redispatch potential according to the second limitation rule (*RDpot_sens_red*) has a noticeable impact on the number of violated constraints.

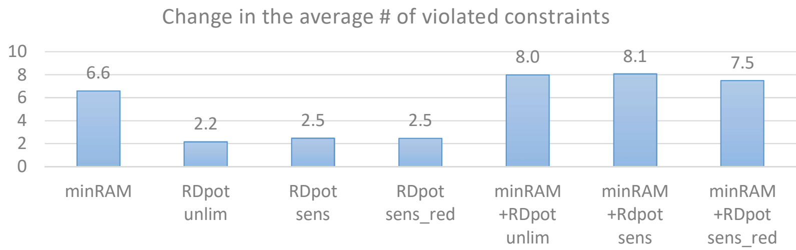


Figure 13: Change in the average number of violated constraints compared to the reference case

Figure 14, showing the change in the average number of **redispatch measures**, and Figure 15, depicting the overall change in **redispatch volume**, exhibit similar patterns. An increased number of violated constraints comes with an increase in redispatch measures and volumes. Under *minRAM*, the redispatch volume rises by 13.2 TWh, which is almost +20% compared to the 67 TWh in the reference case in the Core region.[9] Including redispatch potentials increases volumes by only 5.8 TWh to 8.1 TWh, depending on the limitation rule. Again, the combination of *minRAM* and *RDpot* leads to the highest overall increase.

The higher increase in redispatch volumes under the *minRAM* approach links back to the economic impact. The sweeping definition of minimum margins leads to more violations of grid constraints and a higher number of required redispatch measures. To relieve grid congestion, TSOs must increasingly rely on less efficient redispatch units (i.e., located farther from grid congestion and with lower sensitivity to the congested grid element), leading to higher (specific) redispatch costs.

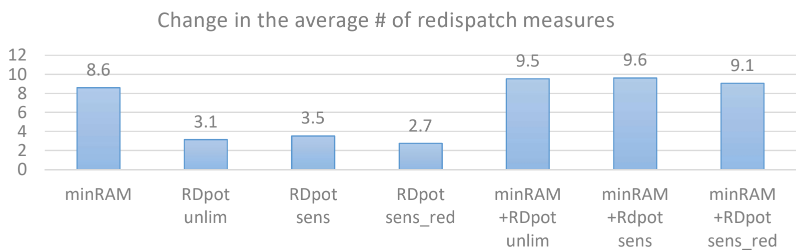


Figure 14: Change in the average number of redispatch measures compared to the reference case

[9] For comparison, the total redispatch volume in the considered model region in 2020 was 47 TWh (ACER and CEER 2021).

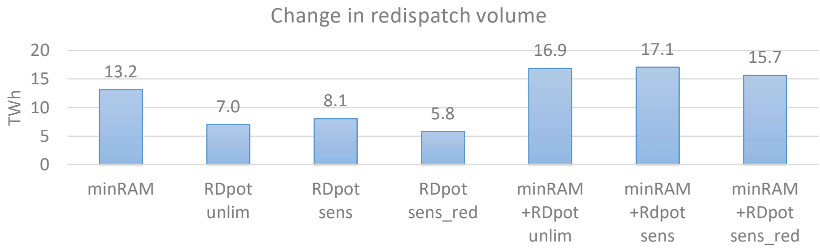


Figure 15: Change in redispatch volume (positive and negative) compared to the reference case

When assessing the outcomes of the analysis, some limitations of the approach must be kept in mind. Notably, the model includes cross-border redispatch, which is currently not perfectly coordinated in the Core region. Regarding integrating the redispatch potential, we may overestimate the resulting benefits in the market-clearing stage. Moreover, the model may underestimate redispatch amounts and costs, affecting all cases. Conversely, negating derogations from the 70% minRAM before 2025 and phase-shifting transformers may imply higher redispatch volumes and costs. Additionally, positive economic real-world implications for investors, which may emerge due to higher price convergence, were not taken into account in this paper. Nevertheless, as the analysis is based on comparing alternative cases relative to a reference case, we do not expect the conclusions to be affected by changes in these assumptions.

5 Conclusions and policy implications

This contribution is motivated by the ongoing debate about further developing the European electricity markets and congestion management. Several recent studies discussed market design options to increase cross-border trade through market-based redispatch mechanisms (Hirth and Schlecht 2020; Schlecht and Hirth 2021; Ehrhart et al. 2022). This paper focuses on an option in which the TSO is responsible for providing the market with the available redispatch potential. It sheds light on the market and system operation effects of minimum margins (i.e., minRAMs) vs determining and integrating redispatch potentials using a large-scale model covering (most of) the Core capacity calculation region.

The results show that the minRAM approach is most effective regarding additional cross-border trade and increasing price convergence. However, a sweeping increase in cross-border capacities can lead to inefficiencies. Moreover, implementing the minRAM approach leads to a higher frequency of violations during grid operation and an increased need for redispatch measures. Our analysis reveals an overall cost increase (i.e., welfare loss) due to the minRAM approach, mainly caused by higher redispatch costs outweighing the benefits of increased cross-border trade.

The inclusion of redispatch potentials is equally effective regarding improved price convergence in the Core region, whereas the increase in cross-border trade is lower than within the minRAM approach. However, using redispatch potentials leads to a more efficient increase in cross-border capacities and helps reduce system costs (or increase welfare) in the Core region. These results align with the findings yielded by a small test system studied by Poplavskaya et al. (2020). The approach mimics the physical reality of the underlying transmission system more closely and hence induces fewer additional redispatch measures compared to the minRAM method.

Based on our analysis, it can be concluded that including redispatch potentials is more efficient for increasing cross-border trade and maintaining system security and should replace the minRAM approach. However, removing the already implemented minRAM method may be politically difficult to push through. This may be even more true since a primary political objective at the European level has always been to facilitate cross-border trade.

If this political objective remains a dominant premise, the already implemented minRAM approach should be extended to include redispatch potentials. In the case of combining the two methods, our analysis indicates an improved net effect of roughly 80 M€ per year. Accordingly, potential efficiency losses due to the minRAM approach can be compensated for by efficiency gains through integrating redispatch potentials into the market clearing. The latter effect is mainly driven by the further increase in the RAM (beyond 70%) by utilizing redispatch potentials when efficient. However, politically accelerated renewables integration plans and sector coupling may significantly increase redispatch volumes and costs in the future if progress regarding network extension remains (too) slow. If this induces a shift in political targets toward system security, including redispatch potentials is a promising alternative for maintaining high cross-border exchanges while limiting redispatch volumes in the transmission grid.

When implementing redispatch potentials in the market-clearing stage, additional aspects are relevant:

- TSOs face a trade-off between defining and giving redispatch potentials to the market and securing grid operation. The selection of generation units for the redispatch potential affects its effectiveness and efficiency. Redispatch potential limitations and selection rules, as proposed in this contribution, may help to overcome or alleviate this issue. Moreover, the definition of the redispatch potential should be embedded in a monitoring process to avoid potentially arbitrary choices made by TSOs.
- Implementing redispatch potentials in the market clearing should align with the respective compensation rules for redispatch. This maintains consistent economic incentives between the market-clearing and redispatch stages.

- The integrated redispatch potential approach may be extended to other market time frames, such as intraday. When moving closer to real time, interactions between utilizing redispatch potentials to increase cross-border trade and the resulting redispatch requirements may yet create challenges for system operation and security. Moreover, day-ahead markets are more critical in terms of trading volumes and potential welfare impacts.
- To fully utilize the benefits of including redispatch potentials, the coordination of redispatch measures between TSOs and across bidding zones should be improved. This improvement would support the efficiency of integrating redispatch potentials and the overall efficiency of the FBMC process, including minRAMs.
- Poplavskaya et al. (2020) mention the issue of potential strategic bidding when generators are often activated for integrated redispatch. However, this contribution assumes only the consideration of redispatch *potentials* and not the *activation* of redispatch units during the market-clearing stage. Nevertheless, the issue of potential strategic bidding remains. In the case of cost-based redispatch, including redispatch potentials does not yet create additional strategic bidding potential. Consequently, changes in the zonal delimitation should rather be used to address structural congestions (that amplify potential strategic bidding).

From a broader perspective, integrating national power markets into a single European market forms a central pillar for a secure, sustainable, and efficient electricity supply. Increasing cross-border trade supports market integration. Grid expansion is the primary measure used to increase cross-border capacities in the long run. In the mid-term, the bidding zone review addresses congestions limiting cross-border trade. Before real-time system operation, the market coupling process and remedial actions—such as redispatch—allocate existing cross-border capacities. Consequently, integrating redispatch potentials in market clearing addresses the short-term optimization of existing capacities and supports cross-border trade and market integration in Europe. According to the results of this study, the design options considered contribute to limiting electricity prices in high-demand and scarcity situations.

The recent strengthening of climate and renewable energy targets is associated with an increased need for transport infrastructure to integrate climate-neutral generation and flexibility into energy systems. Against the background of already delayed grid expansion plans and indecisive past bidding zone reviews, short-term measures, such as implementing redispatch potentials in market clearing, are effective for increasing cross-border trade and support the integration of renewable energy sources.

CRediT authorship contribution statement

Michael Bucksteeg: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing (original draft), writing (reviewing and editing), and visualization. **Simon Voswinkel**: methodology, software, validation, formal analysis, investigation, data curation, writing (original draft), writing (reviewing and editing), and visualization. **Gerald Blumberg**: data curation (grid model), validation, and writing (reviewing and editing).

Acknowledgements

We thank Lars Nolting, Pieter Schavemaker, and Christoph Weber, as well as the participants in the 43rd IAEE International Conference in Tokyo, for their valuable comments and the interesting discussions.

References

50Hertz; Amprion; TenneT; TransnetBW (2019): Grid Development Plan 2030 (2019). Available online at https://www.netzentwicklungsplan.de/en/grid-development-plans/grid-development-plan-2030-2019, updated on 9/15/2022, checked on 9/15/2022.

ACER (2019): Day-ahead capacity calculation methodology of the Core capacity calculation region. Available online at https://documents.acer.europa.eu/Official_documents/Acts_of_the_Agency/ANNEXESTODECISIONOFTHEAGENCYNo022019/Annex%20I%20-%20ACER%20Decision%20on%20Core%20CCM.pdf.

ACER (2022): Action plans: Overview and main characteristics. ACER Report on the result of monitoring the MACZT Derogations. Available online at https://acer.europa.eu/Official_documents/Acts_of_the_Agency/Publications%20Annexes/ACER%20Report%20on%20the%20result%20of%20monitoring%20the%20MACZT%20Generic/ACER%20Report%20on%20the%20result%20of%20monitoring%20the%20MACZT%20Derogations.pdf, checked on 1/17/2023.

ACER; CEER (2021): Annual Report on the Results of Monitoring the Internal Electricity and Natural Gas Markets in 2020. Electricity Wholesale Markets Volume. Available online at https://www.acer.europa.eu/electricity/market-monitoring-report.

Antonopoulos, Georgios; Vitiello, Silvia; Fulli, Gianluca; Masera, Marcelo (2020): Nodal pricing in the European internal electricity market. In *Publications Office of the European Union*. DOI: 10.2760/41018.

Baginski, P.; Bellenbaum, J.; Beran, P.; Broll, R.; Felling, T.; Jahns, C. et al. (2018): Mittelfristprognose zur deutschlandweiten Stromerzeugung aus EEG geförderten Kraftwerken für die Kalenderjahre 2019 bis 2023. Essen. Available online at https://www.netztransparenz.de/portals/1/Content/EEG-Umlage/EEG-Umlage%202019/20181011_Abschlussbericht%20EWL.pdf.

Bertsch, Joachim; Brown, Tom; Hagspiel, Simeon; Just, Lisa (2017): The relevance of grid expansion under zonal markets. In *The Energy Journal* 38 (01). DOI: 10.5547/01956574.38.5.jber.

Bjørndal, Endre; Bjørndal, Mette; Cai, Hong (2014): Nodal pricing in a coupled electricity market. In *11th International Conference on the European Energy Market (EEM14)*, pp. 1–6. DOI: 10.1109/EEM.2014.6861222.

Bjørndal, Endre; Bjørndal, Mette; Cai, Hong; Panos, Evangelos (2018): Hybrid pricing in a coupled European power market with more wind power. In *European Journal of Operational Research* 264 (3), pp. 919–931. DOI: 10.1016/j.ejor.2017.06.048.

Bjørndal, Endre; Bjørndal, Mette Helene; Rud, Linda; Rahimi Alangi, Somayeh (2017): Market Power Under Nodal and Zonal Congestion Management Techniques. In *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3066663.

Bjørndal, Mette; Jornsten, Kurt (2001): Zonal Pricing in a Deregulated Electricity Market. In *The Energy Journal* 22 (1). DOI: 10.5547/ISSN0195-6574-EJ-Vol22-No1-3.

Blumberg, Gerald; Schneller, Christian; Schuster, Henning (2022): Redispatch 3.0: Regulatorischer Rahmen, Markt- und Produktdesign. Zielmodell für eine ergänzende marktbasierte Einbindung kleinteiliger dezentraler Flexibilitäten in den Redispatch-Prozess. Available online at https://www.transnetbw.de/_Resources/Persistent/c/4/6/9/c469f1b0ef6bae7e7bf7260b0b22bdcb29d83db0/221013_Bericht-Redispatch3.0_final1.pdf.

Bundesnetzagentur (2020): Kraftwerksliste der Bundesnetzagentur - Stand: 01.04.2020. Available online at https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/Erzeugungskapazitaeten/Kraftwerksliste/Kraftwerksliste_2020_1.html, updated on 9/15/2022, checked on 9/15/2022.

CREG (2017): Functioning and design of the Central West European day-ahead flow based market coupling for electricity: Impact of TSOs Discretionary Actions. Available online at https://www.creg.be/sites/default/files/assets/Publications/Studies/F1687EN.pdf.

Deilen, Caroline; Felling, Tim; Leisen, Robin; Weber, Christoph (2019): Evaluation of Risks for Electricity Generation Companies through Reconfiguration of Bidding Zones in Extended Central Western Europe. In *The Energy Journal* 40 (01). DOI: 10.5547/01956574.40.SI1.cdei.

Egerer, Jonas; Weibezahn, Jens; Hermann, Hauke (2016): Two price zones for the German electricity market — Market implications and distributional effects. In *Energy Economics* 59, pp. 365–381. DOI: 10.1016/j.eneco.2016.08.002.

Ehrhart, Karl-Martin; Eicke, Anselm; Hirth, Lion; Ocker, Fabian; Ott, Marion; Schlecht, Ingmar; Wang, Runxi (2022): Congestion management games in electricity markets. Mannheim: ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung (ZEW Discussion Papers, 22-060). Available online at https://www.econstor.eu/handle/10419/266651.

Elia Group (2019): Future-proofing the EU energy system towards 2030. Available online at https://www.elia.be/-/media/project/elia/shared/documents/elia-group/publications/studies-and-reports/20191212_future_proofing_eu_system_2030.pdf.

energate (2022): Market data. Available online at https://www.energate-messenger.de/markt/, updated on 9/15/2022, checked on 9/15/2022.

Entso-E (2021a): ENTSO-E Transparency Platform. Available online at https://transparency.entsoe.eu, updated on 9/15/2022, checked on 9/15/2022.

Entso-E (2021b): Ten-Year Network Development Plan 2020. Available online at https://tyndp2020-project-platform.azurewebsites.net/projectsheets, updated on 9/15/2022, checked on 9/15/2022.

European Parliament (2019): Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (Text with EEA relevance.). Available online at http://data.europa.eu/eli/reg/2019/943/oj, updated on 7/13/2022, checked on 7/13/2022.

Felling, Tim (2019): Solving the Bi-level Problem of a Closed Optimization of Electricity Price Zone Configurations using a Genetic Algorithm. In *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3425831.

Felling, Tim; Felten, Björn; Osinski, Paul; Weber, Christoph (2023): Assessing Improved Price Zones in Europe: Flow-Based Market Coupling in Central Western Europe in Focus. In *EJ* 44 (01). DOI: 10.5547/01956574.44.6.tfel.

Felling, Tim; Weber, Christoph (2018): Consistent and robust delimitation of price zones under uncertainty with an application to Central Western Europe. In *Energy Economics* 75, pp. 583–601. DOI: 10.1016/j.eneco.2018.09.012.

Felten, Björn; Felling, Tim; Osinski, Paul; Weber, Christoph (2019): Flow-Based Market Coupling Revised - Part I: Analyses of Small- and Large-Scale Systems. In *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3404044.

Felten, Björn; Osinski, Paul; Felling, Tim; Weber, Christoph (2021): The flow-based market coupling domain - Why we can't get it right. In *Utilities Policy* 70, p. 101136. DOI: 10.1016/j.jup.2020.101136.

Grimm, Veronika; Martin, Alexander; Sölch, Christian; Weibelzahl, Martin; Zöttl, Gregor (2022): Market-Based Redispatch May Result in Inefficient Dispatch. In *EJ* 43 (01). DOI: 10.5547/01956574.43.5.csol.

Grimm, Veronika; Martin, Alexander; SSlch, Christian; Weibelzahl, Martin; ZZttl, Gregor (2018): Market-Based Redispatch May Result in Inefficient Dispatch. In *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3120403.

Grimm, Veronika; Martin, Alexander; Weibelzahl, Martin; Zöttl, Gregor (2016): On the long run effects of market splitting: Why more price zones might decrease welfare. In *Energy Policy* 94, pp. 453–467. DOI: 10.1016/j.enpol.2015.11.010.

Heilmann, Erik; Klempp, Nikolai; Hufendiek, Kai; Wetzel, Heike (2022): Long-term contracts for network-supportive flexibility in local flexibility markets. Marburg: Philipps-University Marburg, School of Business and Economics (MAGKS Joint Discussion Paper Series in Economics, 24-2022). Available online at https://www.econstor.eu/handle/10419/266022.

Henneaux, Pierre; Lamprinakos, Panagiotis; Maere d'Aertrycke, Gauthier de; Karoui, Karim (2021): Impact assessment of a minimum threshold on cross-zonal capacity in a flow-based market. In *Electric Power Systems Research* 190, p. 106693. DOI: 10.1016/j.epsr.2020.106693.

Higher Regional Court Dusseldorf, Decision of 4/28/2015, case number VI-3 Kart 313/12 (V).

Hirth, Lion; Schlecht, Ingmar (2020): Market-Based Redispatch in Zonal Electricity Markets: The Preconditions for and Consequence of Inc-Dec Gaming. In *ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg*. Available online at http://hdl.handle.net/10419/222925.

JAO (2021): Static Grid Model. Core TSOs data. Available online at https://www.jao.eu/static-grid-model, updated on 9/15/2022, checked on 9/15/2022.

Kloubert, Marie-Louise; Schwippe, Johannes; Müller, Sven Christian; Rehtanz, Christian (2015): Analyzing the impact of forecasting errors on redispatch and control reserve activation in congested transmission networks. In *2015 IEEE Eindhoven PowerTech*, pp. 1–6. DOI: 10.1109/PTC.2015.7232716.

Kristiansen, Tarjei (2020): The flow based market coupling arrangement in Europe: Implications for traders. In *Energy Strategy Reviews* 27, p. 100444. DOI: 10.1016/j.esr.2019.100444.

Kunz, Friedrich; Zerrahn, Alexander (2015): Benefits of coordinating congestion management in electricity transmission networks: Theory and application to Germany. In *Utilities Policy* 37, pp. 34–45. DOI: 10.1016/j.jup.2015.09.009.

Lang, Lukas Maximilian; Dallinger, Bettina; Lettner, Georg (2020): The meaning of flow-based market coupling on redispatch measures in Austria. In *Energy Policy* 136, p. 111061. DOI: 10.1016/j.enpol.2019.111061.

Marien, Alain; Luickx, Patrick; Tirez, Andreas; Woitrin, Dominique (2013): Importance of design parameters on flowbased market coupling implementation. In *2013 10th International Conference on the European Energy Market (EEM)*, pp. 1–8. DOI: 10.1109/EEM.2013.6607298.

Martin, Palovic; Christine, Brandstätt; Gert, Brunekreeft; Marius, Buchmann (2022): Strategic behavior in market-based redispatch: International experience. In *The Electricity Journal* 35 (3), p. 107095. DOI: 10.1016/j.tej.2022.107095.

Neuhoff, Karsten; Barquin, Julian; Bialek, Janusz W.; Boyd, Rodney; Dent, Chris J.; Echavarren, Francisco et al. (2013): Renewable electric energy integration: Quantifying the value of design of markets for international transmission capacity. In *Energy Economics* 40, pp. 760–772. DOI: 10.1016/j.eneco.2013.09.004.

N-SIDE (2023): N-SIDE makes pan-European, single-day-ahead coupling possible with EUPHEMIA. Available online at https://energy.n-side.com/resources/case-studies/euphemia.

Open Power System (2020): European power system data. Available online at https://data.open-power-system-data.org.

OpenStreetMap (2021): OpenStreetMap. Available online at https://openstreetmap.org, updated on 9/2/2022, checked on 9/15/2022.

Platts (2018): Platts PowerVision. Available online at https://www.spglobal.com/commodityinsights/en/products-services/electric-power/powervision, updated on 9/15/2022, checked on 9/15/2022.

Poplavskaya, Ksenia; Totschnig, Gerhard; Leimgruber, Fabian; Doorman, Gerard; Etienne, Gilles; Vries, Laurens de (2020): Integration of day-ahead market and redispatch to increase cross-border exchanges in the European electricity market. In *Applied Energy* 278, p. 115669. DOI: 10.1016/j.apenergy.2020.115669.

Schlecht, Ingmar; Hirth, Lion (2021): Dispatch Hub Compensation Schemes - Study on profit impacts and strategic incentives of alternative compensation schemes for Dispatch Hubs in the Flex-in-Market concept. Available online at https://neon.energy/Neon-Dispatch-Hubs-Elia.pdf.

Schönheit, David; Dierstein, Constantin; Möst, Dominik (2021a): Do minimum trading capacities for the cross-zonal exchange of electricity lead to welfare losses? In *Energy Policy* 149, p. 112030. DOI: 10.1016/j.enpol.2020.112030.

Schönheit, David; Kenis, Michiel; Lorenz, Lisa; Möst, Dominik; Delarue, Erik; Bruninx, Kenneth (2021b): Toward a fundamental understanding of flow-based market coupling for cross-border electricity trading. In *Advances in Applied Energy* 2, p. 100027. DOI: 10.1016/j.adapen.2021.100027.

Schönheit, David; Weinhold, Richard; Dierstein, Constantin (2020): The impact of different strategies for generation shift keys (GSKs) on the flow-based market coupling domain: A model-based analysis of Central Western Europe. In *Applied Energy* 258, p. 114067. DOI: 10.1016/j.apenergy.2019.114067.

Trepper, Katrin; Bucksteeg, Michael; Weber, Christoph (2015): Market splitting in Germany – New evidence from a three-stage numerical model of Europe. In *Energy Policy* 87, pp. 199–215. DOI: 10.1016/j.enpol.2015.08.016.

van den Bergh, Kenneth; Boury, Jonas; Delarue, Erik (2016): The Flow-Based Market Coupling in Central Western Europe: Concepts and definitions. In *The Electricity Journal* 29 (1), pp. 24–29. DOI: 10.1016/j.tej.2015.12.004.

Voswinkel, Simon; Felten, Björn; Felling, Tim; Weber, Christoph (2019): Flow-Based Market Coupling – What Drives Welfare in Europe's Electricity Market Design? In *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3424708.

Zerrahn, Alexander; Kunz, Friedrich (2016): Coordinating Cross-Country Congestion Management: Evidence from Central Europe. In *The Energy Journal* Volume 37 (Sustainable Infrastructure Development and Cross-Border Coordination). Available online at https://ideas.repec.org/a/aen/journl/ej37-si3-zerrahn.html, checked on 2/25/2021.

Appendix

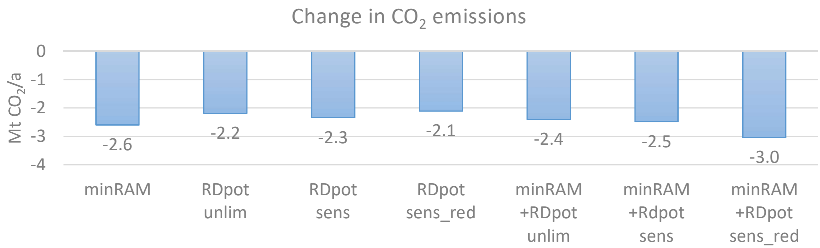


Figure 16: Change in CO_2 emissions compared to the reference case

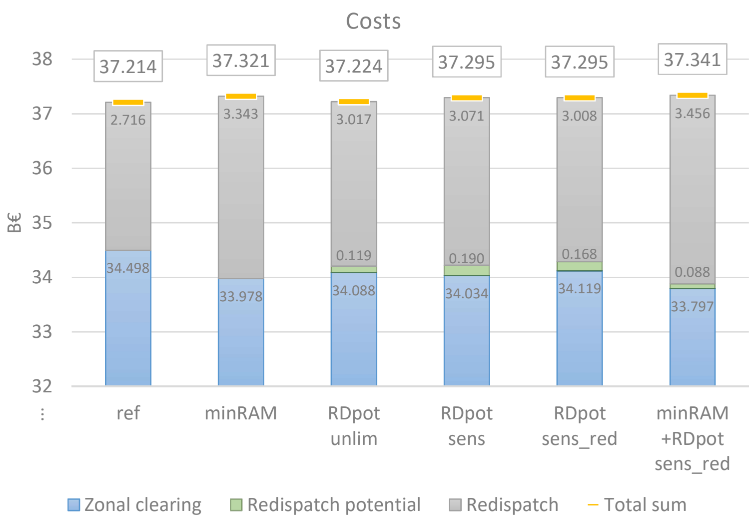


Figure 17: Absolute cost values, including (virtual) costs for redispatch potential

Chapter 5

Sharing congestion management costs among system operators using the Shapley value

by Simon Voswinkel, Jonas Höckner, Abuzar Khalid, and Christoph Weber

published in 2022 in Applied Energy 317

Applied Energy 317 (2022) 119039

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Sharing congestion management costs among system operators using the Shapley value

Simon Voswinkel *, Jonas Höckner, Abuzar Khalid, Christoph Weber

University of Duisburg–Essen, Universitätsstr. 2, 45141 Essen, Germany

ARTICLE INFO

Keywords:
Congestion management
Cost sharing
Shapley value
Approximation methods
Power grids
Game theory

ABSTRACT

With energy generation becoming increasingly decentralized, the need for congestion management across grid voltage levels is also increasing. To enable fair sharing of congestion costs among grid operators, these costs must be allocated to congested grid elements. We propose using the Shapley value for this purpose. The Shapley value is a cooperative game theory concept that was developed to share a total surplus generated by a coalition of players between the players based on their marginal contributions to the coalition. We apply this concept to share the costs of congestion management between grid elements based on their contributions to overall congestion management costs. To reduce the computational complexity of the Shapley value, we introduce two novel simplification approaches and compare them to existing methods using a numerical example based on CIGRE benchmark grids. The first method exploits the fact that the characteristic function for the congestion costs is obtained from an optimal power flow computation (i.e., a constrained optimization problem). It utilizes knowledge about which constraints are non-binding in the optimization to derive the values of related coalitions without calculating them. The second method takes advantage of the fact that the congestion management cost-allocation game is monotone and derives the values of coalitions based on this property. Both methods are implemented and compared to sampling. Using the first method, we are able to reduce computational complexity to less than 20% of that of the original problem while maintaining exact results. Our second approach is not dependent on detailed knowledge of the underlying optimization problem and can reduce the computational time by almost half with exact results and much further when compromising precision. While the methods are presented through an application example, they can be applied to other games with similar properties.

1. Introduction

Congestion management at the transmission system level has been discussed in detail in recent decades [1]. Besides nodal pricing, which has been implemented in parts of the US, zonal pricing in combination with redispatch measures is the most frequently applied strategy for congestion management [2–5]. The increase in renewable capacities has led to significant changes in the electricity system, as congestion management is no longer exclusively limited to the transmission grid. Renewable generation capacities that are connected to lower grid levels are replacing major conventional power plants, which are connected to the transmission grid. Since distribution grids are not commonly designed to transport such large amounts of electricity, congestion management on lower grid levels will become increasingly important [6,7]. Furthermore, transmission system operators will gradually have to rely on flexibility options in the distribution grid to alleviate congestion in transmission grids. Consequently, flexibility options in distribution grids will become more important, and distribution system operators will also play an increasingly important role in congestion management.

In the European electricity system, there are typically different operators for the transmission and distribution grids. However, there are major national differences. Whereas there are only one transmission and one major distribution grid operator in France, Germany has more than 800 distribution grid operators at various voltage levels plus four transmission grid operators [8]. In some cases, all grid levels in a region are operated by different grid operators (i.e., there are up to four different grid operators for the different voltage levels).

Due to the high degree of interdependence between grid levels, close coordination is necessary to exploit synergies between grid operators and ensure efficient congestion management measures across grid levels [9,10]. It is clear that this is a major challenge, especially with over 800 grid operators in Germany. Therefore, an industry solution for

* Corresponding author.
E-mail address: simon.voswinkel@uni-due.de (S. Voswinkel).

https://doi.org/10.1016/j.apenergy.2022.119039
Received 15 October 2021; Received in revised form 16 March 2022; Accepted 25 March 2022
Available online 26 April 2022

coordinating redispatch measures among transmission and distribution grid operators has been developed, which is called Redispatch 2.0 [11]. It serves as a guideline for the implementation of redispatch measures across several grid levels. Specifically, the existing redispatch processes of the transmission grid operators are extended so that the distribution grid operators will take an active role in the redispatch processes in the future.

A key question is how the costs of congestion management (which we will refer to as *congestion costs*) should be allocated between all involved grid operators. This question is particularly relevant in a regulatory framework like the one in Germany, where grid operators charge the congestion costs exclusively to their grid customers. For example, customers of a grid operator with high congestion costs usually have to pay higher grid charges than customers of a grid operator with low congestion costs. Although there has been a law in place since 2017 to gradually align the grid fees of the four transmission system operators, the Grid Fee Modernization Act (NEMoG), this problem still exists at the distribution grid level. Consequently, customers of a grid operator in regions with a particularly high in-feed of renewable energy may have to pay higher grid charges because the grid operator has to resort more frequently to cost-intensive congestion management measures.

In the academic literature, several approaches have been proposed regarding the fair allocation of congestion costs. We focus on a method of cooperative game theory, the Shapley value, to allocate congestion management costs to the grid elements for which congestion management must be performed (and by extension to the grid operators owning the elements). Essentially, the share of total costs allocated to an element is calculated by the Shapley value. It reflects the element's average expected marginal cost contribution to congestion management over all possible combinations of cooperating elements, so called coalitions. In the case of congestion management, applying the Shapley value to use cases with many congested elements quickly leads to deterrent computational efforts, since a multivariable optimization problem has to be solved for each coalition to determine the corresponding costs. Due to this computational complexity, the method is difficult to apply to real-world examples unless approximation and alternative calculation strategies are applied.

2. Contribution

Existing approaches to reduce computational complexity have the disadvantage that either the computational precision of the cost shares is lost, as in the case of approximation methods like sampling, or the methods are only applicable to very specific types of cooperative games (for a review of different methods see for example [12,13]). To the best of our knowledge, no specific method has been developed for a cooperative game with the characteristics of a redispatching problem based on a central optimization problem.

This paper aims to fill this knowledge gap and contributes to the scientific literature in the following ways:

- Two novel methods that simplify the calculation of the Shapley value by exploiting the specific properties of the underlying characteristic function without sacrificing precision are developed:
  - The first method exploits the fact that the characteristic function for the congestion costs is obtained from an optimal power flow computation (i.e., a constrained optimization problem). It utilizes knowledge about which constraints are non-binding in the optimization to derive the values of related coalitions without calculating them.
  - The second method takes advantage of the fact that the congestion management cost-allocation game is monotone and derives the values of coalitions based on this property. In this case, detailed knowledge of the underlying optimization problem is not necessary.

  Both methods allow the derivation of the costs of many coalitions without having to calculate each of them by solving the underlying constrained optimization problem. Since the computation time depends proportionally on the number of optimization problems to be solved, it is shown that significant savings in computational effort can be achieved by the proposed methods.
- The simplification methods are applied to an illustrative CIGRE power grid with seven grid operators and eleven congestions to demonstrate their applicability. We show that the developed calculation methods correctly determine the Shapley value and that the Shapley value is superior to other cost sharing mechanisms, which is in line with existing research.
- The performances of the developed simplification methods are also compared to established approximation methods like sampling. It is shown that the developed methods have significant advantages over the approximation methods that have been previously used.

In summary, the focus and novelty of the paper lies in introducing two new approaches to calculate the Shapley value and enable its use in real world applications. The main focus is not on the superiority of the Shapley value over other cost sharing methods, which has already been shown in several existing papers. Nevertheless, the benefits are addressed in the paper and it is shown that our results are in line with the research that has already been conducted. Although the methods are presented using the example of congestion cost allocation, they could also be used for other application cases with similar characteristics.

The paper is organized as follows. In Section 3, we introduce the Shapley value along with previous energy-related applications and existing approximation methods. Section 4 begins by defining congestion management in game theoretic terms before detailing the main contribution of this paper, the two new simplification methods. The methods are applied to a case study in Section 5 and compared to other approaches. The paper concludes with Section 6.

3. General approach and background

3.1. The Shapley value

The Shapley value was originally introduced to fairly allocate the utility (e.g., money) achieved through the cooperation of all participating players in transferable utility games [14]. However, the method can also be applied to allocate the costs of a grand coalition among its players [15,16]. This section presents the relevant properties of the Shapley value within a cost-allocation problem and derives the mathematical formulation based on [17].

Let a cost-allocation problem be a pair (N, c), where N is the set of players, and $c : 2^N \to \mathbb{R}$ is a map that assigns to every possible coalition $S \subset N$ the cost $c(S)$. The goal is to allocate the costs of the *grand coalition* $c(N)$ among all players $i \in N$. To achieve this, an allocation rule Φ needs to be defined, which assigns to every $c \in C(N)$ a vector $\Phi(c) \in \mathbb{R}^N$ satisfying $\sum_{i \in N} \Phi_i(c) = c(N)$. $C(N)$ is the set of cost-allocation problems with every finite set of players N. Reasonable properties for an allocation rule are symmetry, additivity, and the null agent property:

Symmetry: An allocation rule Φ satisfies symmetry if it holds that $\Phi_i(c) = \Phi_j(c)$ for all $c \in C(N)$ and all $i, j \in N$ that are symmetric in c.

Null Agent Property: Φ satisfies the null agent property if $\Phi_i(c) = 0$ for all $c \in C(N)$ and all agents $i \in N$ with zero costs in c.

Additivity: Φ satisfies additivity if it holds that $\Phi(c+d) = \Phi(c)+\Phi(d)$ for all $c, d \in C(N)$.

The Shapley value is a unique allocation rule for cost-allocation problems that satisfies all three properties, and it can be mathematically formulated as follows:

$$\Phi_i(c) = \sum_{S \in N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (c(S \cup i) - c(S)), \tag{1}$$

where Φ_i is the Shapley value of player i. The Shapley value can be interpreted as the average marginal costs added to all coalitions S by player i. All possible coalitions are considered and weighted with a factor $\frac{|S|!(|N|-|S|-1)!}{n!}$ that represents the frequency of these coalitions. Besides the above-mentioned characteristics, the Shapley value also satisfies the efficiency criterion, which means that it distributes the exact cost of the grand coalition among all agents:

$$\sum_{i=1}^{N} \Phi_i(c) = c(N) \tag{2}$$

In the context of allocation of congestion management costs among grid operators, the properties of the Shapley value also characterize a fair cost allocation mechanism. It is ensured that only the actually incurred costs of congestion management are distributed via the mechanism (efficiency, cf. Eq. (2)) and grid operators without congested lines do not have to bear any costs (null agent). Furthermore, grid operators with exactly the same marginal contribution to the total costs are also allocated the same share (symmetry).

However, to get the exact Shapley value, costs of 2^N coalitions must be calculated, which results in high computational complexity. Thus, the Shapley value is not easily applicable to real-world problems with more than 25 participants [18]. This is especially true in the case of cost allocation between grid operators, since a constrained optimization problem must be solved to calculate the cost for each coalition (see Section 4.1.2).

3.2. Shapley-value: previous energy-related applications

The Shapley value has already been used in different energy-related applications. [19], for example, apply the Shapley value to allocate congestion costs to loads. They show that a solution to the Shapley value can be found for a simple illustrative example, but do not elaborate on the feasibility for a real cost-sharing scenario with many players.

[20] introduce a congestion cost-allocation method based on the Shapley value in the context of a pool market model. The goal of the paper is to allocate congestion costs to transactions through a two-step approach. The first step includes the allocation of costs to congested lines according to the Shapley value. Second, the allocation of line costs to transactions is done by applying a power flow-tracing method. [20] also compare the results of their method with the shadow price and incremental method and conduct a sensitivity analysis for different transmission constraints, generation bids, and load levels. They conclude that the Shapley value is a fair method for cost allocation with the disadvantage of high computational complexity. However, [20] state that the complexity is sufficiently reduced by applying the two-stage process, in which the costs are not allocated to all transactions directly but initially only to the congested lines via the Shapley value. According to the authors, in power systems, there are only a few lines congested at the same operation time, which makes the method feasible for allocating the congestion costs. However, in Germany, for example, a large number of grid elements are frequently overloaded at the same time, especially during periods of high renewable feed-in from wind turbines. As generation capacities are shifting to the distribution grid (cf. Section 1), the number of simultaneously overloaded grid components is likely to increase further. Therefore, the statement of [20] is highly dependent on the power system under consideration. For complex real-world conditions such as described above for the German system, alternative calculation methods are needed to apply the Shapley value.

The Shapley value and related allocation rules have also been applied in other energy-related and non-energy-related areas. Energy-related applications include, for example, loss allocation in transmission and distribution grids [21–26], allocation of fixed grid costs [27–31], allocation of transmission grid investment costs [32–37] and other energy-related applications, such as the allocation of CO2 emissions in a multi-product oil refining industry [38], the distribution of revenues among the participants in a demand response program [39], the assessment of economic signals in microgrids [40], and the provision of a profit-sharing scheme between integrated distributed energy resources pooled in a virtual power plant [41].

As the Shapley value is very well established, there are also many non-energy-related applications. Detailed overviews are given by [42] and recently [43]. An overview with a special focus on cost-allocation problems is given in [17].

3.3. Shapley value: previous approximation and simplification methods

The high computational complexity of the Shapley value can be caused by two aspects: the large number of coalitional values to be computed and/or the complexity of calculating the coalition values, depending on the properties of the characteristic function [13]. Many researchers have already tried to solve these problems, and the resulting approaches can be divided into three groups: first, alternative representation formalisms to efficiently compute the exact Shapley value [44,45]; second, developing exact efficient methods for some specific classes of games [46]; and third, bounded approximate solutions [47–49]. A detailed overview of different strategies to overcome the computational complexity is provided by [12]. Besides the description of complete representation languages for characteristic function games like marginal contribution nets, [12] also present solutions for several combinatorial optimization games, such as induced subgraph games, network flow games, and minimum cost-spanning tree games. Recently, [43] have presented some specific applications where the exact Shapley value can be obtained, avoiding the complex calculation by applying algorithms that rely on the decomposition of a given characteristic function.

To the best of our knowledge, no specific method for a cooperative game with the characteristics of a redispatching problem based on a central optimization problem has been developed yet. Therefore, only approximation and simplification methods for general classes of games in characteristic function form can be considered for the redispatching problem. Out of these, sampling is the most commonly applied approximation method in the literature.

[50] were the first to develop a simple random sampling algorithm, where samples are drawn from the population of all permutations to estimate the Shapley value. Based on this work, several enhancements of sample-based approximation methods were developed. [39,51] propose stratified sampling approaches to reduce the number of samples needed to attain a certain estimation accuracy. Based on their previous paper, [52] also propose stratified random sampling to reduce the variance of the original estimation method. [18] introduce a method called structured random sampling that reduces the average error in the Shapley value approximation by almost 30 percent compared to simple sampling by [50]. In contrast to the preceding methods based on independent sampling, [53] propose an ergodic sampling algorithm that uses pairs of negatively correlated samples to reduce the estimation error, which, in some instances, dramatically improves the quality of the estimation.

While the above mentioned sampling methods bring significant improvements in the computation time of the Shapley value, such approximation methods sacrifice the accuracy of the computed values. While sampling methods have been improved and higher accuracy can be achieved, problem-specific methods can often significantly reduce calculation time without sacrificing accuracy at all. Notably, we are not aware of any problem-specific approaches to improve the Shapley value for congestion cost allocation. Therefore, we propose two new methods that can significantly reduce computational complexity in the context of a congestion management cost-allocation game. The first method exploits the fact that the characteristic function for the congestion costs is obtained from an optimal power flow computation (i.e., a constrained optimization problem). It utilizes knowledge about which constraints are non-binding in the optimization to derive the values of related coalitions, without calculating them. The second method takes advantage of the fact that the congestion management cost-allocation game is monotone and derives the values of coalitions based on this property. Both methods are implemented and compared to sampling. While the methods are presented through an application example, they can be transferred to other games with similar properties.

4. Methodology

This section defines the congestion management cost-allocation game and develops the simplifications that form the contributions of this paper. The main points are quickly summarized as follows.

As detailed in Section 3.1, the Shapley value, while originally developed to allocate utility, can also be used to fairly allocate the costs of a grand coalition among its players. In the case examined in this paper and as will be further explained in Section 4.1.2, the grand coalition is formed by all congested lines in a specific grid load case and the costs of this grand coalition are the costs for relieving these congestions – i.e., the costs of doing redispatch for a specific point in time. The characteristic function (as will be shown in Section 4.2) is an optimal power flow (OPF) calculation, that determines the optimal redispatch configuration for a given set of congested grid elements and the costs associated with that redispatch configuration. When performed for the *grand coalition*, i.e., for all congested elements in a grid load case, this yields the costs for the grand coalition, i.e., the costs that have to be redistributed to the individual grid elements that caused them.

To use the Shapley value for allocating these congestion costs, not only knowledge of the costs of the grand coalition is required, but also of the costs for every other possible configuration the congested elements that make up the grand coalition. Because the optimal power flow used for solving the redispatch problem that is necessary to gain knowledge of these costs (i.e., the characteristic function) is resource intensive, the Shapley value is difficult to compute for all but very small grand coalitions, made up of only a few congested elements. Because solving the underlying optimization problems is by far the most resource intensive part of calculating the Shapley value, reducing the amount of coalitions that have to be calculated proportionally reduces the time needed to calculate the Shapley value — if 50% of the coalitions do not need to be solved by the optimal power flow because the costs are determined with a different, much quicker, method, approximately 50% of the original time is saved.

This is where the contribution of this paper lies: we develop two methods that allow to infer the costs of many coalitions without solving the optimization problem to calculate them. The first method is detailed in Section 4.3.1 and is specific to games where the characteristic function is a linear optimization problem, as is the case in the redispatch problem considered in this paper. The second method (Section 4.3.2) can be applied more widely and applies to all games where the characteristic function is monotone (i.e., costs cannot decrease with new players joining the coalition), as is also the case in the redispatch problem (see also Section 4.2.5).

4.1. Congestion management cost-allocation as a cooperative game

4.1.1. Need for a cooperative solution to the congestion management cost-allocation

Congestions in power grids require coordinated actions by the involved parties. This is particularly true if several grid operators are involved, as the power grids are natural monopolies. A central optimization of all redispatch measures is therefore considered as the most efficient approach [54]. Yet this central optimization does not solve the question of how to allocate the arising costs to the different grid elements or grid operators involved. Here the Shapley value provides a fair cost allocation mechanism. This is valid notwithstanding the fact that grid operators, for example in Germany, have partly followed non-cooperative strategies in the past. In fact, this non-cooperative behavior was primarily encouraged by the poor existing cost-allocation mechanism.[1]

The goals of a cost-allocation mechanism are to set correct incentives and deliver fair results. While it is not necessarily guaranteed that the Shapley value corrects all misaligned incentives in the context of congestion management, it guarantees a fair distribution of the costs and in this way at minimum does not counteract cooperation among participants and decreases the risk of defection of the players from the cooperative congestion management approach.

The Shapley cost allocation approach thus complements the joint minimization of congestion management costs which is imposed on the grid operators by the regulator — based on its regulatory competences established to limit the natural monopoly power.

4.1.2. Definition of the congestion management cost-allocation game

To apply the Shapley value to the problem of allocating congestion management cost, it needs to be defined in game theoretic terms. The resulting game is called the *congestion management cost-allocation game*.

Calculating power flows in electricity grids is complex because the power flow on each line depends on the in-feed and withdrawal on every node in the grid. When certain line flows change, there are also impacts on flows in other lines. Therefore, it is not easy to determine the congestion cost that a operator should pay because all the line flows are interdependent. Since individual measures can have a positive impact on several congested lines at different voltage levels, the allocation of the overall congestion costs may be a challenging task. Thus, the Shapley value is applied to allocate costs based on the marginal contribution of each congested element to overall costs. Since the grid operators are cooperating to conduct the most efficient congestion management, the underlying problem can be formulated as a cooperative game with the following properties:

- Player i: congested line i from a set of congested lines N. Other grid elements like transformers may be overloaded and hence congested as well — yet we focus here on overloaded lines as the most frequent case of congestion.[2]
- Grand coalition N: set of all congested lines in a specific timestep
- Coalitions S: subset of congested lines from all congested lines N. $|S|$ defines the size of a coalition with regard to the number of congestions included.
- Cost function $c(S)$: congestion costs of coalition S (as determined by the optimal power flow (OPF))
- Cost function $c(N)$: overall congestion costs to be allocated (as determined by the OPF) – special case of $c(S)$, where the coalition S is the grand coalition N.

In this game, coalitions are formed by congested grid elements, and $c(S)$ is the cost of addressing a given set of congestions S. We call this approach the *per-element* Shapley value. It is also possible to group certain sets of congested elements together so that they are always regarded as a group instead of as individual players. The groups can be formed according to which grid the congested elements are located in. Because we assume that each grid is managed by a separate grid operator, this variation is labeled the *per-operator* Shapley value. If $|M|$ grid operators are considered in contrast to $|N|$ congested lines,

[1] For example, grid operators had to bear full costs of congestion management measures they instructed, even if other grid operators benefited from the same measures. This created the incentive of free-riding, i.e., grid operators withheld the instruction of measures in order to profit from congestion management measures of other grid operators.

[2] Because lines themselves cannot make decisions, this formulation implies that each line is operated independently by one decision maker. The case where multiple lines are operated jointly by one decision maker is theoretically interesting, but is not part of this analysis.

the number of coalitions that have to be calculated is reduced by a factor of $2^{|N|-|M|}$. While this kind of merging of players allows a faster calculation of Shapley values, it also leads to different results and is not the focus of this paper.[3]

A crucial part of the congestion management cost-allocation game is the characteristic function, which determines the costs $c(S)$ of each coalition. Since we assume coordinated congestion management, the measures and the resulting costs for each coalition are determined by optimal power flow calculations, linear optimization problems which are described in Section 4.2. This is a computationally expensive characteristic function, which makes the computation of Shapley values for even relatively small coalitions a challenge, emphasizing the need for the type of simplifications we develop in this paper.

4.2. Characteristic function

The characteristic function of the congestion management cost-allocation game described in Section 4.1.2 is the optimal power flow performed on the electricity grid, subject to analysis. The optimal power flow is a unit dispatch problem, where additional constraints are given by the network equations, ensuring that limits on line capacities are not exceeded. We perform the optimal power flow taking the market outcomes as starting values; hence, a unit dispatch of zero implies that the original schedule remains unchanged. The underlying market outcomes result from a single grid load case, i.e., a single point in time.

4.2.1. The optimization problem

In our application we use the Matlab toolbox *Matpower* [55] to calculate the optimal power flow (OPF), which results in the optimal redispatch. We apply the linearized DC power flow approximation to avoid issues of non-convexity and high computation times in the optimization. The following equations are adapted from the manual [56].

The vector of optimization variables is

$$x = \begin{bmatrix} \Theta \\ P^G \end{bmatrix}, \tag{3}$$

containing the phase angles Θ, which are not relevant for the interpretation of our specific problem, and the unit power output of each generator P_i^G. In a prior step, the dispatch of the generation units is optimized without line constraints, which corresponds to the zonal market outcome. The power output obtained in the OPF is then redispatched as needed to satisfy the line constraints.

The objective function minimizes the sum of all generation costs, $c_i^G(P_i^G)$, being the cost function of generator i, which in our case always uses a constant marginal cost.

$$\min_{\Theta, P^G} \sum_{i=1}^{n^G} c_i^G(P_i^G) \tag{4}$$

The optimization problem is subject to

$$B_{\text{bus}}\Theta + P^D - L^G P^G = 0 \tag{5}$$

which ensures that the power input and output on each node is balanced and

$$B_{\text{branch}}\Theta - F_{\max} \leq 0 \tag{6}$$

$$-B_{\text{branch}}\Theta - F_{\max} \leq 0 \tag{7}$$

which are the line capacity constraints, with $F_{\max}$ being the maximum line capacity. B_{bus} and B_{branch} contain the imaginary parts of the bus admittance matrix and branch admittance matrix, respectively. P^D is the inelastic power demand on each bus, and the matrix L^G links the generator output P^G to the correct buses.[4]

Lastly, the variables are bounded:

$$\theta_i = \theta_i^{\text{ref}}, \qquad i \in \mathcal{I}_{\text{ref}} \tag{8}$$

$$P_i^{G,\min} \leq P_i^G \leq P_i^{G,\max}, \qquad i = 1 \dots n^G \tag{9}$$

As congestion management is performed from a baseline, the bounds on the power output are the limits from the baseline operating point of each power plant — the room for adjustment. A fully dispatched power plant thus exhibits a $P_i^{G,\max}$ of zero, and $P_i^{G,\min}$ of its full (negative) capacity, as explained in the next section.

4.2.2. Grid model and baseline for congestion management

The underlying grid model specifies the complete grid topology and the power demands $P^{D,0}$ at the nodes for the grid load case under consideration. The market dispatch is then computed by a simple optimization, neglecting the power flow constraints and effectively collapsing all nodes into a single one.[5] With the obtained dispatch information, a power flow calculation is performed, that is, the set of Eqs. (5) is solved with given generation in-feed $P^{G,0}$ to obtain the angles Θ^0 and the corresponding power flow vector $P^{F,0} = B_{\text{from}}\Theta^0$ with elements $P_j^{F,0}$.

With the results of the power flow calculation, the violations of the line capacity limits can then be calculated. The set of all lines with exceeded capacity limits forms the grand coalition N.

In order to perform the redispatch calculation, the grid model has to be adjusted. For the resulting power outputs to be the redispatch quantities – the delta from the baseline operating points – the operating limits of the power plants are adjusted so that they correspond to the room for adjustment left by the original dispatch, resulting in $P_i^{G,\min,\text{RD}}$ and $P_i^{G,\max,\text{RD}}$, the lower and upper adjustment limits for the generators in the redispatch problem:

$$P_i^{G,\min,\text{RD}} = P_i^{G,\min} - P_i^{G,0}, \qquad i = 1 \dots n^G \tag{10}$$

$$P_i^{G,\max,\text{RD}} = P_i^{G,\max} - P_i^{G,0}, \qquad i = 1 \dots n^G \tag{11}$$

The original dispatch $P^{G,0}$ is assigned to the grid nodes via the assignment matrix L^G and subtracted from the corresponding inelastic node demand P^D, resulting in $P^{D,\text{RD}}$, the nodal power demand in the redispatch problem:

$$P^{D,\text{RD}} = P^D - L_k^G P^{G,0} \tag{12}$$

4.2.3. Calculating the coalition values

The value of the grand coalition is calculated by simply performing the OPF calculation on the grid model. To calculate the values for every other coalition, the line capacity limits of the lines *not* belonging to the coalition are modified as follows ($P_j^{F,\max,\text{new}}$ being the modified line capacity limit of a line j):

$$P_j^{F,\max,\text{new}} = \begin{cases} P_j^{F,\max}, \text{ if } P_j^{F,0} \leq P_j^{F,\max} \\ P_j^{F,0}, \text{ if } P_j^{F,0} > P_j^{F,\max}, \end{cases} \qquad j = 1 \dots n^F \tag{13}$$

If the power flow $P_j^{F,0}$ of a line j not belonging to the coalition exceeds the capacity limit $P_j^{F,\max}$, that capacity limit is increased to $P_j^{F,0}$, the power flow in the original dispatch. This relaxation of the line constraints implies that congestions outside of the current coalition do not have to be resolved, but they must not be worsened by any remedial action targeted at the other congestions.

[3] Differences are structural and cannot be compared to errors resulting from an approximation.

[4] In our application, we also consider demand-side flexibilities. These are modeled as generators with negative generation limits and are included in P^G.

[5] The model can also easily be extended to cope with multiple market zones.

4.2.4. Binding and non-binding constraints

The line constraints of a power line become binding if the constraint is a limiting factor during the optimization. A line constraint is non-binding if the line is not fully utilized after the optimization. The dual variable – the shadow price λ_j – of a non-binding constraint is zero [57].

Even lines that are part of the grand coalition – implying that they are overloaded in the baseline power flow – might not be at full capacity after the congestion management optimization. This happens when their congestion is dominated by a different congestion and relieving that congestion leads to a formerly overloaded line not being fully utilized anymore. In that case, its constraint will not be binding, and correspondingly its shadow price will be zero. Because shadow prices are returned by any linear program solver, they can be used to determine which constraints were binding in the optimization.

For a given underlying market outcome and a corresponding subset S of congested lines, a non-binding constraint implies that the result of the redispatch optimization does not change if the capacity limit that forms this constraint is removed.[6] In game theoretic terms: the costs associated with a coalition S (redispatching costs for a set of congested lines) do not change, if the player i (the non-binding line) is removed from the coalition, given that the corresponding constraint is non-binding.[7] We will use this to develop the simplification method in Section 4.3.1.

4.2.5. Monotonicity of the redispatching problem

In a linear optimization with an objective function to be minimized, the objective value cannot decrease when adding new constraints to the existing optimization problem. The program has fewer degrees of freedom with which to achieve an optimal result: either the original solution is then still feasible, which means that the objective function does not change. Or a new solution is needed as the old one is now infeasible. But this new solution must have the same or a higher objective function value. Otherwise, it would already have been preferred in the original problem.

Applied to the redispatching problem this means the costs for redispatching from a given underlying market outcome cannot decrease if a new capacity constraint is added to the optimization. For the congestion management cost-allocation game as defined in 4.1.2, this implies that the cost function is monotone – the costs of a coalition cannot decrease when another player is added. This property will be used to develop the simplification method in Section 4.3.2.

4.3. Proposed approximation and simplification methods

As described earlier, the main challenge in applying the Shapley value is to calculate the costs for all possible 2^N coalitions — the original grand coalition consisting of all congested elements in a grid load case whose associated redispatching costs must be allocated, and all possible sub-coalitions that can be formed by members of the grand coalition along with their respective redispatching costs. Therefore, a reduction of the number of costs that have to be calculated is directly proportional to the reduction of the overall computational complexity. In addition to the use of sampling, there are several problem-specific methods for calculating the Shapley value with reduced computational complexity. In this chapter, two new approaches for a cooperative game with the characteristics of a redispatching problem are proposed.

The first analytical approach uses information about binding constraints from the optimization to determine the optimal overall congestion management measures. A recursive algorithm is applied, which systematically uses knowledge of non-binding line capacity constraints to infer the costs of smaller coalitions without calculating them. This approach is especially suitable for applications in which the values of the characteristic function are determined via optimizations and information about non-binding constraints can be obtained.

The second approach can also be applied if this information is not available or if the characteristic function is not based on an optimization. The corresponding algorithm is based on the monotonicity assumption of the congestion cost function (congestion costs cannot decrease if an additional congested element is added to a given set of congested elements, see Section 4.2.5) and systematically compares the costs of larger coalitions with the costs of associated small sub-coalitions. If these costs are the same or differ only within a defined tolerance, the costs of all intermediate coalitions can be inferred without additional calculations.

4.3.1. Analytical method using non-binding constraints

Description of the method. As described in Section 4.2.4, some line capacity constraints may not be binding in the determination of the value of any given coalition. Non-binding constraints imply that the value of the target function, in this case the congestion costs, does not change if any such restriction is removed. Consequently, all coalitions that differ from a given coalition only in that some subset of the non-binding constraints (players) are removed must have the same value as the original coalition and their respective values do not have to be calculated separately, reducing the computational complexity of the problem.

We call a player j whose constraint is non-binding a *non-binding player*. The group of all non-binding players S_{NBP} of an optimization $f(S)$ (which conceptually corresponds to the cost function $c(S)$) based on the set of players S, is characterized by the shadow prices $\lambda^{f(S)}(j)$ being equal to zero[8]:

$$S_{\mathrm{NBP}}^{f(S)} = \{j \in S | \lambda^{f(S)}(j) = 0\} \tag{14}$$

It is important to mention that the set of non-binding players depends on the optimization problem considered because the constraint corresponding to a player may be binding or non-binding depending on the set of congestions considered in the optimization. Thus, removing a binding player from a coalition may convert non-binding players in the original coalition into binding players of the new coalition. However, removing a non-binding player from a coalition will never change the outcome of the optimization, including which other players are (non-)binding.

The proposed algorithm starts with optimizations of so-called *starter coalitions* and systematically infers the costs of subsets of these starter coalitions based on binding constraints. A starter coalition S_{Start} is a coalition with at least one non-binding player ($S_{\mathrm{NBP}}^{f(S_{\mathrm{Start}})} \neq \emptyset$), such that these players can be removed without changing the value. However, it is not possible to add a player without changing the value. The set of all starter coalitions is labeled as $S_{\mathrm{Start}} \in \mathcal{S}_{\mathrm{Start}}$.

A subset of a starter coalition that does not contain any non-binding players is called *base coalition*

$$S_{\mathrm{Base}}^{f(S_{\mathrm{Start}})} = S_{\mathrm{Start}} \setminus S_{\mathrm{NBP}}^{f(S_{\mathrm{Start}})} = \{i \in S_{\mathrm{Start}} | \lambda^{f(S_{\mathrm{Start}})}(i) \neq 0\} \tag{15}$$

[6] In contrast, binding constraints shape the optimization outcome. A binding constraint implies that the optimization algorithm hits a boundary in search of the optimal solution. Therefore, if a binding constraint is removed, the optimization outcome will be different, as that boundary no longer exists. Instead, one or several other (formerly non-binding) constraints will become binding.

[7] This only holds for a given grid load case. For different load or infeed situations, a completely different set of lines may be congested and form the grand coalition N associated with that new grid load case.

[8] As stated in Section 4.2.4, shadow prices can be used to determine which constraints are binding in the optimization (non-binding constraints have a shadow price of zero). This is not to be confused with other methods for cost allocation discussed in this paper, where the shadow prices themselves may be used to calculate the cost allocation.

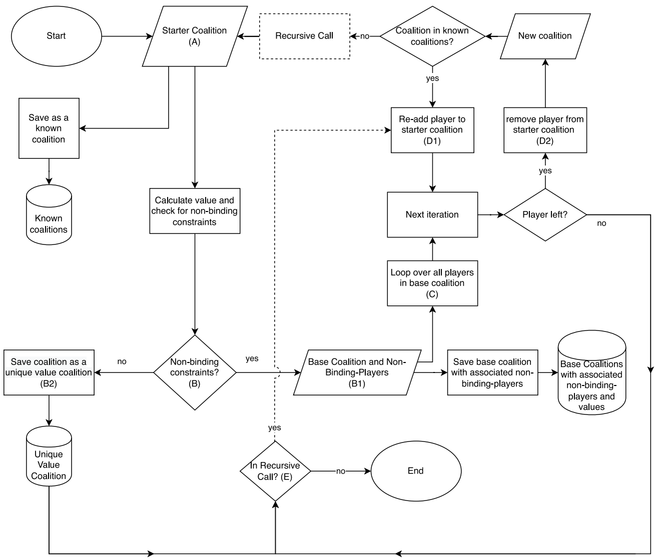


Fig. 1. Simplified process diagram for the analytical method using non-binding constraints.

The costs of base coalitions are the same as the costs of the corresponding starter coalitions and therefore do not need to be calculated again:

$$c(S_{\text{Base}}^{f(S_{\text{Start}})}) = c(S_{\text{Start}}) \tag{16}$$

The set of all base coalitions is labeled as $\mathcal{S}_{\text{Base}}$.

Coalitions that are a subset of a starter coalition and a superset of a corresponding base coalition are called *intermediate coalitions* $S_{\text{Inter}}^{f(S_{\text{Start}})}$ ($S_{\text{Base}}^{f(S_{\text{Start}})} \subset S_{\text{Inter}}^{f(S_{\text{Start}})} \subset S_{\text{Start}}$). They differ from their respective base coalition only in that non-binding players are added, and differ from their respective starter coalition only in that one or more non-binding players are removed. Consequently, they have the same costs as their respective starter and base coalitions:

$$c(S_{\text{Inter}}^{f(S_{\text{Start}})}) = c(S_{\text{Start}}) = c(S_{\text{Base}}^{f(S_{\text{Start}})}) \tag{17}$$

$$\text{where} \quad S_{\text{Inter}}^{f(S_{\text{Start}})} = S_{\text{Base}}^{f(S_{\text{Start}})} \cup T \tag{18}$$

$$\text{with} \quad T \subset S_{\text{NBP}}^{f(S_{\text{Start}})} \tag{19}$$

The set of all intermediate coalitions belonging to a certain optimization $f(S_{\text{Start}})$ is called $\mathcal{S}_{\text{Inter}}^{f(S_{\text{Start}})}$, and the overarching set of all intermediate coalitions that does not need to be calculated is $\mathcal{S}_{\text{Inter}}$.

However, there are some coalitions, where players can neither be added nor removed without changing the value. In that sense, these coalitions are at the same time base and starter coalitions. We call these coalitions *unique value coalitions* and reserve the terms base and starter coalitions for coalitions that can be used to save calculations. The set of all unique value coalitions is labeled as $\mathcal{S}_{\text{UVP}}$.

In summary, the proposed algorithm systematically searches for intermediate and base coalitions using knowledge of non-binding constraints of players of starter coalitions. Thus, only the starter coalitions must be calculated and all coalitions where only non-binding players are removed can be inferred.

Proposed algorithm for implementation. We propose a recursive algorithm. A process diagram is shown in Fig. 1, and a visualization of the recursive loop is shown in Fig. 2. Both figures have common points marked as (A), (B), (…), which will be referred to in the following description of the algorithm. Each level and column in Fig. 2 corresponds to a call to the algorithm, starting again at (A) in Fig. 1. Because of the recursive nature of the algorithm, one branch will be followed until the end of the search tree, after which the algorithm moves up one level and follows the next branch all the way down. The order of the calls in Fig. 2 is marked as *Call (1)*, *Call (2)*, etc.

The first starter coalition is the grand coalition and contains all players. The algorithm will call itself with varying starter coalitions, descending down the levels in Fig. 2 and each time starting at (A) in Fig. 1, respectively. Each starter coalition that has been fed into the algorithm is saved as a *known coalition*, which prevents it from (unnecessarily) being analyzed more than once.

The first step is to solve the optimization problem $f(S_{\text{Start}})$ (and therefore to calculate the value $c(S_{\text{Start}})$) of the input starter coalition. If there are non-binding constraints (players) (B1), the coalition of binding players is saved as a base coalition $S_{\text{Base}}^{f(S_{\text{Start}})}$, together with the non-binding players $S_{\text{NBP}}^{f(S_{\text{Start}})}$ and the value of the starter coalition $c(S_{\text{Start}})$ (which is therefore also the value of the base coalition and all derivable intermediate coalitions, cf. Eqs. (16) and (19)). This information enables the construction of all intermediate coalitions (each corresponding to the base coalition with some subset of non-binding players added, which is the same as the starter coalition with some subset of non-binding players removed), all of which have the value of the starter coalition.

To generate new starter coalitions, the algorithm loops over all players from the discovered base coalition (C) and removes one of these players from the input starter coalition in each iteration. This results in the original starter coalition with one of the binding players removed. Because a binding player is removed, the optimization result will be different, and therefore new players may become binding or

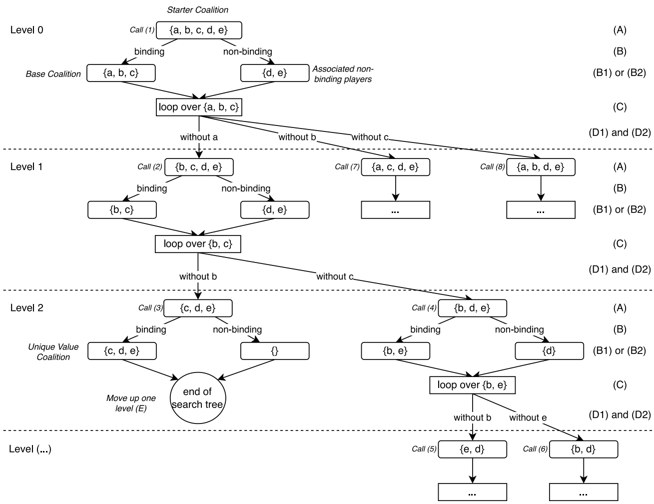


Fig. 2. Schematic view of the recursive loop for a simple example consisting of players a, b, c, d and e.

non-binding.[9] After each iteration, the player is re-added (D1) while a different player is removed (D2), so the coalition does not shrink. In each iteration, if the resulting coalition is not already known, the algorithm calls itself with this coalition as a new starter coalition (entering a new "level" in the stack in Fig. 2 and starting at (A) in Fig. 1).

The algorithm descends along the recursive tree until it finds a unique value coalition ((B2), shown on the left side in level 2 in Fig. 2). This means that no players can be added or subtracted from the coalition without changing its value. This also means that we cannot infer any other coalition values from this coalition.[10] In that case, the algorithm returns to the previous level (level 1) and enters the next iteration (adding player b (D1) and removing player c (D2)). The algorithm is finished once it has performed all iterations in the initial level of the call stack (after finishing *Call (8)*, which was initiated from level 0 in Fig. 2).[11]

4.3.2. Numerical method using value comparison

Description of the method. The approach described above is only applicable if the characteristic function is defined by an optimization problem and if the constraints corresponding to the players can be determined to be binding or non-binding. If this is not the case, the approach cannot be applied. In that case, we propose a numerical approach that can be used for all kinds of monotone characteristic function games and thus a wide range of application examples.

The approach is based on a monotonicity assumption for the cost function (i.e., it is assumed that the congestion costs cannot decrease when additional congestions are considered, see also Section 4.2.5). This means that the following applies:

$$c(S) \geq c(R) \quad \text{if} \quad S, R \subset N \quad \text{and} \quad R \subseteq S \tag{20}$$

This assumption is true for our application to congestion cost allocation because adding a congestion to a coalition of congestions cannot decrease overall costs, as the original congestions must still be resolved. However, while costs cannot decrease, in some constellations the congestion costs $c(S)$ also do not increase ($c(S) = c(R)$) despite the consideration of an additional congestion. This is the case when the measures for resolving the previously considered congestions are already sufficient to resolve the additional congestion. This property is exploited to infer the costs of other coalitions.

If the costs of a small coalition $c(S_{\text{small}})$ and a large coalition $c(S_{\text{large}})$ are equal, then the costs of all medium coalitions $c(S_{\text{medium}})$, which represent a superset of the small and a subset of the large coalition, can be derived without calculating them. This means that

$$c(S_{\text{medium}}) = C \tag{21}$$

$$\text{if} \quad c(S_{\text{small}}) = c(S_{\text{large}}) = C \tag{22}$$

$$\text{and} \quad S_{\text{small}} \subset S_{\text{medium}} \subset S_{\text{large}} \tag{23}$$

By comparing the costs of small and large coalitions, we can save calculations of medium coalition costs. This approach does not need to fulfill special assumptions or requirements for the computation of the coalition costs.

If the first condition in (21) is formulated as a strict equality, the exact Shapley values for all medium coalitions are obtained. However, this condition can be relaxed by introducing a tolerance t:

$$c(S_{\text{small}}) \in [c(S_{\text{large}}) \cdot (1-t); c(S_{\text{large}})] \tag{24}$$

The value of corresponding medium coalitions may then be approximated by a rule yet based on the costs of the small and large coalitions. A simple example of such a rule is to equate the cost of

[9] Cf. footnote 6 in Section 4.2.4.

[10] No players can be subtracted in that case because that would imply that there were non-binding constraints. It is assumed that no players can be added. If that is not the case, the same coalition will later appear (and be overwritten) as a base coalition, or it will be implied as an intermediate coalition. This does not reduce the amount of saved coalitions in our results and is therefore not further elaborated on.

[11] It is not guaranteed that each possible coalition will be discovered as a starter coalition, base coalition, intermediate coalition, or unique value coalition. It is, however, trivial to find the missing coalitions and solve their respective optimization problem.

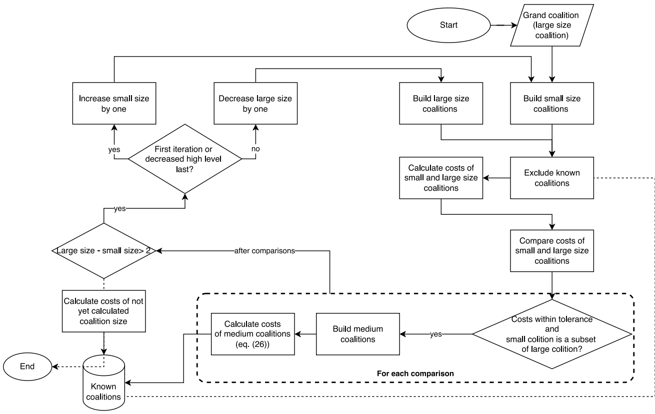


Fig. 3. Simplified process diagram for the numerical approach.

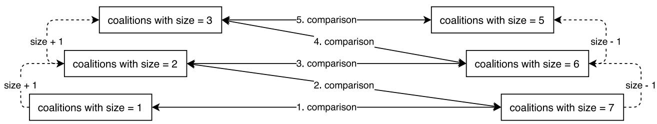


Fig. 4. Schematic visualization of the numerical approach for a simple case with seven players.

medium coalitions to the costs of the small (or the large) coalition. Another rule would be to calculate the mean value of both costs:

$$c(S_{\text{medium}}) = \frac{1}{2}(c(S_{\text{small}}) + c(S_{\text{large}})) \tag{25}$$

On the one hand, by introducing the tolerance t, a larger number of medium coalitions can be identified, and thus less calculation is required. On the other hand, the costs of the medium coalitions are no longer determined exactly, which means that the Shapley values calculated at the end are also no longer exact. This results in a trade-off between calculation complexity and accuracy of the values.

The idea behind the approach is similar to the analytical approach described in Section 4.3.1. The set difference $S_{\text{large}} \setminus S_{\text{small}}$ is comparable to the *non-binding players* in the optimization. In this approach, however, the medium coalitions are not identified by non-binding constraints but rather by a numerical comparison of the cost levels of small and large coalitions. Although this does not require the knowledge of optimization parameters, costs must be calculated for both small and large coalitions. In contrast, with the analytical method there is no need to determine the cost of the base coalitions — which are otherwise similar to the small coalitions in the numerical approach.

Proposed algorithm for implementation. To implement this approach, a stepwise algorithm is proposed. Fig. 3 shows a simplified process diagram, and Fig. 4 shows a schematic general overview of the method. It starts with the calculation of the costs for the smallest and largest coalitions (starting with $|N|$ single coalitions and the grand coalition) and then gradually reduces this gap by increasing or decreasing the coalition sizes step by step. After each calculation, the costs of all coalitions are compared. If equal (within the tolerance) costs are identified and the respective small coalition is a subset of the corresponding larger coalition, a match is found, and all medium coalitions can be derived.

The procedure is briefly described in the following:

1. Calculate coalition cost for coalitions with $|S_{\text{small}}| = 1$
2. Calculate coalition cost for coalitions with $|S_{\text{large}}| = |N|$
3. Compare the values of both levels. If a match is found (within the tolerance), check if the lower coalition is a subset of the higher coalition. Only use coalitions whose values have not been derived themselves.
4. For all pairs of small and large coalitions that fulfill step 3, build all medium coalitions that are a superset of the lower and a subset of the higher coalition, calculate the costs using Eq. (25), and add them and their values to the list of known coalitions.
5. Iteratively lower the gap of the coalition size by increasing (decreasing) the coalition size of small (large) coalitions. Calculate all coalition costs of the respective coalitions, if they are not yet included in the list of known coalitions. After every step of increasing or decreasing the coalition sizes, perform steps 3 and 4 to identify medium coalitions.

This algorithm, along with the algorithm introduced in Section 4.3.1, will be applied to a case study in the next section.

5. Results

The methods for allocating congestion costs and reducing the computational complexity of the Shapley value introduced in the previous sections are tested and compared based on a case study, which is described in Section 5.1. The Shapley value is compared to other cost-sharing methods in Section 5.2, and the results of our proposed algorithms are analyzed in Section 5.3.

5.1. Case study

The case study is based on CIGRE benchmark grids (see Section 5.1.1), and a scenario with eleven grid congestions at different grid levels is analyzed (see Section 5.1.2). On the basis of this case

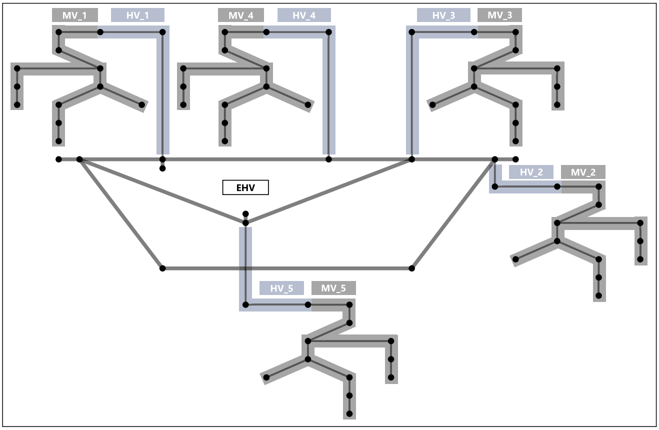


Fig. 5. Topology of grid with regions of responsibility for each grid operator.

study, the Shapley value will be compared to other cost-allocation methods (described in more detail in Appendix) and the two proposed simplification methods for calculating the Shapley value will be applied and their results compared to other existing simplification methods.

5.1.1. CIGRE grid

In order to demonstrate the use of the Shapley value for cost allocation, we use the CIGRE benchmark grids for extra-high (EHV), high (EV), and medium voltage levels (MV) [58]. Benchmark HV and MV grids are added to every EHV node with a load. The MV grids are connected to the EHV grid via an HV subgrid. In our example, there are five distribution grids at the HV level linked to the transmission (EHV) grid and one distribution grid at the MV level linked to each HV grid. The EHV grid is implemented with parameters according to the recommendations laid out in [58, pp. 17], the HV and MV grids according to the recommendations in [58, p. 34]. The model is implemented using Matlab with the open source toolbox Matpower [55]. An illustration of the grid is given in Fig. 5.

The loads at the EHV nodes are reduced by the load connected to the respective (identical) MV grids. Furthermore, 10% of each EHV load is reconnected to an HV node. We assume that this 10% stands for additional, not explicitly modeled, MV grids. To create congestions and flexibility options to resolve them, flexible loads and generators of different sizes and marginal costs are placed in the grids. These flexibility options are detailed in Table 1.

The marginal costs of the flexibility options vary greatly depending on the respective flexibility type. Thus, congestions in some grids may be eliminated by utilizing comparatively cheap flexibility options, while other grids require the use of expensive flexibility options to relieve internal congestions. As a result, it can be cost-efficient to solve congestions by using flexibility options from other grid areas as long as these options have an impact on the respective congested lines. This must be taken into account in a cost-allocation scheme.

5.1.2. Congestion management measures

Before solving the congestion management cost-allocation game, the congestion costs have to be computed. In line with the operation principles in European electricity systems, we first perform a (zonal) market dispatch that disregards the grid constraints. Then, congestion management (also known as redispatch) is performed. If the generation capacities of the illustrative example are optimally dispatched in the zonal market clearing, eleven lines are overloaded (cf. Table 2). These congestions are located at all voltage levels and therefore affect several grids.

Congestion management is then performed to alleviate all grid congestions using an OPF, as described in Section 4.2. The resulting state of the grid is pictured in Fig. 6. Red edges represent congested grid elements (lines or transformers). At colored nodes, flexibility options have been redispatched to alleviate all congestions, with blue representing negative adjustments to the net in-feed (decreasing generation or increasing load) and green representing positive adjustment (increasing generation or decreasing load). The size of the nodes varies with the absolute adjusted energy.

The corresponding optimal congestion management measures are shown in Table 1. The measures, consisting of up- and down-regulation of flexible plants, add up to zero so that the same production level

Table 1
Available flexibility options and their utilization.

Bus ID	Grid	Type of flexibility option	Costs EUR/MWh	Adjustment MWh	Total costs EUR
9	EHV	Conventional	39.38↓	−158.21	−6229.56
9	EHV	Wind Curtailment	−95.10↓	0.00	0.00
10	EHV	Conventional	69.70↑	0.00	0.00
10	EHV	Wind Curtailment	−89.00↓	0.00	0.00
11	EHV	Conventional	55.76↑	22.42	1250.01
12	EHV	Conventional	46.47↑↓	150.90[a]	7011.82
100	HV_1	Wind Curtailment	−89.00↓	−1.58	140.82
104	MV_1	Demand side flexibility	10.00↓	−10.00[a]	−100.00
105	MV_1	Wind Curtailment	−89.00↓	−5.06	450.31
205	MV_2	Demand side flexibility	120.00↑	1.66	198.97
304	MV_3	Demand side flexibility	85.00↑	4.16	353.44
403	MV_4	Demand side flexibility	10.00↓	−1.86	−18.61
404	MV_4	Demand side flexibility	10.00↓	−2.42	−24.20
405	MV_4	Wind Curtailment	−65.70↓	0.00	0.00
408	MV_4	Wind Curtailment	−65.70↓	0.00	0.00
Total		0.00		3033.00	

[a]Fully utilized. Arrows: possible directions for adjustment.

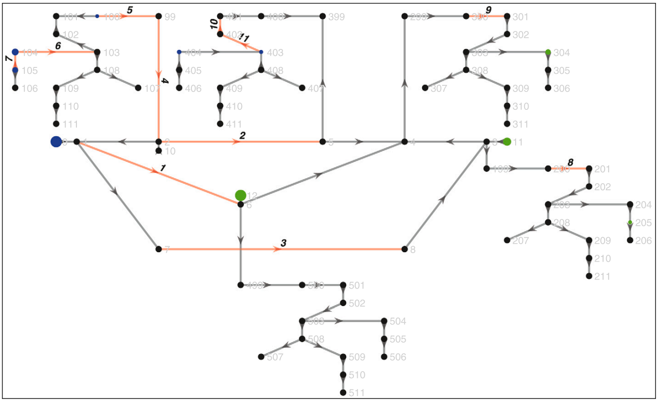


Fig. 6. State of the grid with 11 congestions. Bus IDs in light gray, congested line in red with black numbering. Green: increase of generator power. Blue: decrease of generator power. Size of colored nodes corresponds to size of adjustment.

Table 2
Line data.

Congested lines			Power flows with zonal dispatch			Redispatch results	
ID	Grid	Limit MW	Line Flow MW	Overload MW	Rel. Utilization in %	Line Flow after RD MW	Shadow Price EUR/MW
1	EHV	250	341.41	91.41	137	244.47	0.00
2	EHV	200	218.97	18.97	109	200.00	5.48
3	EHV	500	503.42	3.42	101	444.48	0.00
4	HV_1	130	140.56	10.56	108	123.92	0.00
5	HV_1	150	166.64	16.64	111	150.00	27.27
6	MV_1	25	27.09	2.09	108	12.03	0.00
7	MV_1	25	27.52	2.52	110	22.46	0.00
8	MV_2	25	26.66	1.66	107	25.00	64.24
9	MV_3	25	29.16	4.16	117	25.00	24.91
10	MV_4	25	29.28	4.28	117	25.00	43.21
11	MV_4	25	29.28	4.28	117	25.00	43.21

Note: All columns except *Shadow Price* and *Line Flow after RD* are a result of a power flow calculation. The values in the column *Shadow Price* and *Line Flow after RD* are a result of the optimal power flow.

is maintained. Operators of down-regulated generation plants have to pay back some of the revenues from the original dispatch so that the grid operators generate revenues. Moreover, there are costs for ramping up generation plants or compensating consumers for lower electricity usage. In the optimal scenario, all congestion management measures together induce total costs of 3033 euros that need to be allocated in the congestion management cost-allocation game.

The effect of coordinated congestion management (redispatch) becomes apparent when looking at the distribution grid HV1. Only minor congestion management measures of 1.58 MWh are implemented in that grid region (at Bus ID 100, cf. Table 1) despite overloads of 16.64 MW at congestion ID 5 (cf. Table 2) and 10.56 MW at congestion ID 4. This can be explained by the optimal choice of flexibility options in MV1 as part of coordinated congestion management. The flexibility options at nodes 104 and 105 can eliminate the congestions in the medium-voltage grid (IDs 6 and 7) and alleviate the congestions in the high-voltage grid (IDs 4 and 5). The flexibility option in the high-voltage grid at node 100 has to be applied only marginally because the measures in the medium-voltage grid are not sufficient to completely resolve the congestion of the high-voltage grid. This is a typical case where adequate cost allocation is indispensable, as the measures in one grid can cure the congestions in several grids.

5.2. Cost-allocation methods

The results applying the introduced methods to allocate costs to congested lines[12] are listed in Table 3. In addition to the methods discussed in Appendix, the table shows the costs resulting from coalitions with only one player, the so-called *isolated costs*. These costs can be used as an approximation of the individual costs that arise when each congestion is treated completely independently without considering synergies. However, one has to be careful with this interpretation. The value represents the cost of a single congestion – in a vacuum – and does not consider the competition for the cheapest power plants when several congestions have to be handled simultaneously. As such, it cannot be interpreted as the costs to relieve a congestion independently without engaging in a coalition. It follows that this value does not represent the alternative costs when not engaging in a coalition — in most cases that value would be higher than the isolated costs.

[12] Note that we allocate costs to single congested lines and not on a per operator basis, cf. Section 4.1.2. The results of this *per element* allocation may be aggregated ex post to the grid operator level. Alternatively, per-operator cost allocation could be done directly applying the same principles. However, the results would be structurally different, at least in the case of the Shapley value (cf. footnote 3).

Table 3
Allocated costs per line with different methods (11 congestions).

ID	Grid	Isolated costs	Shapley value	Pro rata	Shadow Price M.	Aggr. allocation
1	EHV	1054.87	488.66	1732.91	0.00	527.43
2	EHV	1508.07	921.12	359.59	466.09	946.18
3	EHV	84.38	11.60	64.79	0.00	42.19
4	HV_1	440.22	145.78	200.14	0.00	220.11
5	HV_1	1264.42	806.01	315.48	2431.79	888.55
6	MV_1	76.32	14.88	39.67	0.00	38.16
7	MV_1	341.98	134.59	47.86	0.00	170.99
8	MV_2	130.22	115.58	31.43	24.81	118.37
9	MV_3	178.72	134.27	78.82	24.13	141.16
10	MV_4	292.96	130.25	81.15	43.09	146.48
11	MV_4	292.96	130.25	81.15	43.09	146.48
Total		5665.12	3033.00	3033.00	3033.00	3386.10

Therefore, the value cannot be used to judge whether joining a coalition would be beneficial for an individual player.

It is noticeable that the costs assigned to individual congestions deviate strongly across allocation methods. While the Shapley value and the aggregated allocation method provide rather similar results, the costs defined by the pro-rata and shadow price methods differ significantly.

The strong deviations of the shadow price method are a result of the method only assigning costs to congestions with non-zero shadow prices. However, congestions 1 and 3 from the transmission grid as well as congestions 4, 6, and 7 in the distribution grid exhibit shadow prices of zero. This phenomenon can be explained by the nature of shadow prices. All of these congestions do not have a direct impact on the total cost of the congestion management, as the elimination of other congestions requires the activation of flexibility options that (more than) eliminate the mentioned congestions as well. However, a fair cost allocation would take these synergies into account, and the corresponding costs would be borne jointly. Additionally, grid operators with these zero shadow price congestions would have no incentive to expand the grid, as the congestion costs would not fall on them. This would only be the case once the grid operator, which instructs and pays for the measures, expands its own grid. Until then, this grid operator pays for congestion management, although other grid operators also benefit from the measures. As a result, a few grid operators bear a very high share of the congestion costs, such as the grid operator of HV1, who has to pay the costs for congestion 5.

The pro-rata method also leads to unfair results since only the absolute amount of the overloads is considered in the cost-allocation method. This is particularly disadvantageous for transmission system operators, who have larger congestions in absolute terms than distribution system operators, as can be observed for congestion 1. The pro-rata method does not consider how expensive the corresponding resolution of the congestions really is. For example, congestion 8 is allocated the lowest cost of only 31.43 EUR, as the overload is the lowest at 1.66 MW. This does not take into account that a very expensive flexibility option on bus 205 (120 EUR/MWh) must be used to resolve this congestion.

The cost components determined by the aggregated allocation method are similar to the cost components of the Shapley value since the calculation takes into account the marginal contributions to the single and grand coalitions, which are also included in the calculation of the Shapley value. However, this method does not meet the efficiency criterion (cf. Eq. (2)) because more costs are allocated to the congested lines (3386.10 EUR) than have actually been incurred by congestion management measures (3033 EUR).

Meanwhile, the Shapley value divides the costs fairly among all congested lines, as explained in detail in Section 3.1. Both the synergies and the costs actually incurred are taken into account via marginal contributions to all possible coalitions. This becomes clear in the assigned cost sharing of congestion 1, which is much lower compared to the pro-rata method, since there are high synergies with required measures for other congestions. Furthermore, a relatively high share of costs is attributed to congestion 8, although the overload is very low. The Shapley value takes into account that the costs are comparatively high to resolve this congestion.

When comparing the various cost-allocation methods, it becomes clear that the Shapley value with its balanced characteristics is a suitable instrument to allocate congestion management costs. However, as already mentioned, the high computational complexity of the Shapley value is a challenge in real-world applications. Accordingly, in the next section, the results of the simplification methods proposed in Section 4.3 are presented.

5.3. Proposed approximation and simplification methods

The discussion of the simplification methods is also based on the example with 11 congestions, although in this case the calculation of the Shapley value using Eq. (1) is still possible without computational problems. Applying the methods to this illustrative example has several advantages. First, the workings of the algorithm can be explained more easily. Additionally, the results can be compared with the "true Shapley value", and thus the accuracy of the methods can be evaluated. However, the algorithms can be applied to cases with a significantly larger number of congestions. A discussion of factors influencing the effectiveness of the algorithms in other settings follows in Section 5.3.4.

5.3.1. Analytical method using non-binding constraints

The approach introduced in Section 4.3.1 leads to a significant reduction in the number of coalitions to be calculated. While the normal calculation method of the Shapley value requires the cost calculation of 2048 coalitions (2^{11}), the calculation of 1720 coalitions can be saved by the analytical method using non-binding constraints. The derived 1720 coalitions consist of 1408 intermediate coalitions and 312 base coalitions. The 328 coalitions to be calculated include 312 starter coalitions and 16 unique value coalitions. In total, the number of coalitions to be calculated is reduced by almost 84 percent, and the method still enables the calculation of the exact Shapley values.

Fig. 7 illustrates the coalitions that can be derived: the base and intermediate coalitions. As detailed in Section 4.3.1, each of these coalitions corresponds to the removal of some or all of the non-binding players from a starter coalition. As these players are non-binding, the cost of the coalition must be the same as the cost of the corresponding starter coalition. Therefore, the costs of these coalitions do not have to be calculated and the total number of calculations is reduced by the number of such coalitions. Our example shows that especially for smaller coalition sizes, a large share of coalitions can be derived without computation. For very small coalition sizes, many of these derivable coalitions are base coalitions, whereas for larger coalition sizes, mostly intermediate coalitions are derivable. It is also clear that the unique value coalitions are particularly prevalent with small coalition sizes. The unique value coalitions are all formed by every possible

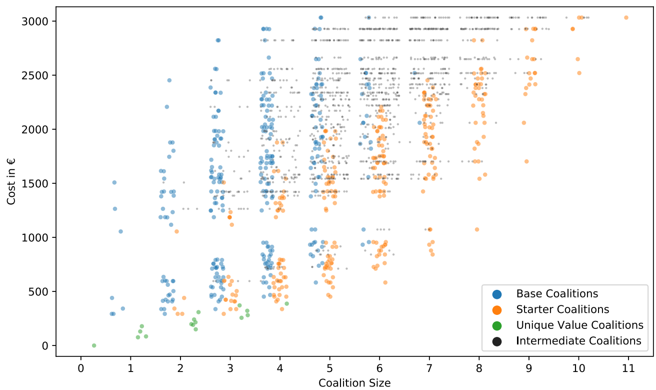


Fig. 7. Needed coalitions for the analytical approach (11 congestions).

Table 4
Comparison of shapley values for congested lines for different approximation methods.

ID	Numerical method: tolerance					Sampling: sample size	
	0%	2.5%	5%	10%	25%	220	40
Relative errors in %:							
1	0	0.02	0.34	0.82	8.48	6.95	15.90
2	0	0.00	−0.23	0.01	−3.17	3.80	8.81
3	0	0.23	24.42	80.23	289.58	15.91	35.92
4	0	−0.03	1.20	4.42	13.99	7.34	17.23
5	0	0.00	−0.14	−0.47	−0.02	2.17	4.97
6	0	−1.37	17.69	60.94	220.48	12.37	28.42
7	0	−0.03	1.30	4.66	8.61	6.99	16.85
8	0	0.00	−2.19	−15.37	−26.22	0.60	1.39
9	0	0.00	−1.99	−11.37	−25.38	1.61	4.06
10	0	0.06	−0.85	0.61	−17.68	7.84	17.59
11	0	0.06	−0.85	0.61	−17.68	7.91	18.11
Absolute errors in EUR:							
1	0	0.08	1.65	4.03	41.45	33.95	77.69
2	0	0.00	−2.10	0.10	−29.20	34.97	81.18
3	0	0.03	2.83	9.31	33.60	1.85	4.17
4	0	−0.04	1.76	6.44	20.39	10.69	25.12
5	0	0.00	−1.10	−3.78	−0.18	17.51	40.03
6	0	−0.20	2.63	9.07	32.81	1.84	4.23
7	0	−0.04	1.76	6.27	11.58	9.41	22.68
8	0	0.00	−2.53	−17.76	−30.30	0.69	1.61
9	0	0.00	−2.68	−15.26	−34.08	2.16	5.46
10	0	0.07	−1.11	0.79	−23.03	10.21	22.92
11	0	0.07	−1.11	0.79	−23.03	10.30	23.59
Necessary coalitions:	1073	1038	914	673	349	1047.4	304.3

Note: Errors for sampling method are standard deviations resulting from 1000 draws of different samples. Necessary coalitions for the sampling method are the mean value of the same 1000 draws.

combination of congestions 3, 6, 8, and 9 (15 combinations). While these congestions are often part of other coalitions, their cheapest solutions on their own do not contribute to the relief of any other congestion.

5.3.2. *Numerical method using value comparison*

The results for the approach introduced in Section 4.3.2 are shown in Table 4. With a tolerance of 0 percent, the calculation of 975 coalitions can be saved by the numerical approach while still calculating the exact Shapley values. While this equates to a relative reduction of almost 48%, the need to calculate small and large coalitions leads to a smaller improvement of the calculation complexity compared to the analytical approach.

Fig. 8 illustrates that the costs of many coalitions with a coalition size of four to eight congestions can be derived. However, in contrast to the analytical approach, no coalitions below a size of four can be derived. This is because in this numerical method, coalitions are derived by comparing the costs of small and large coalition sizes, which sets a lower bound for the size of derivable coalitions. As described in Section 4.3.2, the values of these coalitions can be derived, because each of them is a superset and a subset of a coalition pair, whose values have been calculated and found to be the same (or within a tolerance).[13]

[13] If the two coalitions {a, b} and {a, b, c, d} have the same costs, the coalitions {a, b, c} and {a, b, d} must have the same costs as well. Otherwise,

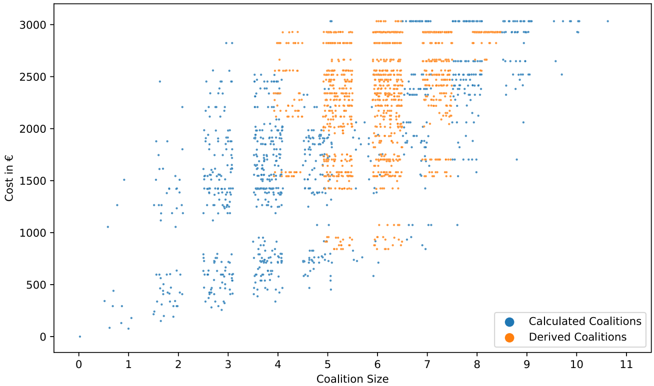


Fig. 8. Needed coalitions for the numerical iterative approach with a tolerance of $t = 5\%$ (11 congestions).

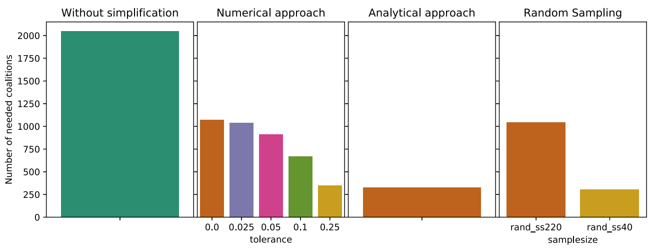


Fig. 9. Number of needed coalitions with different methods, tolerance levels, and sample sizes. rand_ss220 /rand_ss40: Random sampling with sample size 220/40.

The introduction of different tolerance levels can lead to a further reduction in the number of coalitions to be calculated, as displayed in Fig. 9 and Table 4. At a tolerance level of 5%, for example, 914 coalitions have to be calculated, saving over 1100 calculations. However, this reduction of almost 60% comes at the price of imprecise calculations of the Shapley values, which deviate from the true Shapley values, because the values of the derived coalitions are only approximately determined by Eq. (25). The deviations for different tolerance levels are included on the left side of Fig. 10 and in Table 4.

Compared to sampling (i.e., the most commonly applied method to approximate the Shapley value in characteristic function games), the numerical approach still performs quite well. The resulting errors of simple sampling are displayed on the right side of Fig. 10 and in Table 4. For the sampling approach to achieve similar savings to the numerical approach without a tolerance value, a sample size of around 220 is needed. However, while the numerical approach produces exact values at a tolerance of zero, the sampling approach results in errors of up to almost 16%. At a (very high) tolerance level of 25% and the equivalent sample size of 40, errors are still higher. However, in that case sampling performs better than in the numerical approach, where two congestions produce especially high relative errors. In absolute terms, the error is comparable across all congestions, with high relative errors resulting from the small Shapley values assigned to the smaller congestions. Because of the deterministic nature of the error in the numerical approach, it cannot be determined whether high relative errors for some players should be expected generally or if they are outliers.[14]

The error of the approximation methods plays an important role, especially with regard to practical applications like redispatch 2.0. Since large relative errors can occur even with small tolerance values, this could lead to a rejection of the method in practical applications. Especially for small distribution grid operators, even small absolute deviations in relation to the total congestion costs can have a great influence on their operating costs. For this reason, when applying the numerical approach the trade-off between the reduction in the number of coalitions to be calculated and the increasing inaccuracy with larger tolerances needs to be considered.

5.3.3. Impact of the simplification methods on the computational complexity

The results of the previous sections show that the costs for a large number of coalitions do not have to be computed explicitly, but can rather be inferred when using the simplification methods proposed in

costs would have to increase when adding player c to the coalition, but decrease again when additionally adding player d. This would violate the monotonicity assumption, see also Section 4.2.5.

[14] It must be noted that the error from simple sampling is a distribution resulting from 1000 draws, with a mean of zero and a corresponding standard deviation. The latter is used as the error for this method. Meanwhile, the error of the numerical approach is a deterministic error, which is determined by the characteristics of the algorithm. The calculation of the medium coalition values as described in Eq. (25) is one of the influences on this error.

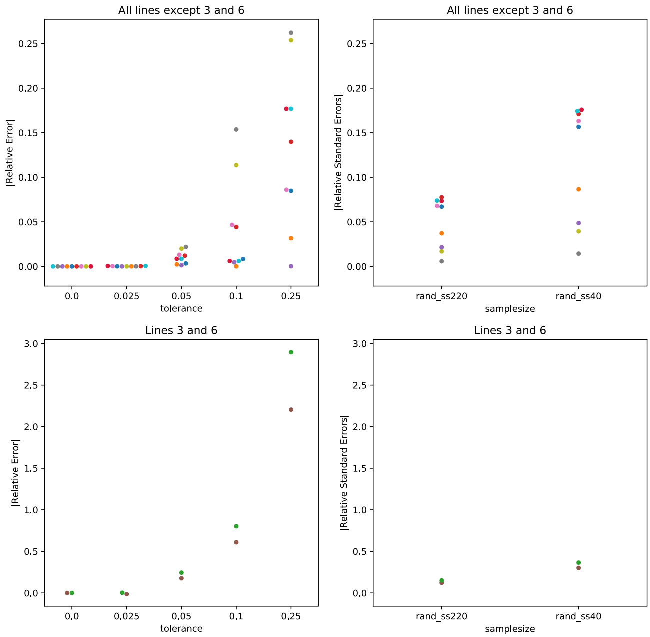


Fig. 10. Errors with different tolerance levels. Left: analytical approach. Right: numerical approach. rand_ss220 /rand_ss40: random sampling with sample sizes 220/40. Lines 3 and 6 are shown separately because their values are much higher and differences between other lines would not be clear if shown in the same graph.

Sections 4.3.1 and 4.3.2. Because explicitly calculating the costs for a coalition involves solving a comparatively resource heavy optimization problem, calculating the costs for all 2^N coalitions (N being the number of congested grid elements) is by far the most resource intensive part of calculating the Shapley value. Consequently, reducing the number of coalitions that have to be calculated by x, approximately reduces the time it takes to calculate the Shapley value by $\frac{x}{2^N}$. Saving half of the coalitions to be calculated reduces the time to calculate the Shapley value by approximately 50%. For the analytical approach, where only 328 out of 2048 coalitions have to be calculated (cf. Section 5.3.1 for details), this implies a calculation time of only 16% of the original calculation time without any simplifications. The numerical approach saves the calculation of 975 coalitions without losing precision, taking only around 47% of the calculation time. Even more time can be saved by sacrificing precision (see Fig. 9).

We have chosen a small example, where each individual optimization takes only the fraction of a second and only comparatively few congested elements are present at the same time. In a real-world application, however, each single optimization is expected to take much longer because of a much more complicated underlying grid model and many more potential redispatching options. Additionally, more congestions are expected to occur at the same time. Consequently, reducing the number of coalitions that have to be calculated will have a direct impact on the feasibility of the Shapley value in a real-world setting.

5.3.4. *Performance of the algorithms in different settings*

The previous sections have shown that the algorithms for reducing the number of coalitions that must be computed are highly effective in the presented case study. However, if these methods are to be implemented in practice to allocate congestion costs, it is necessary to determine whether the observed calculation savings are related to the properties of the specific case study or if they reflect general properties of the approach.

The most important factor determining whether the algorithms perform well is the existence of synergies in the elimination of all congestions. If the synergies are particularly large and several congestions can be eliminated at the same time by certain measures, the algorithms work well because, in this case, non-binding constraints exist, and thus the first algorithm is effective. Of course, this also has a positive effect on the performance of the numerical approach. If there are particularly large congestions, especially in the transmission grid, it is likely that some of the congestions in the lower grid levels will also be solved by the congestion management measures utilized to eliminate the transmission grid congestion. This means that coalitions that include dominant congestions often incur the same costs if the other congestions included in the coalition do not require additional measures.

The stronger the synergies between the measures to eliminate different congestions are, the better the algorithms perform. This is a positive result, as cost-allocation methods are particularly important in situations where measures cannot be assigned directly to the elimination of specific congestions. In the extreme case without any synergies, the corresponding measures could also be directly assigned to the specific congestions, and a complex cost-allocation mechanism would not be necessary.

6. Conclusion

Cooperation among electricity grid operators will become more important in the coming years, as challenges created by the transition to a more sustainable energy system grow. Sharing congestion costs fairly among grid operators will be essential to ensure efficient operation of the system. Congestion cost allocation can be framed as a cooperative game. The central redispatch optimization corresponds to the characteristic function, which determines the costs. The congested elements, which are to be resolved, are then the players forming a coalition. This enables the use of the Shapley value, a mechanism well known from cooperative game theory. It fulfills the central requirement of fairness but requires the calculation of costs for all hypothetical coalitions of congested elements. This raises concerns regarding its feasibility because of its high computational complexity.

To help address this challenge, we have developed two methods that are well suited to the use case of congestion management, which dramatically reduce the amount of resources needed to calculate the Shapley value. The methods work well because congestion management exhibits synergies. In many cases, actions taken to relieve a subset of congestions alleviate other congestions as well. Translated into the game theoretic context, adding more players (congestions) to a given coalition (congestions to be relieved) often does not change the value of the coalition (the costs associated with relieving the congestions). We expect the approaches to yield good results in other domains as well if this key characteristic also holds in these domains.

The first of the two approaches, the analytical approach, reduces the amount of required calculations – and therefore the calculation time – to approximately one-sixth of the original problem in our case. However, it requires further knowledge about each optimization underlying the calculation of the coalition values. Meanwhile, the second approach, the numerical approach, can be applied widely. It relies only on the fact that costs cannot decrease when additional congestions need to be resolved. In our case study, the numerical approach can save almost half of the calculation time by almost halving the number of calculations performed.

While both methods yield precise results, the numerical approach can be tweaked further by introducing tolerances, further reducing the number of coalitions that need to be calculated. This introduces a trade-off, where the potential savings must be evaluated against the inaccuracies that are introduced by the tolerances.

Our approaches compare favorably to standard approximation methods, such as sampling. However, while our method can save a large amount of resources without sacrificing accuracy, our results suggest that when introducing tolerances, errors from the sampling methods increase more steadily with decreasing sample size, while errors in our numerical approach exhibit larger jumps for some of the congestions.

Overall, we believe that the proposed methods make it possible to apply the Shapley value to allocate congestion costs in a real-world setting. Further research should aim to verify their efficacy in a larger case study.

CRediT authorship contribution statement

Simon Voswinkel: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Jonas Höckner:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Abuzar Khalid:** Writing – review & editing. **Christoph Weber:** Supervision, Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The basis for this contribution was developed under contract with TransnetBW GmbH, which awarded the contract with funds from a grant by the German Federal Ministry for Economic Affairs and Energy as part of the SINTEG project C/sells.

SINTEG
SMART ENERGY SHOWCASES
Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

Appendix. Cost allocation methods

The following section provides some more detail on relevant cost allocation methods for congestion management cost as applied in Section 5.2.

The methods that are discussed in literature and in some cases implemented are diverse and vary greatly in their complexity. In addition to the question of how the costs are allocated, the approaches also differ in terms of whom the costs are allocated to. For example, costs can be shared between consumers, suppliers, or both, but the most common approach is to share the costs between final consumers. However, congestion costs can also be allocated to the congested lines or the underlying transactions causing the congestion. Furthermore, a combination of different methods at different levels is also conceivable. For example, congestion costs can first be allocated to grid operators according to a certain method, who then pass the costs on to the consumers in their grid area as uplift costs, as is the case in Germany with the national redispatch of the transmission system operators. We focus on methods of allocating costs to congested elements or the corresponding grid operators as described in the following sections.

Pro rata methods

The simplest method for allocating costs is probably the pro-rata method. It is based on distributing costs in proportion to the size of the congestion as so-called uplift costs among different agents [19,59,60]. The method is also known and applied as postage stamp cost allocation method, which allocates for example transmission costs according to load ratio shares of participants in the market [61]. Typically, the cost share is calculated on the basis of the ratio of the respective connected capacity or the amount of energy withdrawn in relation to the entire system.

$$c_i = C(N) \cdot \frac{P_i^{\text{connected/withdrawn}}}{\sum_i P_i^{\text{connected/withdrawn}}} \tag{A.1}$$

However, the pro rata method has some significant weaknesses as explained in [19,62]. [62] state that though the method is simple to understand and implement, it fails to provide efficient signals because it does not consider the location of consumers. The uplift costs do not provide information about the location of congested lines and hence do not give right signals for investments in the transmission grid. The lack of economic efficiency may also lead to cross subsidization. Despite these disadvantages, the pro rata method can be acceptable if the congestions are infrequent or the congestion costs are small [19].

If congestion management costs are to be distributed among the network operators using the pro rata method, the ratio of the congestion of one network operator to the total congestion of all network operators in the system can be used as a cost sharing key. This procedure corresponds to the so-called congestion account ("Überlastkonto"), which has been in use in Germany to allocate national redispatch costs between the four transmission system operators.

Congestion management costs can also be divided between congested lines in a similar way. [60] introduce the congestion allocation index (CAI), which allocates line congestion costs to market participants based on the load flow that is caused by their injection. To apply this method, the total congestion costs must be allocated to the congested lines. [60] achieve that by simply allocating the costs pro rata based on relative amounts of congestion as indicated in the formula above. In our application, we assign the congestion cost to the congested lines based on the overload observed prior to the congestion management (cf. Table 2):

$$P_j^{F,\text{overload}} = \max(P_j^{F,0} - P_j^{F,\max}, 0) \tag{A.2}$$

For a line j this is the positive difference of the power flow $P_j^{F,0}$ obtained from a power flow computation and the line capacity $P_j^{F,\max}$.

$$c_j = c(N) \cdot \frac{P_j^{F,\text{overload}}}{\sum_{j'} P_{j'}^{F,\text{overload}}} \tag{A.3}$$

Shadow price method

[19] propose an allocation method of congestion management cost, which consists mainly of two steps. In step one, the total congestion cost is allocated to congested lines based on a so-called constraint allocation factor. In step two, the costs for each congested line is allocated to individual transactions using a load allocation factor. As our focus is on the allocation of congestion costs to congested line, only step one of the shadow price method is relevant here.

The constraint allocation factor α_j, that allocates overall costs C_{total} to constraint j, is based on shadow prices μ_j of the underlying optimization problem. It is given as

$$c_j = c(N) \cdot \alpha_j \text{ with } \alpha_j = \frac{\mu_j P_j^{F,\text{overload}}}{\sum_{j'} \mu_{j'} P_{j'}^{F,\text{overload}}}, \tag{A.4}$$

where $P_j^{F,\text{overload}}$ is the overload on a congested line j as defined in Eq. (A.2) and $\sum_j \alpha_j = 1$. The shadow prices μ_j indicate the sensitivity of total costs to changes in congested line capacity, i.e., how much the total costs change when this restriction is relaxed. These prices are non-zero in case of congested lines and zero otherwise.

According to [63] the shadow price method is only suitable for small changes in line flows of congested branches because the relation between total congestion cost and the overload of the congested line is highly nonlinear. For this reason, major changes in overloads are expected to result in significant errors. For this reason [64] propose a modification of this approach which is described in the next section.

Aggregated allocation method

[64] propose the aggregated allocation method which adopts the first step of the shadow price method of [19] used to allocate total congestion management costs to congested lines. [64] calculate the marginal and incremental cost of relieving each congested line, and then assign the average of these two costs to the line as the aggregated cost. This method is also known as incremental method and was used, for example, by [20] to allocate congestion costs. It is based on the fact that the order of relieving congestions has an impact on the cost for the lines due to the nonlinear nature of the cost function in constrained power systems [65]. Therefore the aggregated allocation method considers two components by calculating the cost for each line j based on the marginal and incremental cost. The marginal costs C_j^{mg} of congested line j is the congestion management cost that arise when only the congested line j is alleviated, i.e., in the notation of the Shapley game c({j}).[15] The incremental costs, on the other hand, are the additional costs that arise from the situation in which all congestions J (corresponding to the set of players S in the Shapley game) except j have already been eliminated and j is the last congestion that is eliminated. The incremental costs C_j^{in} of constraint j is

$$C_j^{in} = C(N) - c(N \setminus \{j\}) \tag{A.5}$$

The resulting aggregated congestion management costs that are allocated to line j are therefore

$$C_j = \frac{1}{2} \cdot (C_j^{in} + C_j^{mg}) \tag{A.6}$$

According to [65], this method can lead to improvements compared to the shadow price approach of [19], but it still leads to mismatches. [63] come to the same conclusion and state that the aggregated allocation method still cannot overcome the problem associated with the sequence of the branch congestion removal. They conclude that allocation methods based on shadow prices or marginal prices are unable to produce a fair congestion cost allocation.

[15] In our application we call these costs *isolated costs*.

References

[1] Weibelzahl M. Nodal, zonal, or uniform electricity pricing: How to deal with network congestion. Front Energy 2017;11(2):210–32. http://dx.doi.org/10.1007/s11708-017-0460-z.

[2] Schweppe FC, Caramanis MC, Tabors RD, Bohn RE. Spot pricing of electricity. Boston, MA: Springer US; 1988, http://dx.doi.org/10.1007/978-1-4613-1683-1.

[3] Hogan WW. Contract networks for electric power transmission. J Regul Econ 1992;4(3):211–42. http://dx.doi.org/10.1007/BF00133621.

[4] Bjørndal M, Jørnsten K. Zonal pricing in a deregulated electricity market. Energy J 2001;22(1):51–73, URL www.jstor.org/stable/41322907.

[5] de Vries LJ, Hakvoort RA. An economic assessment of congestion management methods for electricity transmission networks. J Netw Ind 2002;os-3(4):425–66. http://dx.doi.org/10.1177/178359170200300403.

[6] Höckner J, Voswinkel S, Weber C. Market distortions in flexibility markets caused by renewable subsidies – The case for side payments. Energy Policy 2020;137:111135. http://dx.doi.org/10.1016/j.enpol.2019.111135.

[7] Jin X, Wu Q, Jia H. Local flexibility markets: Literature review on concepts, models and clearing methods. Appl Energy 2020;261:114387. http://dx.doi.org/10.1016/j.apenergy.2019.114387.

[8] Andreadou N, Flammini MG, Fulli G, Masera M, Prettico G, Vitiello S. Distribution system operators observatory 2018: Overview of the electricity distribution system in Europe. EUR, Scientific and technical research series, Vol. 29615, Luxembourg: Publications Office of the European Union; 2019.

[9] Bjørndal M, Jørnsten K. Benefits from coordinating congestion management: The Nordic power market. Energy Policy 2007;35(3):1978–91. http://dx.doi.org/10.1016/j.enpol.2006.06.014.

[10] Oggioni G, Smeers Y. Market failures of market coupling and counter-trading in europe: An illustrative model based discussion. Energy Econ 2013;35:74–87. http://dx.doi.org/10.1016/j.eneco.2011.11.018.

[11] BDEW. BDEW-Branchenlösung redispatch 2.0: Datenaustausch-, bilanzierungs- und abrechnungsprozesse. 2020, URL https://www.bdew.de/media/documents/Awh_2020-05-RD_2.0_Branchenl%C3%B6sung_Kerndokument.pdf.

[12] Chalkiadakis G, Elkind E, Wooldridge M. Computational aspects of cooperative game theory. Synth Lect Artif Intell Mach Learn 2011;5(6):1–168. http://dx.doi.org/10.2200/S00355ED1V01Y201107AIM016.

[13] Maleki S. Addressing the computational issues of the Shapley value with applications in the smart grid. 2015.

[14] Shapley LS. 17. A Value for n-person games. In: Kuhn HW, Tucker AW, editors. Contributions to the theory of games (am-28), Vol. II. Princeton: Princeton University Press; 1953, p. 307–18. http://dx.doi.org/10.1515/9781400881970-018.

[15] Shubik M. Incentives, decentralized control, the assignment of joint costs and internal pricing. Manage Sci 1962;8(3):325–43, URL http://www.jstor.org/stable/2627389.

[16] Roth AE, Verrecchia RE. The Shapley value as applied to cost allocation: A reinterpretation. J Account Res 1979;17(1):295. http://dx.doi.org/10.2307/2490320.

[17] Fiestras-Janeiro MG, García-Jurado I, Mosquera MA. Cooperative games and cost allocation problems. TOP 2011;19(1):1–22. http://dx.doi.org/10.1007/s11750-011-0200-1.

[18] van Campen T, Hamers H, Husslage B, Lindelauf R. A new approximation method for the Shapley value applied to the WTC 9/11 terrorist attack. Soc Netw Anal Min 2018;8(1):105. http://dx.doi.org/10.1007/s13278-017-0480-z.

[19] Singh H, Hao S, Papalexopoulos A. Transmission congestion management in competitive electricity markets. IEEE Trans Power Syst 1998;13(2):672–80. http://dx.doi.org/10.1109/59.667399.

[20] Xiao H, Li W. Allocation of congestion cost in a pool based model using Shapley value. In: Asia-pacific power and energy engineering conference, 2009: Parallel als printausg. erschienen / tagungsdaten auch angegeben als: March 28 - 31, 2009. Piscataway, NJ: IEEE; 2009, p. 1–5. http://dx.doi.org/10.1109/APPEEC.2009.4918127.
[21] Lima D, Contreras J, Padilha-Feltrin A. A cooperative game theory analysis for transmission loss allocation. Electr Power Syst Res 2008;78(2):264–75. http://dx.doi.org/10.1016/j.epsr.2007.02.008.
[22] Molina YP, Prada RB, Saavedra OR. Complex losses allocation to generators and loads based on circuit theory and Aumann-Shapley method. IEE Proc, Gener Transm Distrib 2010;25(4):1928–36. http://dx.doi.org/10.1109/TPWRS.2010.2044425.
[23] Azad-Farsani E, Agah S, Askarian-Abyaneh H, Abedi M, Hosseinian S. Stochastic LMP (locational marginal price) calculation method in distribution systems to minimize loss and emission based on Shapley value and two-point estimate method. Energy 2016;107:396–408. http://dx.doi.org/10.1016/j.energy.2016.04.036, URL http://www.sciencedirect.com/science/article/pii/S036054421630442X.
[24] Sharma S, Abhyankar A. Loss allocation for weakly meshed distribution system using analytical formulation of Shapley value. IEEE Trans Power Syst 2016;1. http://dx.doi.org/10.1109/TPWRS.2016.2571980.
[25] Pourahmadi F, Dehghanian P. A game-theoretic loss allocation approach in power distribution systems with high penetration of distributed generations. Mathematics 2018;6(9):158. http://dx.doi.org/10.3390/math6090158.
[26] Amaris H, Molina YP, Alonso M, Luyo JE. Loss allocation in distribution networks based on Aumann–Shapley. IEEE Trans Power Syst 2018;33(6):6655–66. http://dx.doi.org/10.1109/TPWRS.2018.2844740.
[27] Zolezzi JM, Rudnick H. Transmission cost allocation by cooperative games and coalition formation. IEEE Trans Power Syst 2002;17(4):1008–15. http://dx.doi.org/10.1109/TPWRS.2002.804941.
[28] Kattuman P, Green R, Bialek J. Allocating electricity transmission costs through tracing: a game-theoretic rationale. Oper Res Lett 2004;32(2):114–20. http://dx.doi.org/10.1016/S0167-6377(03)00095-6.
[29] Junqueira M, da Costa LC, Barroso LA, Oliveira GC, Thome LM, Pereira MV. An Aumann-Shapley approach to allocate transmission service cost among network users in electricity markets. IEE Proc, Gener Transm Distrib 2007;22(4):1532–46. http://dx.doi.org/10.1109/TPWRS.2007.907133.
[30] Molina YP, Saavedra OR, Amaris H. Transmission network cost allocation based on circuit theory and the Aumann-Shapley method. IEEE Trans Power Syst 2013;28(4):4568–77. http://dx.doi.org/10.1109/TPWRS.2013.2278296.
[31] Zhang W, Wang X, Qi T, Wu X. Transmission cost allocation based on data envelopment analysis and cooperative game method. Electr Power Compon Syst 2018;46(2):208–17. http://dx.doi.org/10.1080/15325008.2018.1444113.
[32] Contreras J, Wu FF. Coalition formation in transmission expansion planning. IEEE Trans Power Syst 1999;14(3):1144–52. http://dx.doi.org/10.1109/59.780946.
[33] Contreras J, Gross G, Arroyo JM, Muñoz JI. An incentive-based mechanism for transmission asset investment. Decis Support Syst 2009;47(1):22–31. http://dx.doi.org/10.1016/j.dss.2008.12.005.
[34] Hasan KN, Saha TK, Chattopadhyay D, Eghbal M. Benefit-based expansion cost allocation for large scale remote renewable power integration into the Australian grid. Appl Energy 2014;113:836–47. http://dx.doi.org/10.1016/j.apenergy.2013.08.031.
[35] Banez-Chicharro F, Olmos L, Ramos A, Latorre JM. Beneficiaries of transmission expansion projects of an expansion plan: An Aumann-Shapley approach. Appl Energy 2017;195:382–401. http://dx.doi.org/10.1016/j.apenergy.2017.03.061.
[36] Banez-Chicharro F, Olmos L, Ramos A, Latorre JM. Estimating the benefits of transmission expansion projects: An Aumann-Shapley approach. Energy 2017;118:1044–54. http://dx.doi.org/10.1016/j.energy.2016.10.135.
[37] Kristiansen M, noz FDM, Oren S, Korpås M. A mechanism for allocating benefits and costs from transmission interconnections under cooperation: A case study of the north sea offshore grid. Energy J 2018;39(6). http://dx.doi.org/10.5547/01956574.39.6.mkri.
[38] Babusiaux D, Pierru A. Modelling and allocation of CO2 emissions in a multi-product industry: The case of oil refining. Appl Energy 2007;84(7–8):828–41. http://dx.doi.org/10.1016/j.apenergy.2007.01.013.
[39] O'Brien G, El Gamal A, Rajagopal R. Shapley value estimation for compensation of participants in demand response programs. IEEE Trans Smart Grid 2015;6(6):2837–44. http://dx.doi.org/10.1109/TSG.2015.2402194.
[40] Lo Prete C, Hobbs BF. A cooperative game theoretic analysis of incentives for microgrids in regulated electricity markets. Appl Energy 2016;169:524–41. http://dx.doi.org/10.1016/j.apenergy.2016.01.099.
[41] Rahmani-Dabbagh S, Sheikh-El-Eslami MK. A profit sharing scheme for distributed energy resources integrated into a virtual power plant. Appl Energy 2016;184:313–28. http://dx.doi.org/10.1016/j.apenergy.2016.10.022.
[42] Moretti S, Patrone F. Transversality of the Shapley value. TOP 2008;16(1):1.
[43] Algaba E, Fragnelli V, Sánchez-Soriano J. Handbook of the Shapley value. Series in operations research, Boca Raton, FL: CRC Press; 2020.
[44] Conitzer V, Sandholm T. Computing Shapley values, manipulating value division schemes, and checking core membership in multi-issue domains. 2004, p. 219–25.
[45] Ieong S, Shoham Y. Marginal contribution nets. In: Riedl J, Kearns M, Reiter M, editors. EC'05. New York, New York, USA: ACM Press; 2005, p. 193–202. http://dx.doi.org/10.1145/1064009.1064030.
[46] Deng X, Papadimitriou CH. On the complexity of cooperative solution concepts. Math Oper Res 1994;19(2):257–66, URL http://www.jstor.org/stable/3690220.
[47] Mann I, Shapley LS. Values of large games, IV: Evaluating the electoral college by Montecarlo techniques. 1960, URL https://www.rand.org/pubs/research_memoranda/RM2651.html.
[48] Owen G. Multilinear extensions of games. Manage Sci 1972;18(5-part-2):64–79. http://dx.doi.org/10.1287/mnsc.18.5.64.
[49] Fatima SS, Wooldridge M, Jennings NR. A linear approximation method for the Shapley value. Artificial Intelligence 2008;172(14):1673–99. http://dx.doi.org/10.1016/j.artint.2008.05.003.
[50] Castro J, Gómez D, Tejada J. Polynomial calculation of the Shapley value based on sampling. Comput Oper Res 2009;36(5):1726–30. http://dx.doi.org/10.1016/j.cor.2008.04.004.
[51] Maleki S, Tran-Thanh L, Hines G, Rahwan T, Rogers A. Bounding the estimation error of sampling-based Shapley value approximation with/without stratifying. 2013, URL http://arxiv.org/abs/1306.4265.
[52] Castro J, Gómez D, Molina E, Tejada J. Improving polynomial estimation of the Shapley value by stratified random sampling with optimum allocation. Comput Oper Res 2017;82:180–8. http://dx.doi.org/10.1016/j.cor.2017.01.019.
[53] Illés F, Kerényi P. Estimation of the Shapley value by ergodic sampling. 2019, URL http://arxiv.org/pdf/1906.05224v1.
[54] Kunz F, Zerrahn A. Benefits of coordinating congestion management in electricity transmission networks: Theory and application to Germany. Util Policy 2015;37:34–45. http://dx.doi.org/10.1016/j.jup.2015.09.009, URL https://www.sciencedirect.com/science/article/pii/S0957178715300254.
[55] Zimmerman RD, Murillo-Sanchez CE, Thomas RJ. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. IEE Proc, Gener Transm Distrib 2011;26(1):12–9. http://dx.doi.org/10.1109/TPWRS.2010.2051168.
[56] Zimmerman RD, Murillo-Sánchez CE. Matpower user's manual. 2019, http://dx.doi.org/10.5281/zenodo.3251118.
[57] Sioshansi R, Conejo AJ. Optimization in Engineering. Berlin, Heidelberg: Springer International Publishing; 2017, http://dx.doi.org/10.1007/978-3-319-56769-3.
[58] CIGRE. Benchmark systems for network integration of renewable and distributed energy resources. Technical brochure, task force c6.04.02, 2014.
[59] Lo KL, Lozano CA, Gers O. JM. Game theory application for determining wheeling charges. In: Lai LL, editor. International conference on electric utility deregulation and restructuring and power technologies, London, UK 4-7 April 2000. Piscataway, NJ: IEEE; 2000, p. 308–13. http://dx.doi.org/10.1109/DRPT.2000.855682.
[60] Dehghan S, Sedighi AR, Moradi MH, Mirjalili HR. The new non-discriminatory strategy for cost allocation in restructured system. In: International conference on power system technology (POWERCON), 2010: Kongr.-Thema: Technological innovations making power grid smarter / parallel als druckausg. erschienen. Piscataway, NJ: IEEE; 2010, p. 1–6. http://dx.doi.org/10.1109/POWERCON.2010.5666583.
[61] Majidi Q. M, Ghazizadeh MS, Afsharnia S. A novel approach to allocate transmission embedded cost based on MW-mile method under deregulated environment. In: IEEE Canada electric power conference, 2008: Kongr.-Thema: Energy innovation. Piscataway, NJ: IEEE; 2008, p. 1–6. http://dx.doi.org/10.1109/EPC.2008.4763344.
[62] Murali M, Sri Divya P, Sailaja Kumari M, Sydulu M. Aumann Shapley method for congestion cost allocation in multilateral transactions framework of restructured power market. Intell Autom Soft Comput 2015;21(1):107–21. http://dx.doi.org/10.1080/10798587.2014.940752.
[63] Wu ZQ, Wang YN, Qing H-S, Ou Yang YX. Continuous integration congestion cost allocation based on sensitivity. IEE Proc, Gener Transm Distrib 2004;151(4):421. http://dx.doi.org/10.1049/ip-gtd:20040711.
[64] Baran ME, Banunarayanan V, Garren KE. Equitable allocation of congestion relief cost to transactions. IEEE Trans Power Syst 2000;15(2):579–85. http://dx.doi.org/10.1109/59.867144.
[65] Yang H, Shi D, Duan X. Congestion cost allocation method based on aumann-Shapley value in bilateral model. In: 2003 IEEE power engineering society general meeting. Piscataway, NJ: IEEE Operations Center; 2003, p. 1002–6. http://dx.doi.org/10.1109/PES.2003.1270448.

Chapter 6

Simplifying the computation of Shapley values for allocating congestion costs in large power grid models

by Simon Voswinkel
submitted to Applied Energy in May of 2023

Simplifying the computation of Shapley values for allocating congestion costs in large power grid models by Simon Voswinkel

Abstract

The ongoing transformation of the energy system challenges the electricity grids throughout Europe, increasing the costs of congestion management. Because the interconnected electricity grids are operated by many different grid operators, fairly allocating these costs is essential. A long-established method for fairly allocating costs is the Shapley value. Based on Voswinkel et al. (2022) [1], this contribution investigates the applicability of a simplification approach for calculating the Shapley value to a realistic grid model, covering the German extra high voltage grid. This *analytical method using non-binding constraints* is applied to a whole year of hourly grid load cases and the factors that govern the efficiency of the algorithm are analyzed. Results show that the effectiveness of the algorithm increases substantially when more overloaded elements are present in a given grid load case, with more than 99% of computational effort being saved in grid load cases with many overloaded elements. Regression analyses indicate that the effectiveness is governed by the extent to which redispatch measures can alleviate multiple overloads at once – the situations where fairly sharing the costs is most important. This contribution thus demonstrates that previously evoked computational difficulties may be overcome and that the operational implementation of the Shapley value should be seriously considered.

Keywords: Congestion Management, Redispatch, Shapley Value, Cost Sharing, Power Grids, Game Theory

Simon Voswinkel

(Corresponding Author)

House of Energy Markets and Finance

University of Duisburg-Essen, Germany

Universitätsstr. 2, 45141 Essen

+49-(0)201 / 183-3127

simon.voswinkel@uni-due.de

www.hemf.net

Contents

List of Figures

List of Tables

1 Introduction

The ongoing transformation of the energy system challenges the electricity grids throughout Europe. Electricity markets in Europe are operated using zonal pricing, which requires redispatch to ensure safe system operation [2, 3]. Since the electricity grid is highly meshed, redispatch measures can affect multiple congestions at once. Coupled with the fact that the interconnected electricity grids are operated by different system operators, the allocation of congestion management costs is an important task.

With almost 900 different grid operators responsible for the electricity grid [4], each levying individual grid fees from its customers, Germany provides a salient example of the need to *fairly* allocate these costs. Its electricity grid, centrally located within Europe and well interconnected, is increasingly congested: Costs for congestion management reached 1.4 billion Euros in 2020 [5] and are projected to have reached 2.3 billion Euros in 2021 [6]. At the same time, the number of congestions in distribution grids is increasing. In 2021, 27.3 % of curtailment measures for renewable energy sources could be attributed to distribution grid-level causes [6].

The Shapley value, a concept from game theory, has long been established as a fair mechanism to allocate utility or cost arising from a coalition among its participants [7, 8, 9]. Recently, [1] demonstrated the use of the Shapley value to allocate congestion costs to the congested grid elements, and by extension to the grid operators responsible for them. Congestion management is commonly based on an optimal power flow (OPF) calculation in its DC formulation, a linear optimization problem that calculates the cost-optimal redispatch of units subject to grid constraints. To compute the Shapley value in the context of congestion management, [1] frame the congestion management problem in a game theoretic context, with the OPF as the cost function and the congested grid elements as players. To calculate the Shapley value, the costs for each possible coalition of players must be known. However, because the cost function is an optimization problem, calculating these costs is in itself computationally expensive. [1] introduce two methods for reducing the number of necessary calculations when applying the Shapley value to linear optimization problems. The more promising of the two methods, called the *analytical method using non-binding constraints* (analytical method), succeeded in reducing the computational burden to less than 20 % of that of the original problem – when applying it to small benchmark grid topologies.

This paper contributes to the existing literature by extending the work in [1] beyond a small benchmark grid for a single grid load case. Hence, this contribution investigates the effectiveness of the simplification method in a more realistic grid topology with 8784 grid load cases covering a whole year. The grid topology and the associated time series are taken from the SimBench project [10], which provides realistic benchmark grid topologies for the German transmission and distribution grids. The paper further investigates the factors determining the performance of the algorithm using regression analyses. The results show that the computational savings achieved with the simplification method increase with the number of congestions. For cases with more than 20 congestions, the computation time is reduced from several weeks to minutes. The regression analyses show that the efficiency of the algorithm depends in large part on the synergies present in the redispatch problem. These are also the situations where fair cost allocation is most important. As such, this paper demonstrates that the Shapley value in the context of congestion management can be used for a realistic grid topology with reasonable computational resources. Therefore, it advances the possibility of implementation considerably.

Section 2 starts by introducing the Shapley value and the algorithm developed in [1] that will be applied in this paper. Further, the methodology for the case study is described. In Section 3, the results are described by first providing an overview and then analyzing the factors that determine the effectiveness of the algorithm using regression analyses. Additionally, the generalizability of the results is discussed. Finally, Section 4 concludes the paper.

2 Methods

In this section, congestion management (Section 2.1) as well as the Shapley value and its application to congestion management (Section 2.2), the simplification methods developed in [1] (Section 2.3), and the methods for the case study (Section 2.4) are described.

2.1 Congestion management

In zonal electricity markets, such as those prevailing in Europe, the initial market clearing leads to a cost-minimal dispatch. As this dispatch may lead to overloads in the power grid, the grid operators perform redispatch operations to cope with the so-called *congestions*.

A stylized example of the dispatch problem and the redispatch problem is shown in Figure 1 (see [1] for the full technical problem formulation). Both the original dispatch and redispatch may be formulated as cost minimization problems. Thereby, the total generation costs are minimized subject to the constraints of the power plants (e.g., maximum capacities) and the load serve constraint, which states that the total generation must equal the total demand. The redispatch problem adds grid constraints, making it an OPF problem.

Because the grid constraints are not present in the original dispatch, this may lead to overloads on grid elements. After redispatch, some elements may still be congested (i.e. fully utilized and constraining the dispatch), but no element may be overloaded.

In Figure 1, there are generators on nodes 1 and 4, which must serve a total load of 100 MW distributed over nodes 2, 3, and 4. Because the generator on node 1 has lower marginal costs than the generator on node 4 and because its capacity is sufficient, it serves the total load of 100 MW. However, this generation pattern would result in overloads of the grid elements {a, b,

Used Symbols

$Pmax$:	Capacity of generator
$Fmax$:	Thermal capacity of line
N:	Total number of generators
c:	Generator costs
Pg_i:	Generation by generator i
Pd:	Demand
F:	Power flow over line

Dispatch problem

$$\min \sum_{i}^{N} c(Pg_i)$$

s.t. power plant constraints

load serve constraint

Redispatch problem

$$\min \sum_{i}^{N} c(Pg_i)$$

s.t. power plant constraints

load serve constraint

grid constraints

Stylized grid **Dispatch outcome** **Redispatch outcome**

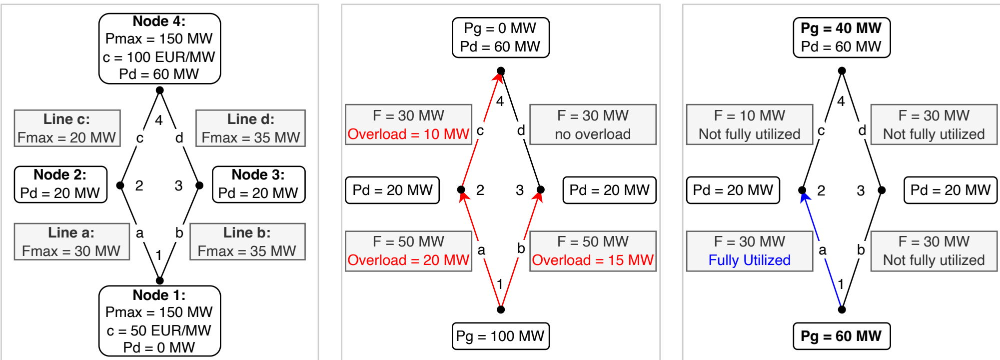


Figure 1: Stylized example of the original dispatch and redispatch. Overloaded elements in red, fully utilized elements in blue. Impedances are assumed to be equal for all lines.

c} (shown in red).[1] To prevent these potential overloads, congestion management is performed, and the output from the generator on node 1 is reduced and replaced by generation on node 4. This redispatch results in extra costs, as the generator on node 4 is more expensive to operate than the generator on node 1 whose output it (partially) replaces. These extra costs are the congestion costs that will be allocated to the original overloaded elements that caused the redispatch and are therefore responsible for the extra costs.

After redispatch, only grid element a is still considered to be congested because it is fully utilized. Grid elements b and c are not fully utilized after redispatch and therefore are not congested. In later sections, they will be referred to as being *non-binding*. Part of the solution to the optimization problem is *shadow prices* for every constraint. The shadow price of a constraint corresponds to the change in the objective function value if the constraint were relaxed by one unit [11]. In the case of non-binding constraints, the shadow price is zero, a fact that will be used in Section 2.3 to develop the simplification method.

[1] The capacities $Fmax$ for each line are chosen for illustrative purposes and do not represent realistic line capacities.

2.2 The Shapley value and its application to congestion management

This section summarizes the concepts explored in [1] to lay the foundations for the main contribution of this paper: the application to a realistic grid and the identification of the success factors of the algorithm.

The Shapley value can be used to allocate the utility or cost of a coalition of cooperating players to the players forming the coalition [7, 8, 9]. Based on [12], the Shapley value Φ_i of player i is defined as

$$\Phi_i(c) = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}(c(S \cup i) - c(S)), \tag{1}$$

where N is the set of all players, $c(N)$ is the cost of coalition N, and S is a subset of the set of all players. The Shapley value can be understood as the average marginal contribution of each player i to each possible coalition, weighted by the frequency of each coalition in the possible permutations arising from the grand coalition.

To use the game-theoretic concept of the Shapley value for allocating congestion costs to the overloaded grid elements, the congestion management problem must be mapped to a corresponding game-theoretic problem. Each time step is treated as its own separate game, with the overloaded grid elements as the players forming the grand coalition in the respective time step. The total costs to be allocated are calculated by performing an OPF calculation for the time step (cf. Section 2.1). A subcoalition of the game-theoretic problem corresponds to a subset of the overloaded elements. The costs for this subset are calculated by setting the capacities of all overloaded elements that are *not* in the coalition to their actual line flows, thereby relaxing the associated constraints, and performing a second OPF calculation. Because the constraints for the elements not in the coalition are relaxed, the resulting costs represent the costs of only performing congestion management for the elements participating in the coalition.

Equation (1) sums over all subsets (subcoalitions) that can be formed from the grand coalition N. As such, the costs $c(S)$ of all coalitions must be known. As described above, the costs are calculated by performing an OPF, which is a linear optimization problem. This makes calculating the costs computationally expensive. Because the number of required cost values grows exponentially with $2^{|N|}$, the extensive computation of all cost figures quickly becomes unfeasible. The next section describes a method for reducing the number of necessary calculations.[2]

2.3 Method to reduce the number of coalitions to be calculated

As described in the previous section, the main driver of the computational complexity of the Shapley value in the context of congestion management is the expensiveness of calculating the

[2]The number of cost values that are required for the Shapley equation (1) will be called the number of *required* calculations or coalitions. The number of cost values that still have to be calculated after using the simplification algorithm will be called the number of *necessary* calculations or *calculated* coalitions.

costs for each subcoalition of congested grid elements. However, with the methods developed in [1], not every coalition has to be calculated. While two methods are proposed in [1], this paper focuses on the more promising method called the *analytical method using non-binding constraints* (analytical method). This method is chosen because it yields exact results and also has a higher savings potential. The analytical method is described in the remainder of this section.

The solution to a linear optimization problem (e.g., in an OPF reflecting the redispatch) contains primal and dual variables. The dual variables are called shadow prices, and each one corresponds to a constraint of the optimization problem. The variables reflect the marginal change in the value of the objective function for a marginal relaxation of the associated constraint.[3] A shadow price of zero indicates that the associated constraint is not binding, such that a relaxation of the constraint would not affect the value of the objective function [11].

The simplification method utilizes the knowledge about which constraints are non-binding to reduce the number of coalitional values that must be explicitly calculated by solving the associated OPF. If the constraint associated with a grid element is non-binding, this constraint can be removed from the optimization problem without changing the outcome – corresponding to removing a player from a coalition. Because the outcome is not changed, the costs must be unchanged as well. Therefore, it is not necessary to solve the OPF without the corresponding grid element to learn its costs – they can be inferred from the coalition *with* the corresponding grid element.

The procedure can be illustrated with the following example: Let the grand coalition consist of players $\{a, b, c, d, e\}$, and let the shadow prices of players d and e be zero. Players d and e can be removed from the grand coalition without changing its solution. It follows that

$$c(\{a, b, c\}) = c(\{a, b, c, d\}) = c(\{a, b, c, e\}) = c(\{a, b, c, d, e\}), \tag{2}$$

with $c(S)$ being the cost of coalition S. In this case, $\{a, b, c\}$ is called the *base coalition*, because it contains no non-binding players. Instead of four coalitions, only one must be calculated, namely, the one for the set $\{a, b, c, d, e\}$, because (in this example) this calculation yields the information that the shadow prices of the players d and e are zero.

The developed algorithm recursively searches for base coalitions by removing *binding* players from coalitions, making note of non-binding players and therefore the costs that can be inferred without calculating the associated costs via the OPF along the way.

In [1], this method was applied to a small grid topology consisting of CIGRE benchmark grids ([13]) of multiple voltage levels with promising results. The test case included one time step with 11 congestions, normally requiring the solution of $2^{11} = 2048$ OPFs. Applying the simplification method showed promising results. Of the 2048 coalitions, only 328 coalitions had to be calculated, while the remaining 1720 coalitions could be inferred.

[3] The shadow price of a power flow constraint of a grid element represents the cost savings in the objective function if more power was allowed to flow over the grid element.

In the remainder of this paper, I investigate the applicability of this simplification method to a realistic grid topology over a whole year, and I analyze the factors determining the achievable benefits.

2.4 Methodology of the case study

2.4.1 Scenario data

The main focus of this paper is to investigate the effectiveness of the simplification method using non-binding constraints, developed in [1], in a realistic grid setting. For this purpose, a grid dataset, including grid topologies, time series data, power plants, renewable capacities, and storage units from the SimBench project ([10]) is used. According to the authors, SimBench "is intended as a benchmark to test, publish and compare methods and algorithms for various use cases, [...], [including] the fields of grid planning, operation and simulation"([10]), making it especially suitable for applying the algorithm discussed above.

From the many available grid configurations, the grid code *1-EHVHV-mixed-all-1-no-sw* was chosen. It includes the German extra-high-voltage (EHV) grid and two explicitly represented high-voltage (HV) grids connected to it. The grids are in the near-future "tomorrow" configuration. Where underlying grids are not explicitly represented, separate aggregated time series for demand, storage units, and renewable infeed are applied to the existing grid nodes. This enables the individual adjustment of these components, an important characteristic for redispatch calculations. To initially handle the SimBench data, the associated Python package was used.

The time series data originally have a quarter-hourly resolution for a whole year. They are resampled to hourly resolution by averaging over the 15-minute segments forming each hour. Costs for conventional power plants are assigned by determining efficiency ranges for the technology types available in the SimBench dataset. Together with CO2 emission factors, fuel prices, and other variable costs (see Table 1), ranges for marginal costs for each technology type are calculated.[4] The cost for curtailing renewable infeed is set to -50 EUR/MWh, ensuring curtailment is used as a last resort measure. Costs for individual power plants are randomly assigned from evenly spaced points from the marginal cost range for the respective technology. This pragmatic approach was chosen instead of manually matching individual power plants to their existing counterparts, where the true technological parameters would have still been unknown. Because the system costs (and redispatch costs) as such are not the focus of this paper, this trade-off is acceptable. The location of installed power plant capacities is shown in Figure 2.

[4] The efficiency ranges, emission factors, and other variable costs are taken from the ENTSO-E Midterm Adequacy Forecast 2020 Dataset ([14]). Fuel costs represent future notations for the year 2025, averaged over the last quarter of 2021. For the technology Waste, parameters have been chosen to reflect the must-run nature of waste power plants. The CO_2 price is set to 72.29 EUR/t, also representing the average notation for European Carbon Futures in the last quarter of 2021.

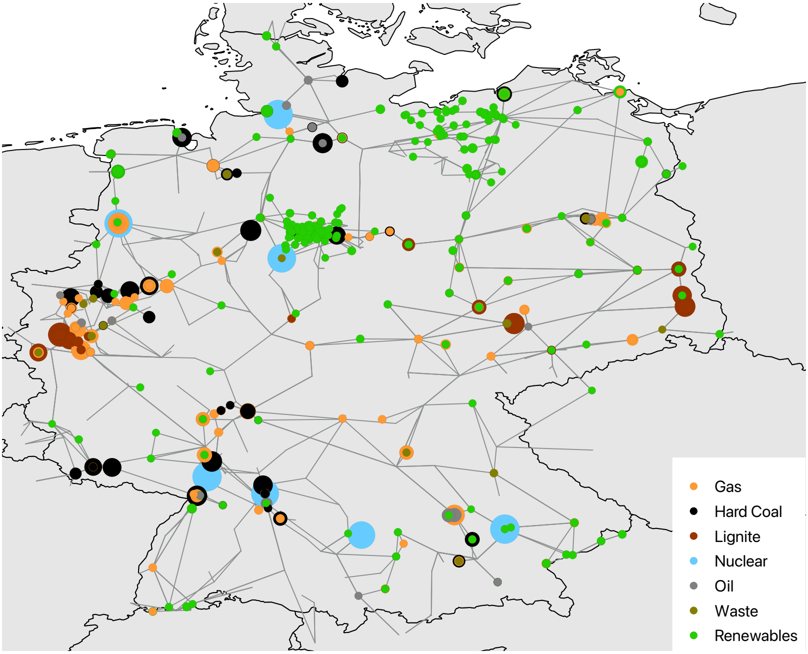


Figure 2: Map of scenario with installed capacities

Table 1: Assumption for costs of conventional power plants

Technology	Efficiency range		CO_2 Emission Factor	Fuel Price	Other variable costs
	Minimum	Maximum	t/MWh_{th}	EUR/MWh_{th}	EUR/MWh_{el}
Nuclear	0.30	0.35	0	1.70	9
Gas	0.33	0.60	0.21	24.92	1.60
Hard Coal	0.30	0.46	0.34	11.85	3.30
Lignite	0.30	0.46	0.36	4.24	3.30
Oil	0.25	0.43	0.28	33.98	3.30
Waste	1	1	0	0	2.45

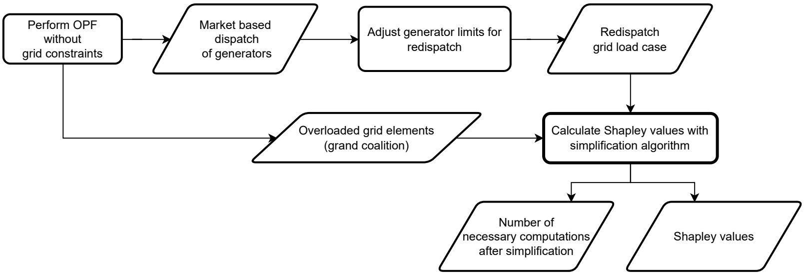


Figure 3: Process for performing the case study for a single grid load case

2.4.2 Process of calculating Shapley values

With the scenario data described in Section 2.4.1, the process for calculating the Shapley values for each time step can be performed. The process is shown in Figure 3. It has to be performed for each grid load case. First, an OPF without grid constraints is performed.[5] This represents a market solution in a zonal market configuration, such as the one in Germany. With the market-based dispatch of the generators, the redispatch case is prepared by adjusting the generator limits, as described in detail in [1]. Additionally, the grid elements that would be overloaded in the market solution can be determined from the initial OPF without grid constraints. These grid elements are the players forming the grand coalition for the grid load case under investigation. This information and the scenario data are supplied to the algorithm that calculates the Shapley value using the simplifications, as described in Sections 2.2 and 2.3.

3 Results

This section details the results of the case study and further analyses based on these results. The results of applying the simplification algorithm are detailed in Section 3.1. Further analyses of these results are performed in Section 3.2. Section 3.3 discusses whether the results can be generalized beyond the case study.

3.1 Overview

Figure 4 shows the number of samples within the 8784 time steps for each observable number of overloaded elements. The overloaded elements are the players forming the grand coalition in

[5] While time series data for power plants are available in the underlying dataset, they are disregarded to ensure consistency between the market solution and the redispatch solution, as determined by the OPF. Time series data are, however, used for demand, renewable infeed, and storage units.

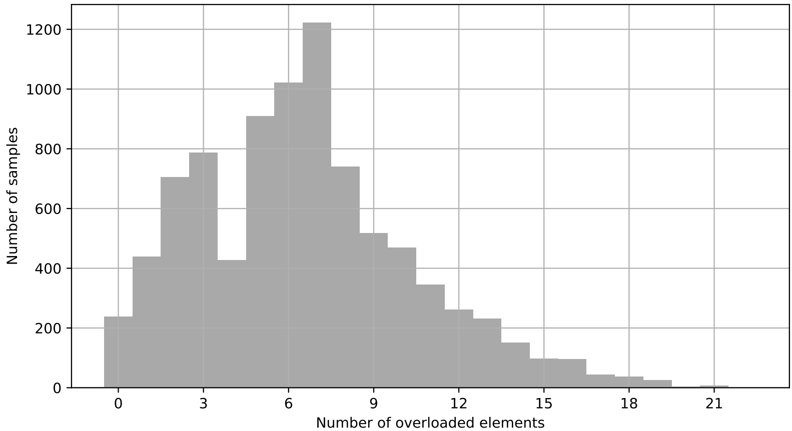


Figure 4: Number of samples within one year classified by the number of overloaded elements.

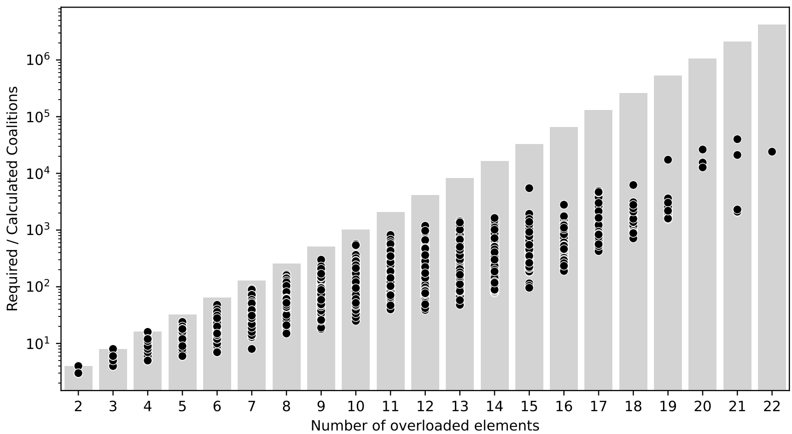


Figure 5: Number of required coalitions (bars) vs. number of calculated coalitions (circles).

each time step. The number of overloaded elements can therefore also be called the game size. The largest number of overloaded elements in any time step is 22, but this amount occurs only once. The average game size is 6.56, while the median is six overloaded elements. There are 679 time steps with no or one overloaded element. Because there is no savings potential in these time steps, they are excluded from further analysis, leaving 8105 time steps to analyze.

As shown in Figure 5, the simplification algorithm continues to be effective when applied to the realistic dataset described in Section 2.4.1. For each number of overloaded elements, the figure shows the number of coalition values that are required to calculate the Shapley values as bars (also called required coalitions) and the number of coalition values that must necessarily be calculated when using the simplification method as circles (also called calculated coalitions). Because of the exponential nature of the number of required coalitions, the y-axis is presented with a logarithmic scale. Additionally, Table 2 shows statistics for the required and calculated coalitions for each game size.

The number of calculated coalitions relative to the number of required coalitions decreases with the number of overloaded elements. The relative savings – the number of coalitions that do not have to be calculated due to the simplification method relative to the total number of required coalitions – are shown in Figure 6. The savings increase with the number of overloaded elements. At or above 10 overloaded elements, the mean savings exceed 90 % of required coalitions. For a game size (number of overloaded elements) of 19, the largest game size with more than 10 samples, only 2 709 coalitions out of the required 524 288 have to be calculated on average, representing around 0.5 % of the required coalitions.

Over the whole year, the simplification algorithm saves a total of around 67.6 million calculations of the objective function, representing 98.68 % of the calculations otherwise required. With an average time per coalition calculation of 0.03 seconds, this implies time savings of around 547 hours, or 3.25 weeks.[6]

3.2 Factors determining the effectiveness of the algorithm

Considering the range of calculated coalitions for each game size – shown in Table 2 and Figure 5 – a question arises: What determines the effectiveness of the algorithm in any specific time step?

Part of the answer lies in the number of binding constraints in the grand coalition. Figure 7 shows a boxplot of the number of binding constraints in the grand coalition for each number of overloaded elements. A binding constraint in the grand coalition indicates that after redispatch the grid element corresponding to this constraint (which was overloaded in the initial market result) is still congested, i.e. fully utilized. A non-binding constraint, however, indicates that the corresponding

[6]The calculations were performed on a machine with an Intel Core i9-9900K CPU.

Table 2: Statistics for calculated coalitions of each game size

size	required	count	calculated coalitions					mean savings (%)
			mean	std. dev.	min	median	max	
0	1	239	1.0	0.0	1	1	1	0.0
1	2	440	2.0	0.0	2	2	2	0.0
2	4	706	3.3	0.4	3	3	4	18.3
3	8	788	6.1	0.6	4	6	8	24.0
4	16	427	10.7	2.1	5	12	16	33.0
5	32	910	15.4	3.0	6	16	24	51.9
6	64	1021	22.7	5.7	7	23	48	64.6
7	128	1222	30.7	10.8	8	29	92	76.0
8	256	741	48.3	22.5	15	44	166	81.1
9	512	518	76.6	42.1	18	64	300	85.0
10	1024	470	99.1	62.3	25	74	568	90.3
11	2048	345	153.0	124.6	40	106	820	92.5
12	4096	261	198.8	151.4	39	155	1190	95.1
13	8192	232	283.1	221.2	48	211	1430	96.5
14	16 384	152	396.2	335.3	79	307	1634	97.6
15	32 768	98	536.4	604.9	96	418	5482	98.4
16	65 536	96	531.8	407.6	190	364	2798	99.2
17	131 072	44	1349.6	1203.1	426	765	4910	99.0
18	262 144	37	1780.6	945.4	715	1582	6240	99.3
19	524 288	25	2708.6	3096.2	1566	2016	17 405	99.5
20	1 048 576	4	17 127.3	6253.1	12 760	14 698	26 354	98.4
21	2 097 152	7	10 287.9	14 929.2	2070	2186	40 114	99.5
22	4 194 304	1	24 127.0		24 127	24 127	24 127	99.4

grid element is no longer fully utilized after redispatch, even though it was overloaded in the initial market results.[7]

The number of binding constraints in the grand coalition is lower than the number of overloaded elements in all samples with a game size of five or more. The maximum number of binding constraints occurs at a game size of 15 with 10 binding constraints. The tapering off of the number of binding constraints can be explained as follows: A grid load case with high overall power flows results in overloads in different areas of the grid and on multiple grid elements. Synergies in redispatch mean that there is not one individual set of measures designed to relieve overloads on each individual grid element. Therefore, relieving high absolute overloads may lower the flows over formerly overloaded smaller grid elements below their capacity, meaning they are no longer binding. Synergies in relieving congestion are the reason for applying the Shapley value to share the resulting costs in the first place.

[7] For an explanation, consider two power lines connected in series. Assume that both lines have a different thermal rating and that both are overloaded in the market result. In this case, the line with the lower rating limits the total power that can flow over both lines. After redispatch, the line with the higher rating will no longer be fully utilized, and the associated constraint will be non-binding, while the line with the lower rating will be fully utilized and have a binding constraint.

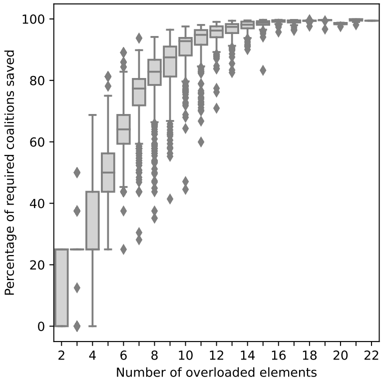


Figure 6: Percentage of required coalitions saved for each number of overloaded elements.

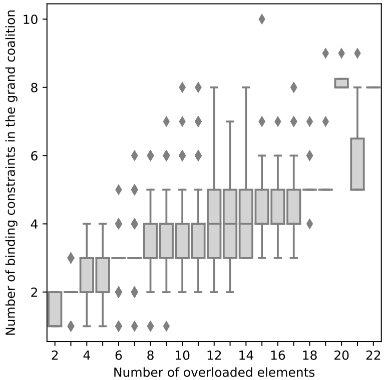
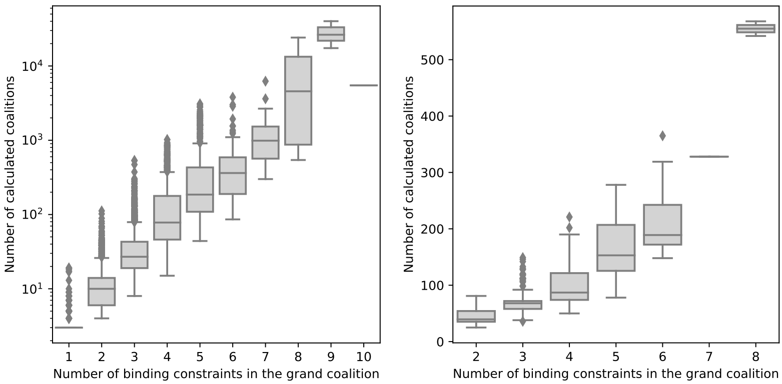


Figure 7: Number of binding constraints in the grand coalition for every number of overloaded elements.



(a) Across all game sizes.

(b) Only game size of 10.

Figure 8: Number of calculated coalitions for every number of binding constraints in the grand coalition.

Figure 8a shows a box plot of the relationship between the number of binding constraints in the grand coalition and the number of calculated coalitions after simplification. The y-axis is scaled logarithmically. There is a strong correlation between the number of binding constraints and the number of calculated coalitions, even without taking into account the game size. However, as with the relationship between the game size and the number of calculated coalitions (Figure 7), there is a large variance within each number of binding constraints, with values spread over three orders of magnitude in some cases. Figure 8b provides a filtered view of Figure 8a, showing only the data for a game size of 10.[8] There is still an obvious correlation between the number of binding constraints in the grand coalition and the number of calculated coalitions, but the variance within each number of binding constraints remains quite large.

To analyze the influences of the number of overloaded elements and the number of binding constraints in the grand coalition on the number of coalitions that have to be calculated, statistical analyses are performed. Four different exponential functions are fitted using non-linear least squares, with the last two being variations of the same underlying model:

$$n_{\text{calculated}} = a \cdot \exp\left(b \cdot n_{\text{overloaded}}\right) + \epsilon \tag{3}$$

$$n_{\text{calculated}} = a \cdot \exp\left(c \cdot n_{\text{binding}}\right) + \epsilon \tag{4}$$

$$n_{\text{calculated}} = a \cdot \exp\left(b \cdot n_{\text{overloaded}} + c \cdot n_{\text{binding}}\right) + \epsilon \tag{5a}$$

$$n_{\text{calculated}} = a \cdot \exp\left(b \cdot n_{\text{overloaded}} + d \cdot (n_{\text{overloaded}} - n_{\text{binding}})\right) + \epsilon \tag{5b}$$

The dependent variable $n_{\text{calculated}}$ is the number of calculated coalitions when using the simplification algorithm, while the explanatory variables are the number of overloaded elements $n_{\text{overloaded}}$ and the number of binding constraints in the grand coalition n_{binding}. Equation 5*b* is a reformulation of equation 5*a* that explicitly includes the difference between the number of overloaded elements and the number of binding constraints.

The results of the regression analysis are shown in Table 3. Using the number of overloaded elements or the number of binding constraints in the univariate models yields coefficients of determination R^2 of 0.498 and 0.398, respectively, suggesting that both variables on their own only have limited predictive power. This is consistent with the figures shown previously, where large variances can be observed.

When including both explanatory variables simultaneously, as specified in Equations (5*a*) and (5*b*), an R^2 of 0.97 is obtained. This suggests that most of the variance contained in the results can be explained by the interaction between the number of overloaded elements and the number of binding constraints in the grand coalition. The coefficients for model (5*b*) show explicitly that for any given game size, an increasing gap between the number of overloaded elements and the number of binding constraints in the grand coalition decreases the number of coalitions that must be calculated.

[8]Other game sizes exhibit similar patterns.

Table 3: Regression results for a nonlinear regression using least squares

	Dependent Variable:			
	Calculated Coalitions			
	(3)	(4)	(5a)	(5b)
Constant (a)	0.077	8.459	0.203	0.203
Number of overloaded elements (b)	0.571		0.326	0.908
Binding constraints in grand coalition (c)	-	0.801	0.583	-
Δ Overloaded-Binding (d)	-	-	-	−0.583
R^2	0.498	0.398	0.970	0.970
Observations	8105	8105	8105	8105

This formulation also has an implication that, at first sight, may seem implausible: Assuming a gap of zero, the number of saved coalitions becomes negative for game sizes larger than seven. However, this is not a realistic assumption. In the underlying dataset, there is no game size larger than four without non-binding constraints. With realistic combinations of game size and non-binding constraints, the estimation yields plausible results. While the coefficients should not be used to estimate the number of calculated coalitions in other grid models, it is expected that the general relationship holds – i.e. that the number of binding constraints in the grand coalition and the difference between the number of overloaded elements and binding constraints in the grand coalition jointly largely determine the savings that can be achieved with the simplification algorithm.

3.3 Generalization of the performance claims

The algorithm has shown its merits in two very different settings: the original benchmark grid used in [1] and the realistic grid investigated in this contribution. Nevertheless, even the realistic grid does not include a detailed modeling of the topology of lower voltage levels. Instead, the relevant units and time series are aggregated at the higher voltage nodes. Regarding further generalization of the results, the following question arises: What are the implications of adding more low-voltage grids to the grid topology?

The answer depends on the specific details of the low-voltage grids to be added. The nature of low-voltage grids is such that each individual overload is much smaller than the overloads in the transmission grid because power flows in general are much smaller. Further, the causes of overloads may often be the same as in the transmission grid: an abundance of infeed from renewable energy sources. Consequently, large synergies are to be expected, where actions that solve overloads in the underlying low-voltage grids are also beneficial for overloads in the transmission grid. Whether or not these synergies translate to a large number of non-binding constraints in these low-voltage grids depends on the specific location of the low-voltage grids and the redispatch measures available in them.

For suitably located grids, the measures taken to cure their overloads are also the cheapest measures available to cure the larger overloads in higher voltage levels. In this case, overloaded elements from low-voltage grids may be non-binding, resulting in a large gap between the number of elements that are overloaded before redispatch and the number of elements that are still congested (and therefore binding constraints) after solving the redispatch optimization. As Section 3.2 has shown, this gap seems to be largely responsible for the savings that can be achieved with the algorithm. Hence, it should continue to work well and may even become more effective when including low-voltage grids in the grid model.

However, the opposite case is also plausible: Depending on the exact location of the low-voltage grids, the available measures to cure overloads in these grids may also reduce overloads in the transmission grid, but they may not be the cheapest way to do so. In that case, while there would still be synergies, measures in the low-voltage grid would only be used as far as is required to cure the specific congestions. Other measures would be used to resolve remaining overloads on other grid elements. In this way, the element in the low-voltage grid would remain binding. Consequently, in this situation, an overload would be added, but the gap size would remain constant.

In any case, substantially increasing the number of overloaded elements may be problematic. The relationship between the number of calculations still necessary when using the simplification algorithm and the number of overloaded elements remains exponential, albeit with a smaller growth rate than in the original problem.

4 Conclusion

Fairly allocating costs arising from congestion management will become increasingly important, with congestions moving to lower grid levels and affecting more grid operators. The Shapley value has been suggested for allocating these costs to each affected grid element and by extension to the grid operators that operate them. However, using the Shapley value requires performing a large number of calculations, as the costs for alternative redispatch settings must be determined. Previous work [1] has developed a simplification algorithm that substantially reduces the number of calculations that must be performed and has demonstrated the effectiveness of the algorithm on a small benchmark grid with one grid load case.

This contribution has aimed to assess the merits of the simplification algorithm (1) in a grid with a more realistic size and (2) with a greater variety of grid load cases. To this end, a detailed representation of the German transmission grid with two high-voltage grids and scenario data from the SimBench Project [10] have been used. The analytic simplification algorithm using non-binding constraints from [1] has been applied to calculate the Shapley values for all 8784 hours of a sample year.

The results show that the benefits of the simplification algorithm in a realistic grid exceed those observed in the initial tests in the benchmark grid. The algorithm continues to substantially reduce the number of calculations that have to be performed, and savings increase with the number of overloaded elements in a given time step. In time steps with more than 15 congestions, more than 99% of the originally required calculations can be saved in most cases. Regression analyses show that the achievable savings depend on the gap between the number of overloaded elements in a given time step and the number of constraints associated with these elements that are binding in the redispatch optimization.

The performance of the algorithm is expected to translate well to other grid models, as long as there are substantial synergies achieved in congestion management. This is also the use case where applying the Shapley value to allocate the associated costs is most beneficial, meaning the algorithm works best where it is needed the most.

Regarding the translation of the operative and often decentral processes that are implemented in practice into the central planner approach that is the basic assumption of the OPF calculation, further research is needed. Additionally, further operational constraints of the units used for redispatch, such as unit commitment decisions, minimum run times, and minimum stable operation limits, may also have to be taken into account. How these constraints should be considered – or worked around – requires further investigation.

Overall, this contribution represents an important step toward enabling the usage of the Shapley value to allocate costs from congestion management. Further research should work towards implementing the Shapley value with the simplification algorithm in the operative processes to replace the current cost allocation methods.

5 Acknowledgements

I thank Christoph Weber for his valuable comments and interesting discussions.

References

[1] S. Voswinkel, J. Höckner, A. Khalid, C. Weber, Sharing congestion management costs among system operators using the shapley value, Applied Energy 317 (2022) 119039. doi:10.1016/j.apenergy.2022.119039.

[2] M. Bjørndal, K. Jørnsten, Zonal pricing in a deregulated electricity market, The Energy Journal 22 (1) (2001) 51–73.
URL www.jstor.org/stable/41322907

[3] L. J. de Vries, R. A. Hakvoort, An economic assessment of congestion management methods for electricity transmission networks, Journal of Network Industries os-3 (4) (2002) 425–466. doi:10.1177/178359170200300403.

[4] N. Andreadou, M. G. Flammini, G. Fulli, M. Masera, G. Prettico, S. Vitiello, Distribution system operators observatory 2018: Overview of the electricity distribution system in Europe, Vol. 29615 of EUR, Scientific and technical research series, Publications Office of the European Union, Luxembourg, 2019.

[5] Bundesnetzagentur, Bundeskartellamt, Monitoringbericht 2021 (Mar. 2022).
URL https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Monitoringberichte/Monitoringbericht_Energie2021.pdf?__blob=publicationFile&v=6

[6] Bundesnetzagentur, Bericht netzengpassmanagement gesamtes jahr 2021 (Jul. 2022).
URL https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/Engpassmanagement/Zahlen%20Ganzes%20Jahr2021.pdf

[7] L. S. Shapley, 17. a value for n-person games, in: H. W. Kuhn, A. W. Tucker (Eds.), Contributions to the Theory of Games (AM-28), Volume II, Princeton University Press, Princeton, 1953, pp. 307–318. doi:10.1515/9781400881970-018.

[8] M. Shubik, Incentives, decentralized control, the assignment of joint costs and internal pricing, Management Science 8 (3) (1962) 325–343.
URL http://www.jstor.org/stable/2627389

[9] A. E. Roth, R. E. Verrecchia, The shapley value as applied to cost allocation: A reinterpretation, Journal of Accounting Research 17 (1) (1979) 295. doi:10.2307/2490320.

[10] S. Meinecke, D. Sarajlić, S. R. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, A. Moser, M. Braun, Simbench—a benchmark dataset of electric power systems to compare innovative solutions based on power flow analysis, Energies 13 (12) (2020). doi:10.3390/en13123290.

[11] R. Sioshansi, A. J. Conejo, Optimization in Engineering, Springer International Publishing, Berlin, Heidelberg, 2017. doi:10.1007/978-3-319-56769-3.

[12] M. G. Fiestras-Janeiro, I. García-Jurado, M. A. Mosquera, Cooperative games and cost allocation problems, TOP 19 (1) (2011) 1–22. doi:10.1007/s11750-011-0200-1.

[13] CIGRE, Benchmark systems for network integration of renewable and distributed energy resources, Technical Brochure, Task Force C6.04.02 (2014).

[14] ENTSO-E, Midterm adequacy forecast 2020 dataset (Nov. 2020).
URL https://eepublicdownloads.entsoe.eu/clean-documents/sdc-documents/MAF/2020/MAF%202020%20-%20Dataset.xlsx

Chapter 7

Conclusions

This thesis comprises five articles that are included in Chapters 2 to 6. In this concluding chapter, the contributions of these articles to the research questions raised in Section 1.2 are highlighted and summarized.

7.1 How can different market designs be modeled with linear programming and what are the strengths and weaknesses?

Electricity market modeling and electricity market design play a role in each of the five articles included in this thesis – either market designs are themselves subject to investigation, such as in Chapter 2, Chapter 3, and Chapter 4, or market modeling is indirectly supporting other investigations in Chapter 5 and Chapter 6. In each case, the market is modeled as an optimization problem.

Linear programming (LP) and mixed integer programming (MIP) are two types of optimization that have long been established as methods to model energy markets (Pfenninger et al. 2014). Both have distinct advantages and drawbacks. Linear problems are comparatively quick to solve, allowing for investigations with a large scope, such as the modeling of international electricity markets in Europe. Drawbacks include, for example, the inability to include binary variables, which, among others, are used to represent unit commitment decisions for power plants: Because each variable must be continuous, it is not possible to constrain a power plant to produce at a minimum level or nothing at all. Mixed integer problems allow for binary variables and can accordingly represent operational constraints much more realistically. The drawback is a solution time that can be orders of magnitude longer than a comparable linear model. In comparison, simple linear problems can even be solved graphically.

Within this thesis, Chapter 3 provides the most substantial contribution to the research question. In this chapter, a framework is developed to model each stage of Flow-Based Market Coupling (FBMC), the mechanism to implicitly allocate cross-border transmission capacity in Central Western Europe (CWE) and Central Eastern Europe (CEE). FBMC constraints are included in the algorithm of the day-ahead auction (cf. Section 1.1.3). Their implementation requires the calculation of remaining available margins (RAMs) and power transfer distribution factors (PTDFs) in advance, these are also called flow-based parameters. To calculate the flow-based parameters, transmission system operators (TSOs) use the expected market results, which are gathered from experience and reference situations. After the auction, TSOs perform redispatch to manage congestion and ensure stable grid operation.

When modeling FBMC, gathering the required input for the calculation of the flow-based parameters, i.e., the capacity calculation, presents a circular problem: The capacity calculation requires knowledge about the probable market outcome, but the probable market outcome depends on the capacity calculation. The developed framework approaches this problem by iterating through multiple stages: First, the nodal problem is solved, corresponding to modeling nodal pricing (cf. Section 1.1.5). The results are used as a first approximation of the expected market outcome for the capacity calculation. With the calculated flow-based parameters, the zonal problem is solved. The results are better suited to represent the expected market outcome, since – like the actual market outcome – they are the result of the zonal algorithm. The results are used for a second iteration of the capacity calculation, yielding a more realistic set of flow-based parameters. The zonal problem is solved a second time with these flow-based parameters, now representing the actual market clearing of the day-ahead auction. Afterward, the results are used as the starting point for the last stage, another nodal calculation this time representing the redispatch problem.

The article in Chapter 4 delves into a refinement of the market design and adds more iterations to the model, (cf. Figure 3 in Chapter 4). In this article, the consideration of measures available for later redispatch, so-called redispatch potential (RDpot), in the day-ahead auction is investigated. RDpot is considered by extending the zonal optimization problem. The constraints are adjusted so that a higher flow over a given capacity constraint is possible if there are enough potential measures to alleviate resulting overloads via a more fine-grained approach in the redispatch stage. The use of RDpot adds cost to the objective function and is therefore used only if it reduces overall costs. The available RDpot is an input parameter for the zonal optimization problem. It is determined by first solving the zonal problem *without* RDpot and evaluating the resulting dispatch – unutilized generator capacity enables positive potential, meaning generation can be increased, while already utilized capacity corresponds to negative potential, meaning generation can be decreased. The zonal problem is then solved again to represent the auction *with* RDpot.

The article also evaluates different rules to determine the available (potential) measures. One of these rules only considers capacity under RDpot that is not used for redispatch in the reference case, i.e., when RDpot is absent from the zonal auction. Applying this rule therefore also necessitates the solution of the redispatch problem after the zonal problem without RDpot. In summary, the simulation of including the RDpot in the market design requires one extra calculation of the zonal problem and, depending on the rule used to determine RDpot, one extra calculation of the redispatch problem.

Each stage is modeled as a linear problem. This includes the nodal calculations (serving as the original input for the calculation of the flow-based parameters as well as providing the basis for the redispatch problem), which are performed using a linearized form of the normally non-linear load flow equations. Additionally, each time step that is considered in these calculations (out of usually 8760, representing a full year in hourly resolution) is treated as a separate optimization problem, isolated from other time steps. The linear formulation of these problems along with the isolated consideration of each time step allows for a consistent comparison of different market design choices. While it is not unusual to use a model with coupled time steps for zonal calculations, and even for MIP models to be used in zonal calculations, this is hardly possible when performing nodal calculations for a full year. However, using differently specified models for the calculation of nodal and zonal designs makes certain comparisons inconsistent. If the redispatch problem is subject to fundamentally different constraints than the zonal problem, it is not evident how to isolate effects (and costs) resulting from diverging problem formulations and costs resulting from the evaluated market designs.

However, the use of linear problems without time step coupling presents its own set of problems. As stated above, LP models do not allow for unit-commitment decisions, thereby preventing the consideration of factors like start-up costs and allowing unrealistic part load operation of generators. Disregarding time step coupling significantly impedes the realistic simulation of electricity storage, such as pump hydro storage and battery storage. These are included by using the output of a more detailed model that remains static and does not vary between different simulations. This is, however, still much less precise than modeling storage assets endogenously and introduces an error that increases with the importance of such storage technologies for the overall system. Overall, however, this error remains consistent across all comparisons performed with the model framework, and therefore the comparisons are expected to be consistent as well.

A slightly different and much simpler approach to modeling electricity markets is chosen in Chapters 5 and 6. In these articles, the market design itself is not the subject of the evaluations. However, because the articles focus on redispatch, the analyses required a market result on which to perform redispatch. Both articles only consider a single bidding zone (a simple benchmark model in Chapter 5 and the German bidding zone (without Luxembourg) in Chapter 6), therefore simulating the market auction without any kind of market coupling. In this case, the zonal problem reduces to a simple cost minimization without any grid constraints. This is equivalent to a nodal problem with infinite line capacities, allowing for the same model to be easily used for redispatch and the zonal problem. Both articles use a linear optimization problem with linearized power flow equations to enable the application of the Shapley value, which would be far too computationally expensive to calculate otherwise (cf. Section 7.2). The drawbacks of linear optimization compared to mixed integer models in general and the use of the linearized DC formulation of the power flow problem instead of the physically more accurate non-linear power flow equations are therefore present but fairly inconsequential in this analysis. Even if implemented in a real-world application it is used only as an ex-post analysis and is not relied on for securing the operational security of the grid. Therefore, enabling the usage of Shapley values for congestion cost allocation far outweighs the inaccuracy that is introduced by the use of linearized models.

The article in Chapter 2 focuses on analytical analyses and is not backed by a comprehensive computer model. The presented analyses are based on graphical analyses of the merit order in a flexibility market. Yet, these again correspond to a type of graphical solution to a (simple) linear optimization problem. This illustrates the wide-ranging applicability of linear programs and the transparency they can provide in problem-solving.

7.2 How can the use of Shapley values and linear programming contribute to achieving a fair allocation of congestion costs?

As laid out in Section 1.1.4, the costs resulting from congestions in the grid can be either implicitly or explicitly allocated, depending on the extent to which congestions have been addressed in the market clearing and how much ex-post redispatch is required. The distributional effects of implicit and explicit cost allocation differ: On the one hand, Shapley values calculated using linear programming techniques can be used to fairly allocate congestion costs explicitly. On the other hand, linear programming can be used to determine distributional effects of implicit cost allocation, which may in turn further inform the discussion about fair cost allocation.

The explicit cost allocation required by ex-post redispatch is the subject of the articles in Chapter 5 and Chapter 6. Continuing the theme of using linear programming in the analysis of electricity market designs, Chapter 5 first demonstrates the use of the Shapley value, a concept from cooperative game theory, for fairly allocating congestion costs and then introduces two methods to speed up the comprehensive calculation of the Shapley value. Both methods (for clarity and brevity called the *analytical* and *numerical* methods) build on features inherent to linear optimization.

Fairly allocating costs can be interpreted in several different ways. In the context of allocating explicit costs arising from redispatch, a fair allocation may entail allocating the costs to the parties responsible for the redispatch in the first place in a way that (1) costs are recouped and (2) costs are allocated relative to the share that the parties had in causing them. These two points are discussed later in this section. Regarding the parties responsible for redispatch, two class of parties may be distinguished: One class of parties consists of grid users, i.e., electricity generators and demanders. Another class is comprised of the grid operators responsible for the congested grid elements. As the grid within a bidding zone in a zonal market design is meant to be mostly free of congestion (cf. Section 1.1.3), it is sensible to assign the responsibility for redispatch costs to the grid operators instead of the grid users. On the other hand, the grid operators will recoup their costs by charging network fees to the grid users, which again raises questions of fairness. However, the analyses in this thesis are limited to the grid operation level.

The details of calculating the Shapley value are explained in Chapter 5. It can be used to allocate costs or profits arising from a grand coalition of players to each player according to the individual player's contribution to the profits or costs.[1] The Shapley value has multiple properties that cause it to be commonly referred to as being fair: Only actually incurred costs (and these costs exactly) are distributed to the players, equivalent players are assigned the same Shapley value, and players not contributing to the costs are not charged. Additionally, the Shapley value is well suited to problems with synergies, making it a good fit for congestion management: Its calculation considers the contribution of each player to the costs of each possible coalition of players in all permutations of these coalitions, i.e., being added to the coalition first, last, or somewhere in between. In electricity grids, overloads on large transmission lines may overshadow overloads on smaller lines, and resolving these large overloads may resolve the smaller ones as well. Nevertheless, arguably the operators of the smaller lines should be allocated some of the costs for relieving the congestion. The Shapley value can be used to determine exactly how much of the costs should be allocated to each line.[2]

To allocate congestion costs, the individual overloaded grid elements are defined as players, and the grand coalition is formed by all the elements that without action would be overloaded in any given time step.[3] The costs to be allocated to the overloaded grid elements are the redispatch costs for the time step. Calculating the Shapley value requires knowledge of not only the costs for the grand coalition, but the costs of every possi-

[1] In this context, *Shapley value* is used as the name of the concept. However, the Shapley value also refers to the specific part of the profits or costs allocated to a player. Each player has their own specific Shapley value and the sum of the Shapley values is the total profits or costs to be allocated.

[2] As public information about the ownership of individual lines is limited and given that the technical aspects of congestion require the consideration of individual lines, the focus is subsequently placed on single grid elements as players and not on the economic entities (grid operators) they belong to.

[3] In this section, *overloaded elements* always refers to elements that *would* be overloaded, if redispatch were not performed. Overloaded elements are also congested. After redispatch, elements can still be *congested* (i.e., fully utilized and constraining the dispatch), but redispatch removes all overloads.

ble sub-coalition as well. Sub-coalitions can be constructed by removing players from the grand coalition. In the context of allocating congestion costs, this corresponds to removing the need to perform redispatch for a specific grid element. This is accomplished by modifying the constraint of the grid element so that its utilization does not correspond to being overloaded.[4] With N players (grid elements) in the grand coalition, the number of total possible coalitions is 2^N. Because the costs for each sub-coalition are calculated by solving a (linear) optimization problem (the redispatch problem), the calculation of 2^N coalitions is challenging. The two methods developed in Chapter 5 simplify the calculation of the Shapley value by reducing the number of sub-coalitions for which costs must be calculated.

The *numerical* method uses a general feature of optimization theory: In any minimization problem, the value of the objective function cannot decrease when adding a new (or tightening an existing) constraint and cannot increase when removing (or relaxing) a constraint. As detailed above, sub-coalitions for the redispatch problem are constructed by relaxing the constraints of the grid elements that are no longer part of the coalition. Accordingly, if two coalitions of different sizes share the same costs and the smaller coalition is a subset of the larger coalition, any combination of players not in the smaller coalition can be removed from the larger coalition without changing its value: The value cannot increase by removing a player, and it also cannot decrease, because the smaller coalition has the same value and all of its players are also part of the larger coalition.

This insight is utilized by systematically calculating coalitions of different sizes and comparing their values. Details of the algorithm are described in Chapter 5. If only precisely matching values are considered, the algorithm yields exact results. Further computational savings can be achieved by introducing a tolerance for costs to be considered equal. However, this reduces the precision of the calculated Shapley value.

The analytical method follows a related approach but depends on the specifics of the underlying optimization problem. It utilizes the fact that any non-binding constraint in a continuous optimization problem (i.e., without integer variables) can be removed from the problem formulation

[4] An example: The power flow calculation shows a flow of 1100 MW on a transmission line with a capacity of 1000 MW – an overload of 100 MW. To remove this grid element from the grand coalition, its capacity is adjusted to 1100 MW.

without changing the optimal value (Sioshansi and Conejo 2017). In the context of the redispatch problem, a non-binding constraint corresponds to a grid element, that (while being overloaded without redispatch) after performing redispatch is no longer fully utilized. In this case, removing this grid element from the grand coalition by the means described above will not change the value of the objective function (the costs for performing redispatch). Whether a constraint is binding can be determined by examining the so-called shadow price of the constraint, which is returned as part of the solution of the optimization problem. The shadow price signifies the change in the objective function if the associated constraint is relaxed. If the shadow price is zero, the constraint is not binding. Because the players in the coalition correspond to constraints in the redispatch problem, binding and non-binding constraints translate into binding and non-binding players. Knowing that any combination of non-binding players can be removed from a coalition without changing the costs, calculating the costs of any coalition that can be constructed with the binding players and any subset of the non-binding players is not required. Starting with the grand coalition, the algorithm strategically builds new coalitions looking for non-binding players and thereby inferring as many coalition costs as possible.

Both methods were implemented, demonstrated, and compared to other established simplification methods on a benchmark grid topology (CIGRE 2014) with 11 overloads, normally requiring the calculation of 2048 coalition costs. With the numerical method without a tolerance, 975 of the 2048 coalitions do not have to be calculated. With a tolerance of 5%, 1100 of 2048 coalitions do not have to be calculated, but at the expense of reduced precision of the Shapley value. The analytical method performs much better, only requiring the calculation of 328 out of 2048 coalitions. However, the numerical method relies on the knowledge of coalition costs alone, while the analytical approach requires knowledge of further optimization outputs (the non-binding constraints). In settings where this knowledge is not readily available, e.g., because of pre-existing software tools, using the analytical approach may not be possible, leaving the numerical approach as a fallback. Compared to established simplification techniques, such as sampling, results were promising but mixed. The numerical method with no applied tolerance yields exact results, while sampling induces errors of almost 16% when choosing a sample size that requires a similar number of calculations. However, when increasing the tolerance in the numerical method, the resulting (relative) error can be very large for small absolute values. This is currently not predictable, yet could be investigated by further research.

The analytical method was further analyzed and validated in Chapter 6. It extends the application from the benchmark grid with one time step in Chapter 5 by using data for the whole German grid as published in a dataset from the Simbench project (Meinecke et al. 2020) and hourly time series data for a full (leap) year, covering 8784 hours. This (1) allows the demonstration of the effectiveness of the algorithm in a more realistic setting and with more instances (grid load cases), and (2) allows analyses of the factors influencing the effectiveness of the algorithm.

Results show that the share of coalitions that do not have to be calculated increases with the number of players in the grand coalition, i.e., the number of overloaded elements in any given time step. In a time step with 22 overloaded elements, only 24 127 out of 4 194 304 coalitions had to be calculated, saving 99.4 % of calculations. This, however, occurs only in a single instance. For a grand coalition size of 19 (arising in 25 of 8784 time steps) an average of 2708.6 out of 524 288 coalitions had to be calculated, with numbers varying between 1566 and 17 405. Regression analysis ($R^2 = 0.97$) reveals the main influences to be the number of overloaded elements (the size of the grand coalition) and the number of binding constraints in the grand coalition (the number of formerly overloaded elements that are at full capacity after redispatch): Bigger differences between the size of the grand coalition and the number of binding constraints lead to increased savings when applying the simplification algorithm.

Because of the demonstrated consistently high savings of the analytical method, it can be considered a step toward increasing the feasibility of implementing the Shapley value in a practical setting.

While Chapters 5 and 6 have used linear programming and fair cost allocation methods to distribute the costs of ex-post congestion management, the distributional effects of the part of congestion management that is implicit in FBMC can be analyzed with the methods developed in Chapters 3 and 4.

By preventing the dispatch of the cheapest technologies where grid constraints prohibit the transmission, FBMC influences prices and the distribution of rents between producers and consumers in addition to introducing congestion rents, which accrue to the TSOs. In Chapter 3, multiple methods for calculating so-called GSKs (representing the generation pattern for changes of the net position of a bidding zone) are analyzed. Furthermore, the effects of minRAM (mandating minimum margins available for trade on critical network elements), the consideration of internal critical network elements, and TSO safety margins on critical network elements are considered. Chapter 4 additionally analyzes the effects of including the known potential for redispatch. Comparisons between the results of these scenarios allow for the calculation of the distributional effects of congestion management, which may serve as input for further discussion about the fairness of certain design choices.

While producer and consumer rents are trivial to calculate (being the volume-weighted difference between the electricity price and generation costs or value of lost load[5] respectively), the calculation of congestion rents and especially the allocation of congestion rents to individual bidding zones are more involved. CWE CIA WG (2020) describes the methodology as it is currently employed. Congestion rents are not part of the analyses undertaken in Chapters 3 and 4. However, the developed methods allow for the implementation of this calculation because the employed LP optimization yields all the required parameters such as shadow prices of the capacity constraints and prices for the individual bidding zones. Further research could also analyze the effects on producer and consumer rents at a bidding zone level using the methods developed in these chapters.

7.3 What are the welfare implications of the trade-off between increasing trading possibilities and accounting for grid constraints?

In nodal pricing, the constraints posed by thermal capacities and voltage limits of grid elements are directly considered in the market clearing algorithm (cf. Section 1.1.4). However, this comprehensive consideration may severely limit trading possibilities. Zonal pricing uses simplified grid constraints to increase trading possibilities but requires redispatch to ensure that the physical constraints are met. Recently, several amend-

[5]The value of lost load can be understood as the maximum acceptable electricity price for consumers before accepting service interruptions, cf. Weber et al. (2022, p. 186).

ments to FBMC in Europe have aimed to further increase trade by posing restrictions on the constraints themselves (ACER 2019). The most prominent of these amendments is the minRAM stipulation, which requires that RAMs – the margins of line capacities that remain available to the market – must exceed 70 % of the thermal line capacity. Furthermore, considering internal branches as CNECs is discouraged by requiring declaration well in advance and detailed written justification including the analysis of all other options. Furthermore, the consideration of the so-called redispatch potential (RDpot), analyzed in Chapter 4 based on the concept first described by the Belgian TSO Elia (Elia Group 2019), is also meant to increase trade.

The trade-off between increasing trade and disregarding grid constraints is studied in Chapters 3 and 4. As explained in Section 7.1, the developed model framework allows consistently comparing model outcomes of different market designs. In Chapter 3 this framework is used to compare different designs options of FBMC to each other and to nodal pricing, in Chapter 4 it is used to study the impact of the implementation of redispatch potential. The studies calculate the system costs and do not directly calculate welfare. However, as described in Section 1.1.5, when considering inelastic demand, minimizing costs equates to maximizing welfare. Consequently, increased costs translate directly into decreased welfare and vice versa.

Both chapters consistently show that nodal pricing always outperforms zonal pricing. This result is well backed by the literature, where nodal pricing often serves as the first-best benchmark (Bjørndal and Jørnsten 2001; Bjørndal and Jørnsten 2007; Oggioni and Smeers 2013; Felling et al. 2023). There are two direct causes for higher costs in a zonal setting with redispatch compared to nodal pricing: (1) redispatch should adjust the market result only as far as it is required for safe grid operation and (2) additional costs arise from opportunity costs and other factors stemming from the TSOs directly ordering the dispatch of specific assets (cf. Section 1.1.5). Both points are reflected by an adjusted objective function in the models used in Chapters 3 and 4. Firstly, a general penalty for redispatch volume is added to the objective function. This reflects the objective of the TSOs to keep the redispatched volume to a minimum. Secondly, the marginal costs of the power plants are multiplied by penalty factors to account for the increased costs described above.

Chapter 3 places FBMC between two extremes of accounting for grid constraints: On the one end, *nodal pricing* incorporates all physical constraints, limiting trade to ensure a physically feasible market result. On the other end, grid constraints are disregarded completely in an *unlimited trade* scenario. The cost difference between nodal pricing and the unlimited trade scenario is considered as the maximum benefit of implementing nodal pricing. Different FBMC implementation scenarios are evaluated based on their cost savings compared to the unlimited trade scenario, and these savings are in turn expressed as a percentage of the theoretically achievable savings. In an optimal setting, with no congestion within bidding zone borders, FBMC achieves 87 % of the cost savings achieved by nodal pricing. However, this number drops to 59 % when the actual congested elements within bidding zones are considered. This shows that FBMC is highly dependent on bidding zone configuration. Without appropriate bidding zone configurations, other substantial changes, such as different methods for calculating generation shift keys (GSKs, cf. Section 1.1.3) have little effect on overall system costs.[6] Splitting the German bidding zone is a measure that has long been discussed as part of a solution to the suboptimal bidding zone configuration within the scope of FBMC, but especially the German government has opposed this idea. The discussion around this topic continues to evolve. The Expert Commission for the Monitoring Process "Energy of the Future" recently recommended to the government to drop its resistance should the current bidding zone review of the European Network of Transmission System Operators for Electricity (ENTSO-E) arrive at a positive conclusion regarding the effects of a split of the German bidding zone (Löschel et al. 2023). Furthermore, the German federal state of Schleswig-Holstein has officially asked for the Federal Council (Bundesrat) to adopt a position in favor of a reevaluation of a bidding zone reconfiguration (Schleswig-Holstein 2023). Keeping a united German bidding zone can be viewed as a trade-off between the benefits (a uniform wholesale price and frictionless trade within Germany) and lower total costs when accounting for the physical grid constraints. The consideration of minRAM significantly increases redispatch volumes

[6] Recently, Schönheit et al. (2020) have found large impacts of different strategies for determining GSKs on the flow-based domain. While this at first seems to contradict the findings in Chapter 3, even a substantially different flow-based domain does not necessarily lead to different total system costs, especially including redispatch. Quantifying these effects as costs is beyond the scope of their analysis.

by relaxing the constraints imposed on the market clearing algorithm. However, depending on the parametrization of redispatch costs, the effect on costs varies in the chosen setting between a slight increase in costs to a substantial decrease. Since the parametrization that leads to a substantial decrease in costs is a sensitivity analysis and not a realistic scenario, no clear conclusions may be drawn from the observed costs. With an updated grid, updated scenario data, and a parametrization of redispatch costs between the extremes considered in Chapter 3, in Chapter 4, costs for minRAM rise to a much larger extent (107 million Euros in Chapter 4 compared to 10 million Euros in Chapter 3). Together with the substantially increased volumes already observed in Chapter 3 (independent of the cost parametrization), minRAM appears to negatively impact the system on balance. Similar results are observed by Schönheit et al. (2021a), who calculate cost increases of 7.25 %.

Chapter 4 implements potential redispatch measures into the zonal clearing and compares the results to FBMC without any adjustments as well as to minRAM and combinations of both. Considering potential redispatch measures, a concept based on the dispatch hubs introduced by Elia Group (2019) and further analyzed by Schlecht and Hirth (2021), increases trade by enabling the market algorithm to allow otherwise infeasible outcomes if redispatch units are known to be available to "correct" the result on a finer granularity where required.

Results show that this more targeted approach (compared to minRAM) of considering potential redispatch measures manages to increase trade while at the same time reducing overall costs. If combined with minRAM, trade increases further, but costs increase. As minRAM is unlikely to be abolished, the article recommends complementing the already implemented minRAM approach with the consideration of potential redispatch measures in order to decrease costs and further increase trade.

As stated above, both studies clearly show that nodal pricing performs better in every analyzed scenario. However, limitations apply: The reason for preferring increased trade to a more detailed consideration of grid constraints is the general economic insight that larger markets foster competition. Where nodal pricing is implemented, actors face a reduced level of competition. Consequently, these markets require tight regulation to prevent the abuse of market power. The models discussed above consider functioning markets without any market power even in the nodal pricing case. As such, the models are by design unable to reveal the benefits of zonal pricing compared to nodal pricing. While nodal pricing can serve as an effective benchmark for different implementation options of zonal pricing, the results should not be overinterpreted in the sense that they show the overall superiority of nodal over zonal pricing.

7.4 What contributes to the efficient operation of zonal electricity markets with congestion management?

Due to the financial volume and practical importance of European electricity markets, their efficient operation is crucial. Section 7.3 has already shown that policies such as minRAM can increase the system costs arising from FBMC. Under the assumption that all relevant factors are covered in the underlying models, more expensive solutions are less efficient than cheaper alternatives. In the context of this section, costs will therefore serve as a proxy for the efficiency of the overall process – the limitations of this assumption will also be addressed.

All papers contribute to the question at hand. Many of the implications arising from Chapters 3 and 4 have already been discussed in the context of Section 7.3: Increasing trade to the detriment of grid stability requires extensive redispatch measures, which generally increases costs, thereby reducing welfare and efficiency. However, it must be noted that using the overall costs as a proxy for market efficiency makes the absolute values sensitive to model parameterization. As noted in Section 1.1.5, the redispatch problem is different from the nodal problem by changing the objective function: A penalty for overall volume and penalty factors applied to the variable costs of generators are added. Without these penalties, the zonal problem in combination with redispatch always leads to a nodally optimal solution, which is the option with the overall lowest system costs (cf. Section 1.1.5). This implies that the zonal solution may be adjusted even where it is physically feasible, an undesirable outcome. Conversely, absolute values arrived at by using penalty factors must be treated with caution (cf. Section 7.3).

Chapters 5 and 6 introduce the Shapley value as an allocation method of the costs arising from redispatch. As already discussed in Section 7.2, an important property of the Shapley value is its fairness: Costs are distributed according to the marginal contribution of all contributors in all possible permutations, thereby taking into account synergies. This represents a major improvement on the status quo at the time of writing the articles, where parts of the redispatch costs (curtailment of renewables) were allocated to whoever first initiated the measure (cf. Section 1.1.4). In a situation where several different grid operators might benefit from a measure, this leads to obvious incentives for delaying initiation as long as possible in the hope that another grid operator moves first. This obviously disincentivizes coordinated actions and consequently decreases overall efficiency. Conversely, using the Shapley value as proposed in Chapters 5 and 6 ensures that the cost allocation mechanism does not impede an efficient overall solution.

Chapter 2 shows that the subsidies required to drive investments into renewable energy sources (RES) hinder the efficient operation of local flexibility markets, a form of market-based redispatch, by distorting prices. RES in Germany receive a guaranteed remuneration regardless of whether they feed in or are curtailed by the grid operator – in the latter case the remuneration is paid as a reimbursement for lost revenues. However, in the case of *voluntary* curtailment, as would occur in a market-based redispatch scheme, no reimbursement is granted. Therefore the bids of RES must reflect these lost revenues. This, however, leads to an inefficient solution: The guaranteed remuneration is paid in any case and should not influence the market result. Considering the remuneration in the bids increases overall costs by inflating bids far beyond true marginal costs. It can also lead to an inefficient selection of bids, because affected units may become uncompetitive compared to other units, even though marginal costs are lower.

The paper addresses this issue by proposing the introduction of side payments. Side payments are payments outside of the market in the amount of the lost subsidies when participating in the redispatch market. This enables the unit operators to exclude the subsidies from the bid calculations, thereby contributing to an efficient operation of the redispatch market.

An important factor limiting the efficiency of any market design is the effect of market power, which has not been explicitly considered in the articles forming this thesis. The assumption that total costs can serve as a proxy for efficiency, stated at the beginning of this section, rests on the further assumption that all relevant factors are considered in the model. As it stands, the statements regarding efficiency must be limited to comparisons where the effects of market power can be expected to be comparable, i.e., different designs of zonal pricing. Regarding the possible market power of RES in flexibility markets, Chapter 2 suggests the implementation of price caps to ensure that the system costs of the flexibility market do not exceed the costs under the alternative command and control strategy.

Regarding mechanism designs for the efficient operation of zonal electricity markets with congestion management, the following key findings may be derived from the articles in this thesis:

1. Artificially increasing cross-border capacities beyond the physical realities, such as practiced with the minRAM policy, can reduce the efficiency of the overall market design.

2. Adding nodal features such as the consideration of potential redispatch measures into the zonal market clearing algorithm can decrease costs (thereby increasing overall efficiency) while increasing overall trade.

3. The mechanism for allocating the costs of redispatch should not hinder the coordination of redispatch, which would make redispatch less efficient.

4. Subsidies for renewables must be considered when designing redispatch markets, for example, by introducing side payments.

Bibliography

50Hertz Transmission GmbH, Amprion GmbH, TenneT TSO GmbH, and TransnetBW GmbH (2021). *Netzentwicklungsplan Strom 2035, Version 2021*. URL: https://www.netzentwicklungsplan.de/sites/default/files/2022-11/NEP_2035_V2021_2_Entwurf_Teil1.pdf (visited on 04/09/2023).

ACER – Agency for the Cooperation of Energy Regulators (2019). *Day-ahead capacity calculation methodology of the Core capacity calculation region*. Annex 1 to ACER Decision 02-2019 on Core CCM. URL: https://acer.europa.eu/Official_documents/Acts_of_the_Agency/ANNEXESTODECISIONOFTHEAGENCYNo022019/Annex%20I%20-%20ACER%20Decision%20on%20Core%20CCM.pdf (visited on 09/29/2019).

Bjørndal, M. and K. Jørnsten (2001). "Zonal Pricing in a Deregulated Electricity Market". In: *The Energy Journal* 22.1, pp. 51–73. URL: www.jstor.org/stable/41322907 (visited on 04/03/2019).

Bjørndal, M. and K. Jørnsten (2007). "Benefits from coordinating congestion management: The Nordic power market". In: *Energy Policy* 35.3, pp. 1978–1991. DOI: 10.1016/j.enpol.2006.06.014.

Bundesnetzagentur (2022). *Bericht Netzengpassmanagement Gesamtes Jahr 2021*. URL: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/Engpassmanagement/Zahlen%20Ganzes%20Jahr2021.pdf (visited on 10/04/2022).

Bundesnetzagentur and Bundeskartellamt (2022). *Monitoringbericht 2021*. URL: https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Monitoringberichte/Monitoringbericht_Energie2021.pdf?__blob=publicationFile&v=6 (visited on 10/04/2022).

CIGRE (2014). "Benchmark Systems for Network Integration of Renewable and Distributed Energy Resources". In: *Technical Brochure, Task Force C6.04.02*.

CWE CIA WG (2020). *Congestion income allocation under Flow-Based Market Coupling.* URL: https://www.jao.eu/sites/default/files/2021-05/CWE%20FBMC%20AP%20Annex%2014.32%20CID%20report%2012%20months%20after%20DE-AT%20v20200710.pdf (visited on 02/16/2023).

E-Bridge, IAEW, and OFFIS (2014). *Moderne Verteilnetze für Deutschland (Verteilernetzstudie).* URL: https://www.bmwk.de/Redaktion/DE/Publikationen/Studien/verteilernetzstudie.pdf?__blob=publicationFile&v=1 (visited on 04/09/2023).

EEG (2017). *Gesetz für den Ausbau erneuerbarer Energien (German code for the expansion of renewable energy) from July 21st, 2014 (BGBl. I p. 1066), last modified by article 8 of the act from August 8th, 2020 (BGBl. I p. 1728).*

Eicke, A. and T. Schittekatte (2022). "Fighting the wrong battle? A critical assessment of arguments against nodal electricity prices in the European debate". In: *Energy Policy* 170, p. 113220. DOI: 10.1016/j.enpol.2022.113220.

Elia Group (2019). *Future-proofing the EU energy system towards 2030.* URL: https://www.elia.be/-/media/project/elia/shared/documents/elia-group/publications/studies-and-reports/20191212_future_proofing_eu_system_2030.pdf (visited on 03/08/2023).

ENTSO-E (2023). *Single Day-ahead Coupling.* URL: https://www.entsoe.eu/network_codes/cacm/implementation/sdac/ (visited on 04/11/2023).

EnWG (2023). *Energiewirtschaftsgesetz (German code for electricity and gas supply) from July 7th, 2005 (BGBl. I p. 1970, 3621), last modified by article 3 of the act from January 4th, 2023 (BGBl. 2023 I No. 9).*

European Comission (2015). *Comission Regulation (EU) 1222/2015 of 24 July 2015 establishing a guideline on capacity allocation and congestion management.*

Felling, T., B. Felten, P. Osinski, and C. Weber (2023). "Assessing Improved Price Zones in Europe: Flow-Based Market Coupling in Central Western Europe in Focus". In: *The Energy Journal* 44.01. DOI: 10.5547/01956574.44.6.tfel.

Felten, B., P. Osinski, T. Felling, and C. Weber (2021). "The flow-based market coupling domain - Why we can't get it right". In: *Utilities Policy* 70, p. 101136. DOI: 10.1016/j.jup.2020.101136.

Hirth, L., I. Schlecht, C. Maurer, and B. Tersteegen (2019). *Cost- or market-based? Future redispatch procurement in Germany.* URL: https://www.bmwk.de/Redaktion/EN/Publikationen/Studien/future-redispatch-procurement-in-germany.pdf?__blob=publicationFile&v=3 (visited on 01/31/2023).

Hogan, W. W. (1992). "Contract networks for electric power transmission". In: *Journal of Regulatory Economics* 4.3, pp. 211–242. DOI: 10.1007/BF00133621.

JAO – Joint Allocation Office (2022). *Launch of Flow-Based Market Coupling in the Core region enhances energy transition.* Press Release. URL: https://www.jao.eu/sites/default/files/2022-06/Core%20DA%20FB%20MC%20go-live%20press%20release.pdf (visited on 01/31/2023).

Löschel, A., V. Grimm, F. Matthes, and A. Weidlich (2023). *Stellungnahme zum Strommarktdesign und dessen Weiterentwicklungsmöglichkeiten.* Expertenkommission zum Monitoring-Prozess "Energie der Zukunft". URL: https://www2.wiwi.rub.de/wp-content/uploads/2023/02/Stellungnahme-zum-Strommarktdesign-und-dessen-Weiterentwicklungsmoeglichkeiten.pdf (visited on 02/22/2023).

Meinecke, S., D. Sarajlić, S. R. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, A. Moser, and M. Braun (2020). "SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis". In: *Energies* 13.12. DOI: 10.3390/en13123290.

NEMO Committee (2020). *Euphemia Public Description. Single Price Coupling Algorithm.* URL: https://www.nordpoolgroup.com/globalassets/download-center/single-day-ahead-coupling/euphemia-public-description.pdf (visited on 04/13/2023).

OLG Düsseldorf – Oberlandesgericht Düsseldorf (2015). *Beschluss vom 28.04.2015 - VI-3 Kart 313/12 (V).* URL: https://openjur.de/u/867722.html (visited on 04/12/2023).

Oggioni, G. and Y. Smeers (2013). “Market failures of Market Coupling and counter-trading in Europe: An illustrative model based discussion”. In: *Energy Economics* 35, pp. 74–87. DOI: 10.1016/j.eneco.2011.11.018.

Pearson, S., S. Wellnitz, P. Crespo del Granado, and N. Hashemipour (2022). “The value of TSO-DSO coordination in re-dispatch with flexible decentralized energy sources: Insights for Germany in 2030”. In: *Applied Energy* 326, p. 119905. DOI: 10.1016/j.apenergy.2022.119905.

Pfenninger, S., A. Hawkes, and J. Keirstead (2014). “Energy systems modeling for twenty-first century energy challenges”. In: *Renewable and Sustainable Energy Reviews* 33, pp. 74–86. DOI: 10.1016/j.rser.2014.02.003.

Schlecht, I. and L. Hirth (2021). *Dispatch Hub Compensation Schemes - Study on profit impacts and strategic incentives of alternative compensation schemes for Dispatch Hubs in the Flex-in-Market concept*. URL: https://neon.energy/Neon-Dispatch-Hubs-Elia.pdf (visited on 03/08/2023).

Schleswig-Holstein (2023). *Antrag des Landes Schleswig-Holstein: Entschließung des Bundesrates „Industriestandort Deutschland stärken, Produktion klimarelevanter Technologien hochfahren“*. URL: https://dserver.bundestag.de/brd/2023/0113-23.pdf (visited on 04/12/2023).

Schönheit, D., C. Dierstein, and D. Möst (2021a). “Do minimum trading capacities for the cross-zonal exchange of electricity lead to welfare losses?” In: *Energy Policy* 149, p. 112030. DOI: 10.1016/j.enpol.2020.112030.

Schönheit, D., M. Kenis, L. Lorenz, D. Möst, E. Delarue, and K. Bruninx (2021b). “Toward a fundamental understanding of flow-based market coupling for cross-border electricity trading”. In: *Advances in Applied Energy* 2, p. 100027. DOI: 10.1016/j.adapen.2021.100027.

Schönheit, D., R. Weinhold, and C. Dierstein (2020). “The impact of different strategies for generation shift keys (GSKs) on the flow-based market coupling domain: A model-based analysis of Central Western Europe”. In: *Applied Energy* 258, p. 114067. DOI: 10.1016/j.apenergy.2019.114067.

Schweppe, F. C., M. C. Caramanis, R. D. Tabors, and R. E. Bohn (1988). *Spot Pricing of Electricity.* Boston, MA: Springer US. DOI: 10.1007/978-1-4613-1683-1.

Sioshansi, R. and A. J. Conejo (2017). *Optimization in Engineering.* Berlin, Heidelberg: Springer International Publishing. DOI: 10.1007/978-3-319-56769-3.

Ventosa, M., Á. Baíllo, A. Ramos, and M. Rivier (2005). "Electricity market modeling trends". In: *Energy Policy* 33.7, pp. 897–913. DOI: 10.1016/j.enpol.2003.10.013.

Weber, C., D. Möst, and W. Fichtner (2022). *Economics of Power Systems: Fundamentals for Sustainable Energy.* Cham, Switzerland: Springer International Publishing. DOI: 10.1007/978-3-030-97770-2.

Wilson, R. (2002). "Architecture of Power Markets". In: *Econometrica* 70.4, pp. 1299–1340. DOI: 10.1111/1468-0262.00334.

Zimmerman, R. D., C. E. Murillo-Sanchez, and R. J. Thomas (2011). "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education". In: *IEE Proceedings - Generation, Transmission and Distribution* 26.1, pp. 12–19. DOI: 10.1109/TPWRS.2010.2051168.