Université catholique de Louvain
École Polytechnique de Louvain

Center for Operations Research and Econometrics

Models and Algorithms for Quantifying and Mitigating the Inefficiency of Zonal Pricing: Short and Long-Term Measures

Quentin Lété

Thesis submitted in partial fulfillment of the requirements for the degree of
Docteur en sciences de l'ingénieur et technologie

Supervisor:
Anthony Papavasiliou
(UCLouvain, Belgium)

Jury:
Yves Smeers (UCLouvain, Belgium)
Raphaël Jungers (UCLouvain, Belgium)
Mathieu Van Vyve (UCLouvain, Belgium)
Karsten Neuhoff (Technical University of Berlin, Germany)
Mette Bjørndal (Norwegian School of Economics, Norway)

Chair:
Mathieu Van Vyve (UCLouvain, Belgium)

PhD Organization

Quentin Lété
UCLouvain
École Polytechnique de Louvain
Center for Operations Research and Econometrics

Thesis Supervisor

Anthony Papavasiliou
Associate Professor, UCLouvain
École Polytechnique de Louvain
Center for Operations Research and Econometrics

Supervisory Committee

Yves Smeers
Professor Emeritus, UCLouvain
Center for Operations Research and Econometrics

Karsten Neuhoff
Professor, Technical University of Berlin
Institute for Economics and Law

Abstract

The necessity of the energy transition and its progressive implementation revealed a number of challenges for the planning and operation of electric power systems, many of which are the subject of active research. These challenges are very diverse and range from specific operational issues to market-oriented questions, essential to policy makers. Among the open issues that relate to market design, one has been the subject of particularly intense debate in Europe since the liberalization: the market-based allocation of transmission capacity.

Currently, the market is organized around the principle of zonal pricing, which is characterized by the delimitation of regions on the power network, that are called zones, inside which the electricity price is the same in the wholesale market. This principle is opposed to nodal pricing, in which the price is differentiated between every location.

In this dissertation, we present detailed models and algorithms for analyzing the efficiency of zonal pricing. The dissertation is divided into two parts.

The first part focuses on the short-term efficiency and, in particular, on the impacts of transmission switching in zonal pricing. A two-stage model of the short-term European market is presented, that accounts for transmission switching in both the day-ahead and real-time stages. We propose a cutting-plane algorithm for solving the resulting model, which belongs to the class of adaptive robust optimization problems with mixed-integer recourse, with the particularity that the inner-level problem has the structure of an interdiction game.

The second part of the dissertation investigates the efficiency of zonal pricing in the long run, when investment decisions in generating capacity are accounted for. First, we propose a model of the long-run equilibrium of zonal pricing with flow-based market coupling, which is the current methodology used in practice in Europe for allocating transmission capacity. We show that this methodology introduces new inefficiencies in the context of capacity expansion. Then, we investigate whether additional locational instruments, such as capacity and energy-based signals or market-based re-dispatch, can restore the efficiency of investments in zonal pricing.

For each of these three main contributions, we present simulation results of our models on a large-scale instance that represents the Central Western European network and we discuss the implications of our results on the efficiency of zonal electricity pricing.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Anthony Papavasiliou. I am thankful for his accessibility, his very regular and always relevant guidance and the confidence that he always placed in my ability to answer difficult questions. His support helped me to overcome much more easily the periods of doubts of which a PhD thesis is inevitably made. With the Energy group, Anthony built a true research group at CORE that triggered many interactions, both scientific and informal. I am very grateful to him for that, and for his example in putting human relationships at the center of the thought process in research.

I sincerely thank Prof. Yves Smeers for his guidance throughout the development of this work. After about one year into my PhD, Prof. Smeers became almost like a second supervisor to me and our regular interactions have had an enormous influence on my research. I thank him for his enthusiasm for the work and for his constant inspiration for identifying the right research question that came next. I also express my gratitude to Prof. Smeers for sharing his knowledge in the crash courses in economics or operations research into which our meetings were sometimes turned. I learned a lot from these interactions.

My thanks go to Prof. Karsten Neuhoff for his role as part of the doctoral supervisory committee as well as of the thesis jury. I am grateful for the relevant questions and comments that he raised at each of our interactions. I also thank him for welcoming me in Berlin for the Future Power Market Platform meeting, that allowed me to get immersed into the current policy discussions among stakeholders at the beginning of my PhD.

I would like to thank Professors Mette Bjørndal, Raphaël Jungers and Mathieu Van Vyve for agreeing to be part of the thesis jury and for the feedback they provided on my research in the process. I extend my special thanks to Prof. Bjørndal for the useful follow-up interactions, to Prof. Jungers for acting as the secretary and to Prof. Van Vyve for chairing the jury.

I am very grateful to Ignacio Aravena for giving me the opportunity to spend a research stay at the Lawrence Livermore National Laboratory. I thank him for making it such a fun time and for sharing with me some of his computing and programming skills. I learned a lot while working with Ignacio, and his research had a big influence on this thesis.

Many people at CORE made it such a nice place to work. I am grateful to the members of the Energy group, for the interesting informal discussions on optimization and electricity markets as well as the fun times at conferences and group dinners: Yuting, Gilles, Ilyès, Céline, Simone, Daniel, Loïc, Jacques, Jehum and Nicolas. Thanks to our OR and Econ friends for the great atmosphere during the numerous lunches and WABs: Henri, Antoine, Alessandro, Matteo, Sefane, Robin, Niranjana, Martial, Frank, Juan, Thomas^2.

Besides research, I warmly thank my family, friends and housemates for their support, for the great moments spent together over the past four years and the many more to come. Special thanks go to Rémy, for our discussions on the importance of Science, that helped me giving meaning to what I was doing, to Auguste, for always being eager to listen to me talking about my research and, of course, to Benjamin for his deep friendship. To Antoine and Thibault: the three of us have developed such a strong connection, I am immensely grateful for that. To my parents, Christine and Pascal, thanks for everything. I admire you for putting forward human values in your personal and professional life. You are true role models to me.

The financial support for this thesis was provided by the Fonds de la Recherche Scientifique - FNRS through a Research Fellowship.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context and motivation

Climate change is arguably one of the most important challenges that our generation is facing. This is due to the disastrous damages that it causes to our living environment, and will cause even more in the future. Moreover, the transformation of our societies that is required, should we want to tackle it seriously, is significant. In response, ambitious goals have been set by governments worldwide, and in particular within the European Union, that can be summarized by two medium and long-term targets from the European Commission: a reduction of greenhouse gas emission by 2030 to at least 55% compared to 1990 [Eur20] and a climate-neutral union (net-zero greenhouse gas emission) by 2050 [Eur18a].

Crucial to achieving these goals is the evolution of how we produce and consume energy. On the production side, the use of low-emission energy sources needs to be expanded. On the consumption side, energy efficiency has to be improved. Electricity, as an energy vector that simultaneously allows for the exploitation of renewable energy sources for production and zero-emission in consumption, has a key role to play in this energy transition. A significant part of the success of the transition will thus depend on how much and how fast one can integrate renewable resources to the electricity grid. This is accompanied by dramatic changes to electric power systems, which become more decentralized and face an increase in uncertainty and variability of production and consumption patterns.

The transformation of power systems places significant pressure on the electricity market design and in particular on how the market deals with transmission constraints, which is the focus of the present dissertation. The main characteristics of the European market design regarding transmission capacity allocation are indeed inherited from the early days of the liberalization of the sector and the creation of the internal market for electricity, in the early 2000s, a period that largely preceded the aforementioned transformations. The increasing pressure on the market design is witnessed by numerous changes of practices and regulations related to transmission capacity allocation that took place in the past few years. These changes are detailed in section 1.2, after introducing some basic notions and nomenclature related to locational electricity pricing. We should simply note at this stage that market-based transmission capacity allocation is currently an active topic of discussions among European stakeholders.

Fundamentally, part of the reason why certain questions like the ones that are the subject of this dissertation are not yet settled, more than 30 years after the first experience of deregulation in Europe (the market in England and Wales opened in 1990) is the inherent complexity of electric systems. Indeed, electricity has certain properties that raise challenges: it cannot be stored efficiently and obeys complex physical laws such as Kirchhoff's laws, which are central when dealing with transmission. When we add to that its crucial role in our economy and daily lives, and the sizes of the resulting systems (the electrical power grid is often called *the largest machine ever built*), we face the design of a system of high complexity where technological and economical considerations are intertwined.

For this reason, advanced modeling tools and algorithmic methods are required in order to study the electricity market, probably more than for many other markets. The concept of mathematical optimization, with its powerful capabilities for modeling both physical phenomena and economic agents and with its set of mature algorithms for solving diverse problem classes, stands out as a central concept for studying electricity markets. This is why optimization is ubiquitous in the analysis of electricity markets, and this work is no exception.

In this dissertation, our goal is to contribute to the ongoing policy discussions regarding transmission capacity allocation in European electricity markets, using optimization models and algorithms. In particular, the thesis is articulated around the study of zonal electricity pricing, which is the building block of the European methodology for allocating transmission capacity. Before diving into the core of the work, we propose to start by presenting some background information on zonal pricing in section 1.2 and on complementarity problems in section 1.3. Then, section 1.4 introduces the structure of this document and summarizes our contributions.

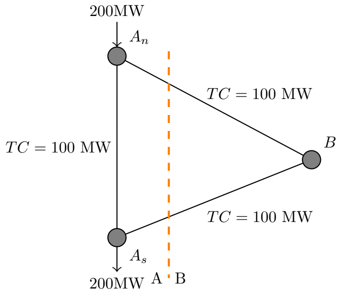


Figure 1.1: Basic three-node network used in the example of section 1.2.1.

1.2 Background on zonal electricity pricing

In this section, we start by presenting in subsection 1.2.1 some basic notions about electricity transmission that motivate the concept of zonal electricity pricing. Then, in subsection 1.2.2, we describe how zonal pricing can be formulated mathematically. In subsection 1.2.3, we summarize the evolution of transmission capacity allocation in the European electricity market and list the changes to the design that took place recently. Finally, we list in subsection 1.2.4 some open issues related to zonal pricing.

1.2.1 Basics

The value of electricity depends, by nature, on the location of its production or consumption. In fact, this is true for most commodities, as there is a cost associated to transporting a product from its production site to its consumption location. However, the locational nature of its value is particularly important for electricity due to the two following reasons that are related to its fundamental properties: (i) as electricity cannot be stored efficiently, one needs an electric power grid to transport electricity almost instantaneously from the production units to the consumption sites, which is costly to build and operate; (ii) the physical laws to which electricity abides impose some restrictions on how it can be transported in an electricity grid.

Let us illustrate this last point on a simple example. Consider a three-node transmission network where each pair of nodes is linked with a transmission line that has a capacity of 100MW, as shown in Figure 1.1. There is a generating unit with a capacity of 200MW in node A_n and a load that consumes 200MW in node A_s. If electricity would obey the same physical laws as a transportation network, it would be possible with the network of Figure 1.1 to carry all 200MW of production to the load by simply transporting 100MW via line $A_n - A_s$ and 100MW via lines $A_n - B$ and $B - A_s$. This is however not possible in an electricity grid, as the flows on the network must obey the so-called power flow equations. Using the DC approximation of these equations, which is generally deemed an acceptable representation for market models of transmission networks, the flows on the network respect the equivalent of Kirchhoff's second law: the sum of the power flows weighted by the inverse of the susceptance of the line on any closed loop of the network must be equal to zero. In our example, assuming that all lines have the same susceptance, one thus has that the sum of the flows on line $A_n - B$ and on line $B - A_s$ must equal the flow on line $A_n - A_s$. This implies that the flow on the path $A_n - B - A_s$ is actually limited to 50MW and the network cannot carry more than 150MW from A_n to A_s.

The analysis provided here is a simplification that omits two important aspects of power systems: (i) Phase Shifting Transformers (PST) and (ii) AC power flows. PSTs are specialized types of transformers that are used to control the flows on electric transmission grids. They can be used to regulate the voltage phase angle difference between two nodes of the network and, consequently, reduce the impact of Kirchhoff's laws on congestion. Although the number of PSTs installed in the European transmission network is expected to rise in the coming years, the equipment remains costly, which prevents it from being installed on every transmission line[1]. Moreover, the electricity network obeys alternating current power flows. In the analysis of the illustrative example presented above, as well throughout the present thesis, we assume that the power flow equations can be accurately described by the DC approximation of the power flow equations. These simplified equations are commonly used in technical-economic studies of electric power systems because of their simplicity and generally low approximation errors when applied on static analyses of high-voltage transmission networks. The DC approximation, however, neglects reactive power management, voltage stability and transmission losses, and one must be careful with its validity outside normal operating conditions. The reader is referred to [PMVDB05] and references therein for further details on the DC approximation and its validity.

Following basic economic principles, the locational nature of the value of electricity implies that its price should be differentiated among locations, i.e. among the nodes of the transportation network. In our illustrative example, assume that there is also a generating unit with a capacity of 200MW in node A_s, but that its cost of production is higher than that of the unit in node A_n. Let us say, for instance, that the generator in A_n is a wind farm that has zero marginal cost of production and that the generator in A_s is a gas unit that produces at 100€/MWh. The price of electricity in one location corresponds to the increase in the total production cost of the system when the demand marginally increases in that location. In A_n, the price should thus be 0€/MWh, as an increase of demand in A_n can be served by the unused capacity of wind generators, that have zero marginal cost. An additional MW of demand in node A_s, however, cannot be served by the wind generators due to transmission constraints. The price in that node should thus be equal to the

[1] See `https://www.entsoe.eu/Technopedia/techsheets/phase-shifting-transformers` for further specifications of PSTs in the context of the European transmission grid.

marginal cost of the gas generator, i.e. 100€/MWh. Such an electricity market, that differentiates the price of electricity among all nodes of the transmission network, is said to implement *nodal pricing*.

In contrast, the European electricity market follows a different principle, called *zonal pricing*. In zonal pricing, the nodes are aggregated into a set of zones, and the prices are only differentiated among zones. In our example, let us assume that nodes A_n and A_s are grouped into a single zone, that we call A. The prices between A_n and A_s cannot be differentiated anymore. There is a single price in zone A (which in our example can be anything between 0 and 100€/MWh), and a price in zone B.

Crucially, zonal pricing introduces ambiguity about how the transmission constraints should be formulated mathematically. This is what we discuss in the next section.

1.2.2 Mathematical formulations

As we want to integrate transmission constraints into economic models that study the electricity market, the mathematical object that we are interested in is the set of power injections and withdrawals in the different locations that the network can support. In nodal pricing, this set has as many dimensions as the number of nodes in the transmission network. The set of feasible nodal net injections under the DC approximation, that we denote as $\mathcal{R}$, can be written mathematically as follows:

$$\mathcal{R} = \left\{ r \in \mathbb{R}^{|N|} \;\middle|\; \exists (f,\theta) \in \mathbb{R}^{|L|} \times \mathbb{R}^{|N|} : \sum_{l \in L(n,\cdot)} f_l - \sum_{l \in L(\cdot,n)} f_l = r_n \quad \forall n \in N, \right.$$
$$\left. -F_l \leq f_l \leq F_l,\; f_l = B_l \left(\theta_{m(l)} - \theta_{n(l)}\right) \quad \forall l \in L \right\} \tag{1.1}$$

In this equation, r_n is the net injection in node $n \in N$, f_l is the power flow on line $l \in L$, F_l is the transmission limit on line l, B_l is the susceptance of line l and θ_n is the voltage angle at node n. Additionally, $m(l)$ and $n(l)$ denote the adjacent nodes of line l, respectively in the outgoing and incoming direction, and $L(m,n)$ is the set of lines from node m to node n. Note that this set is unambiguous: it uniquely represents the set of feasible nodal net injections under the DC approximation. In particular, the formulation of equation (1.1) is equivalent to the formulation with Kirchhoff's second law that we mentioned in the previous section.

The same cannot be said, however, for zonal pricing. The fundamental set that should be defined in zonal pricing is the set of zonal net positions that are feasible for the network, that we denote by $\mathcal{P}$, analogous to $\mathcal{R}$. There exist two main ways of defining $\mathcal{P}$ that we describe successively below: (i) Available Transfer Capacity Market Coupling (ATCMC) and (ii) Flow-Based Market Coupling (FBMC).

Available Transfer Capacity Market Coupling In zonal pricing based on ATCMC, one defines a set of interconnectors which correspond to a single fictitious transmission line for every pair of zones that are connected in the initial network. A maximum capacity is defined for every interconnector and a simple transportation model is used on the network of zones linked by interconnectors. The feasible set of zonal net positions under this model can be written as follows:

$$\mathcal{P}^{\text{ATCMC}} = \left\{ p \in \mathbb{R}^{|Z|} \;\middle|\; \exists f \in \mathbb{R}^{|A|} : \sum_{a \in A(z,\cdot)} f_a - \sum_{a \in A(\cdot,z)} f_a = p_z \quad \forall z \in Z, \right.$$
$$\left. - ATC_a \leq f_a \leq ATC_a \quad \forall a \in A \right\} \tag{1.2}$$

where p_z is the net position of zone $z \in Z$, f_a is the flow on interconnector $a \in A$ and ATC_a is the capacity of interconnector a. The ATCMC methodology implicitly assumes that the exchanges between each pair of zones are independent, which is not the case in reality. A large export of electricity between, for instance, Germany and France will increase the loading in the Belgian network, which in practice reduces the export capacity of Germany to Belgium. The FBMC methodology has been proposed in order to account for the dependence of exchanges between connected zones.

Flow-Based Market Coupling In FBMC, one identifies a set of network elements (transmission lines and transformers) that are highly impacted by interzonal exchanges. These network elements are referred to as critical branches and are denoted by CB. Two parameters are then defined: (i) the remaining available margin, that we denote by RAM_{cb}, a parameter that evaluates the capacity of critical branch $cb \in CB$ that can be used for cross-border exchanges and (ii) the power transfer distribution factor of line cb with respect to zone z, denoted by $PTDF_{cb,z}$, which measures the expected change in power flow on line cb resulting from a unit increase in the net position of zone z. Using these two parameters, one can write the feasible set of zonal net positions in FBMC as follows:

$$\mathcal{P}^{FBMC} = \left\{ p \in \mathbb{R}^{|Z|} \;\middle|\; \sum_{z \in Z} p_z = 0, \; \sum_{z \in Z} PTDF_{cb,z} \cdot p_z \leq RAM_{cb}, \; \forall cb \in CB \right\} \tag{1.3}$$

The method is described in further detail in the literature [VBD16], as well as in documentation published by the European transmission system operators (TSOs) [Eur18b]. FBMC is the methodology currently used in a large part of Europe for defining the feasible set of zonal net positions. As noted in [ALPS21], the following challenges emerge when modeling FBMC for the purpose of policy analysis:

1. The flow-based polytope $\mathcal{P}^{FBMC}$ is characterized by parameters (CB, $PTDF_{cb,z}$, RAM_{cb}), the definition of which differs among TSOs. This makes it difficult to represent exactly current TSO practice, particularly since the market clearing outcome is sensitive to these parameters [MLTW13].

2. The method used in practice faces a circularity problem: the parameters that are used for defining the flow-based domain are computed from a forecast of the outcome of the market, which depends itself on the value of these parameters.

Instead of building a model that attempts to replicate current practice, [ALPS21] proposes a definition of the flow-based domain that does not depend on arbitrary parameters. The idea of [ALPS21] is to define the flow-based polytope as the projection of the set of feasible nodal net injections, given a forecast of the demand and the existing capacity in every node, onto the space of zonal net positions. This model is called *FBMC with exact projection* and its polytope is denoted by $\mathcal{P}^{\text{FBMC-EP}}$. Mathematically, it can be formulated as follows:

$$\begin{aligned}\mathcal{P}^{FBMC-EP} = \Bigg\{ p \in \mathbb{R}^{|Z|} \,\Bigg|\, & \exists(\bar{v}, f, \theta) \in [0,1]^{|G|} \times \mathbb{R}^{|L|} \times \mathbb{R}^{|N|} : \\ & \sum_{g \in G(z)} Q_g \bar{v}_g - p_z = \sum_{n \in N(z)} Q_n \quad \forall z \in Z, \\ & \sum_{g \in G(n)} Q_g \bar{v}_g - \sum_{l \in L(n,\cdot)} f_l + \sum_{l \in L(\cdot,n)} f_l = Q_n \quad \forall n \in N, \\ & -F_l \leq f_l \leq F_l,\ f_l = B_l \left(\theta_{m(l)} - \theta_{n(l)}\right) \quad \forall l \in L \Bigg\}\end{aligned} \tag{1.4}$$

In this equation, Q_g corresponds to the capacity of generator $g \in G$; $G(n)$ (respectively $G(z)$) is the set of generators at node n (respectively zone z); Q_n is the forecast demand at node $n \in N$; $N(z)$ is the set of nodes that belong to zone z and $\bar{v}_g$ is the fraction of the capacity of generator g that is used for producing. This set includes all zonal net positions for which there is at least one set of generator output levels that are feasible under the full network model. In this way, FBMC with exact projection allows all trades that are feasible with respect to the real network to be cleared, in compliance with legislation [Eur09], and ensures that no trade that provably leads to a violation of transmission constraints will be cleared[2]. The reader is referred to [ALPS21] for a more detailed description of $\mathcal{P}^{FBMC-EP}$, as well as its implications on zonal market clearing.

[2] Annex I of [Eur09] establishes that "... *TSOs shall endeavour to accept all commercial transactions, including those involving cross-border-trade* ..." (Article 1.1) and that "... *TSOs shall not limit interconnection capacity in order to solve congestion inside their own control area, save for the abovementioned reasons and reasons of operational security* ..." (Article 1.7).

Nodal	Zonal
$\min_{v\in[0,1],r} \sum_{g\in G} P_g Q_g v_g$ s.t. $\sum_{g\in G(n)} Q_g v_g - r_n = Q_n \, [\rho_n]$ $r \in \mathcal{R}$	$\min_{v\in[0,1],p} \sum_{g\in G} P_g Q_g v_g$ s.t. $\sum_{g\in G(z)} Q_g v_g - p_z = \sum_{n\in N(z)} Q_n \, [\rho_z]$ $p \in \mathcal{P}$

Table 1.1: A comparison of nodal and zonal economic dispatch problems. Variable between brackets denote dual variables of the associated constraint. In this case, ρ_n and ρ_z represent, therefore, respectively the nodal and zonal electricity prices.

Nodal and zonal economic dispatch In their simplest form, nodal and zonal electricity markets can be described by the so-called *economic dispatch problem*. Given price-quantity bids (P_g, Q_g) by generators $g \in G$, its goal is to determine the acceptance level v_g of each bid and prices (nodal ρ_n or zonal ρ_z) at the competitive equilibrium. Using the feasible sets $\mathcal{R}$ and $\mathcal{P}$ introduced in this section, we can write the economic dispatch problem as in Table 1.1. As can be observed in this comparison, the main difference between the two problems lies in the variable that controls power balance. Whereas a nodal economic dispatch problem decides on the nodal net injections $r_n \in \mathbb{R}^{|N|}$ so that power balance is enforced at every node of the network, power is balanced at the zonal level in zonal pricing, using the zonal net positions $p_z \in \mathbb{R}^{|Z|}$ as the main control variables.

One may wonder whether there is any relationship between the feasible sets of acceptance/rejection of bids of the two designs. Let us define these sets formally as follows:

$$\begin{aligned}\mathcal{V}^{\text{Nodal}} = \{v \in [0,1]^{|G|} \;\mid\; &\exists r \in \mathbb{R}^{|N|} : \\ &\sum_{g\in G(n)} Q_g v_g - r_n = Q_n, \quad r \in \mathcal{R}\}\end{aligned}$$

$$\begin{aligned}\mathcal{V}^{\text{FBMC}} = \{v \in [0,1]^{|G|} \;\mid\; &\exists p \in \mathbb{R}^{|Z|} : \\ &\sum_{g\in G(z)} Q_g v_g - p_z = \sum_{n\in N(z)} Q_n, \quad p \in \mathcal{P}^{\text{FBMC}}\}\end{aligned}$$

$$\begin{aligned}\mathcal{V}^{\text{FBMC-EP}} = \{v \in [0,1]^{|G|} \;\mid\; &\exists p \in \mathbb{R}^{|Z|} : \\ &\sum_{g\in G(z)} Q_g v_g - p_z = \sum_{n\in N(z)} Q_n, \quad p \in \mathcal{P}^{\text{FBMC-EP}}\}\end{aligned}$$

In general, there is no relationship between the feasible sets that correspond to the nodal and zonal designs. A priori, it is thus not possible to conclude that the feasible set of a zonal design is a relaxation of that of a nodal design, or vice-versa. In the specific case of zonal with FBMC-EP, however, the following relationship holds:

$$\mathcal{V}^{\text{Nodal}} \subseteq \mathcal{V}^{\text{FBMC-EP}}$$

That is, the feasible set of flow-based market coupling with the exact projection methodology is a relaxation of the feasible set of nodal pricing.

This dissertation contributes to the development of mathematical formulations of zonal pricing in three different ways:

1. The set $\mathcal{P}^{FBMC-EP}$ is extended in order to account for transmission line switching in chapter 2.

2. We present a new formulation for the feasible set of zonal net positions that is based on the fundamental principle of zonal pricing in chapter 3 and we discuss its advantage for the efficiency of long-term equilibrium in zonal pricing.

3. We propose a formulation that accounts for the new developments of European legislation regarding transmission capacity allocation in chapter 4.

1.2.3 Evolution of transmission capacity allocation

We provide here a brief description of the evolution of regulations and practices related to transmission capacity allocation in Europe. This summary, largely based on [Mee20], as well as on some of our published papers [ALPS21, LSP22], demonstrates that transmission capacity allocation is a very active topic among stakeholders.

The European electricity market developed in line with the objective of the Single Market enacted in 1986 in the Single European Act, at the time of the European Community. The first significant step towards a single electricity market was made with Regulation (EC) No 1228/2003 which required the adoption of a market-based approach for transmission capacity allocation. The zonal model emerged as the preferred model as it enabled an alignment between bidding zones and national borders. This was politically appealing and was seen as technically justified: national networks tend to be more developed than international connections and it was expected that congestion would appear mainly at the borders. This correspondence between bidding zones and member states still remains today, with a few exceptions, as illustrated on Figure 1.2.

Explicit auctioning was first considered in reaction to this regulation, but was quickly abandoned for implicit auctioning due to its inefficiency. The Nordic countries (Denmark, Finland, Norway and Sweden) were the first to implement implicit auctioning using a methodology that is referred to as *market splitting*. The idea is that an algorithm clears the market with a single price for the entire area. If the resulting power flows violate some constraints, the area is split into smaller zones with a different price in each zone. In Western Europe,

Figure 1.2: Delimitation of the different bidding zones in the European electricity market. Bidding zones are mostly aligned with national borders, with exceptions in Norway, Sweden, Denmark, Ireland and Italy. Source: [Mee20].

the first initiative was taken by Belgium, France and The Netherlands. These member states developed a methodology called trilateral *market coupling*, that was launched in 2006. The idea was similar to the Nordic method but the market clearing algorithm worked in one step and directly computed one price per zone. Since then, the market coupling initiative has grown. Today, 98,6% of EU consumption is coupled [ENT22].

The trilateral market coupling initiative adopted ATCMC as the preferred transmission capacity allocation methodology. For the reasons mentioned in section 1.2.2, FBMC is claimed to be more efficient. It is now defined as the default methodology in the Capacity Allocation and Congestion Management (CACM) guideline from the Commission (Commission Regulation (EU) 2015/1222), although ATCMC is still allowed. The first initiative of moving to FBMC was taken by the countries of the Central Western European (CWE) network area that includes Austria, Belgium, France, Germany, Luxembourg and The Netherlands, with a go-live in May 2015. The FBMC methodology has been extended in June 2022 to seven countries in central Europe and it is expected that the Nordic region will also follow this transition in the near future.

Although FBMC is claimed to be more efficient, it is often noted that this comes at the cost of higher complexity, with risks of undermining transparency [VBD16, Aus20]. This is due in part to the fact that TSOs have a lot of freedom in how they compute the FBMC parameters and that the market outcome tends to be sensitive to these parameters [MLTW13]. As noted in [Mee20], it has for instance been observed that some TSOs tends to limit cross-border capacity in order to decrease congestion management costs. As a result, the European

Commission introduced a new rule for capacity calculation in FBMC in the recent recast of the electricity regulation (Regulation (EU) 2019/943). This rule states that the RAM of a critical line should be equal to at least 70% of its total capacity.

In addition to a constant improvement of the capacity calculation methodologies themselves, an important tool that can be used for increasing the efficiency of capacity allocation in Europe is the modification of the delimitation of bidding zones. In this spirit, the CACM regulation instructs the Agency for the Cooperation of Energy Regulators (ACER) to assess the efficiency of current bidding zone configuration every three years. The first review was published in 2018 and resulted in a splitting of the German-Austrian bidding zone, while the second review is ongoing.

1.2.4 Inefficiencies of zonal pricing

The opinion about zonal pricing among EU stakeholders is quite divided. On the one hand, certain TSOs believe that a transition to nodal pricing would be beneficial. This is the case of the Polish TSO, which currently investigates the possibility of a transition to nodal pricing inside its own bidding zone, and of the British TSO that, in a recent study aiming at assessing the fitness of the British market design for the energy transition, advocated in favor of nodal pricing [Nat22]. On the other hand, nodal pricing is largely contested for various reasons by many stakeholders, as thoroughly discussed in [ES22].

The situation is however quite different among academics, where an almost complete consensus in favor of nodal pricing has developed over the years. Many papers in the scientific literature have focused on describing the inefficiencies associated to zonal pricing and arguing in favor of nodal pricing in order to mitigate them. We present here a summary of the types of inefficiencies that have been identified.

The first papers on the subject in the European literature clearly identified the difficulties associated to correctly defining the basic parameters for the zonal design like zone delimitation [BJ01] and the capacities of interconnectors [ES05]. Inefficiencies related to the distortion of the locational signal for investment were first studied in [DW11] using a stylized 2-node network with one congested transmission line. [HL15] also discuss the distortion of the long-term signal. They show that equivalence in terms of efficiency between nodal and zonal pricing can only be achieved under restrictive assumptions in the short term but that this comes at the cost of decreasing long-term efficiency. [NHN11] describe the distortion of incentives associated to zonal pricing. This distortion applies both for the TSO in terms of defining available capacity and also for market participants in terms of participating truthfully in re-dispatch. [OS13] show that the splitting of grid operations among different TSOs and their incomplete coordination in re-dispatch and countertrading also undermine overall efficiency. Finally, inefficiencies associated to sub-optimal unit commitment decisions are described in [vdWH11, AP17, ALPS21].

The present dissertation also contributes to this literature in chapter 3: we identify a new type of inefficiency of zonal pricing in the long term, due to the interactions between zonal transmission constraints and capacity investment.

1.3 Background on complementarity problems

An important mathematical concept used in this thesis is that of complementarity problems. These problems, than can be viewed as a generalization of continuous optimization problems on convex sets, provide a natural framework for analyzing a large series of equilibrium problems, among which economic equilibrium on the electricity market. In this section, we provide some definitions and known results on complementarity problems that will be useful throughout the thesis.

Definition 1.1. *[FP03, Definition 1.1.2] Given a cone K and a mapping $F : K \to \mathbb{R}^n$, the complementarity problem, denoted $CP(K, F)$, is to find a vector $x \in \mathbb{R}^n$ satisfying the following conditions:*

$$K \ni x \perp F(x) \in K^*,$$

where the notation $\perp$ means "perpendicular" and K^* is the dual cone *of K defined as:*

$$K^* \equiv \{d \in \mathbb{R}^n : v^\top d \geq 0 \quad \forall v \in K\};$$

that is, K^* consists of all vectors that make a non-obtuse angle with all the vectors in K.

Definition 1.2. *A mixed complementarity problem (MCP) corresponds to a special case of the complementarity problem $CP(K, F)$ where K is the special cone $\mathbb{R}_+^{n_1} \times \mathbb{R}^{n_2}$, with $n_1 + n_2 = n$.*

Definition 1.3. *A mixed linear complementarity problem (MLCP) corresponds to a special case of an $MCP(K, F)$ where F is an affine mapping.*

Definition 1.4. *A linear complementarity problem (LCP) corresponds to a special case of the complementarity problem $CP(K, F)$ where K is the special cone $\mathbb{R}_+^n$ and F is an affine mapping. Let the mapping F be defined by*

$$F(x) \equiv Mx + q, \quad \forall x \in \mathbb{R}^n,$$

for some vector $q \in \mathbb{R}^n$ and some matrix $M \in \mathbb{R}^{n \times n}$. The $LCP(q, M)$ is thus defined as the problem of finding $x^* \in \mathbb{R}^n$ such that

$$0 \leq x^* \perp Mx^* + q \geq 0$$

Transformation Note that it is always possible to convert an MLCP into an LCP in an extended space. Indeed, consider the following MLCP defined on $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}_+^{n_1} \times \mathbb{R}^{n_2}$:

$$\begin{aligned} 0 \le x_1 &\perp M_{11}x_1 + M_{12}x_2 + q_1 \ge 0 \\ x_2 \text{ free } &\perp M_{21}x_1 + M_{22}x_2 + q_2 = 0 \end{aligned}$$

Then, defining $x_2^+ = \max(0, x_2)$ and $x_2^- = \min(0, x_2)$, one can check that it is equivalent to the following *LCP* defined on the extended space $\begin{bmatrix} x_1 \\ x_2^+ \\ x_2^- \end{bmatrix} \in \mathbb{R}_+^{n_1+2n_2}$:

$$\begin{aligned} 0 \le x_1 &\perp M_{11}x_1 + M_{12}x_2^+ - M_{12}x_2^- + q_1 \ge 0 \\ 0 \le x_2^+ &\perp M_{21}x_1 + M_{22}x_2^+ - M_{22}x_2^- + q_2 \ge 0 \\ 0 \le x_2^- &\perp -M_{21}x_1 - M_{22}x_2^+ + M_{22}x_2^- - q_2 \ge 0 \end{aligned}$$

Existence In order to prove existence of solutions to our models, we will in this thesis rely recurrently on an existence theorem that applies to copositive matrices.

Definition 1.5. *A matrix $M \in \mathbb{R}^{n\times n}$ is said to be* ***copositive*** *if $x^\top Mx \ge 0$ for all $x \in \mathbb{R}_+^n$*

Theorem 1.1. *[CPS09, Theorem 3.8.6] Let $M \in \mathbb{R}^{n\times n}$ be copositive and let $q \in \mathbb{R}^n$ be given. If the implication*

$$[v \ge 0, Mv \ge 0, v^\top Mv = 0] \Rightarrow [v^\top q \ge 0]$$

is valid, then the $LCP(q, M)$ has a solution.

In other words, Theorem 1.1 states that a sufficient condition for existence of a solution is the positivity of $v^\top q$ for all solutions v of the homogeneous counterpart of the $LCP(q, M)$, when M is copositive.

1.4 Structure of the dissertation and contributions

The overarching goal of this dissertation is to contribute to the ongoing policy discussions regarding transmission capacity allocation in the European electricity market. The dissertation is structured into three chapters grouped into two parts. The first part focuses on short-term models of the electricity market and contains chapter 2. In the second part, we focus on the efficiency of zonal pricing in the long term, when generation capacity investment is introduced. This second part contains chapters 3 and 4.

Part I - Short-term efficiency of zonal pricing

- *Chapter 2: Impacts of transmission switching in zonal pricing*
  The physical laws to which electricity abides in a transmission network imply a seeming paradox: the total costs of the system can sometimes be reduced by disconnecting one or more transmission lines. This brings additional flexibility to the system operator who can optimize the topology in order to manage congestion. It has sometimes been claimed by certain stakeholders that markets that implement zonal pricing are better suited than those with nodal pricing for taking advantage of this additional flexibility. In this chapter, our goal is to assess this claim by proposing a model of zonal pricing that accounts for transmission line switching. We then perform simulations of both nodal and zonal pricing models with transmission switching on a large-scale instance calibrated against the CWE system and comment on the impacts of transmission switching in zonal pricing. This chapter is based on the following publications:

  ◇ Q. Lété, A. Papavasiliou, Impacts of Transmission Switching in Zonal Electricity Markets - Part I, *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 902-913, March 2021

  ◇ Q. Lété, A. Papavasiliou, Impacts of Transmission Switching in Zonal Electricity Markets - Part II, *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 914-922, March 2021

Part II - Long-term efficiency of zonal pricing

- *Chapter 3: An analysis of zonal pricing from a long-term perspective*
  In an era of energy transition, it is crucial to ensure that the design of the short-term electricity market provides sufficient cash flows to producers so as to allow the investment of the right technology at the right location. In this chapter, we revisit the question of capacity allocation in zonal markets from a long-term perspective. We model the capacity expansion problem in zonal markets with FBMC and demonstrate that the classical result of equivalence between centralized and decentralized formulations in transmission-constrained markets ceases to hold in this case. We then perform simulations of the capacity expansion problem with nodal pricing and three variations of zonal pricing and compare these different policies based on their long-term efficiency. The results of this chapter have been published as follows:

  ◇ Q. Lété, Y. Smeers, A. Papavasiliou, An analysis of zonal electricity pricing from a long-term perspective, *Energy Economics*, Volume 107, March 2022, 105853

- *Chapter 4: Locational instruments for efficient power generation investment under zonal pricing*

In this chapter, we abstract from the dichotomy of nodal vs zonal pricing. Instead, we take zonal pricing as a given, and investigate whether long-run efficiency could be restored by means of additional market-based instruments. Three types of instruments are considered: capacity, energy and re-dispatch markets. We formulate the long-run economic equilibrium under these three policies under a unifying modeling framework and compare their performance, both from a theoretical as well as empirical perspective. We find that theoretical conditions under which efficiency would be restored exist for all three policies, but that the strictness of these conditions renders their practical implementation difficult and therefore, a full recovery of efficiency unlikely. The results described in this chapter are the focus of the following work:

◇ Q. Lété, Y. Smeers, A. Papavasiliou, Locational instruments for efficient power generation investment under zonal pricing, *Working Paper*

Part I

Short-term efficiency of zonal pricing

2 Impacts of transmission switching in zonal pricing

2.1 Introduction

Transmission switching refers to the possibility for the system operator to disconnect lines in the electricity grid if it can help to improve operations. The fact that disconnecting a line in an electricity grid can improve operations may seem counter-intuitive at first glance. In fact, this is not specific to electricity networks. Dietrich Braess, a German mathematician, found that adding a road to a road network can increase the overall journey time. This result is now known as Braess' paradox. It has been shown that the apparent paradox in electric power systems has a similar interpretation as Braess' paradox in traffic equilibria [BJ08, BJ05]. The idea is that power flows in an electric grid emerge from an equilibrium situation where the total reactive power losses are minimized, just as traffic flows in a road network emerge from the equilibrium situation where each agent minimizes its travel time. The intuition of why removing lines can increase operational efficiency in power systems scheduling is that the flow of power cannot be controlled arbitrarily, but is subject to physical laws that are summarized in Kirchhoff's laws, which give rise to the power flow equations. Switching the lines of a network affords operators a certain degree of control on how the power flow equations are expressed, and on the resulting configuration of flows in an electric power network.

With the increasingly important integration of renewable energy into the electricity grid, the need for deploying tools to actively manage the network is greater than ever. As an additional source of flexibility for the system operator, transmission switching has received considerable attention by the research community. O'Neill *et al.* formulated the problem of employing transmission switching in order to improve operational efficiency as a mathematical optimization problem in [OBH$^+$05], and dubbed the term *optimal transmission switching* (OTS). Fisher *et al.* [FOF08] demonstrate how this problem can be formulated as a mixed-integer linear program if we consider the DC approximation of power flow. They find that transmission switching can achieve a cost reduction of 25% on a test network with 118 nodes. Hedman *et al.* [HOFO10] focus on quantifying these gains when the N-1 security criterion is accounted for. They observe that the cost reduction reduces to 16% on the same network when the N-1 security criterion is taken into account, which remains a significant improvement.

These promising early results, emanating from the US research community, were not followed by a large-scale adoption in the US electricity market, where the use of transmission switching remains anecdotal. This is in contrast with the European market, where transmission switching is used more extensively. A recurrent argument for explaining this difference that is often put forward among EU stakeholders is that zonal pricing is better suited for taking advantage of transmission switching [RTE19, ENT21]. The reasoning is that transmission constraints impose a more important computational burden to nodal market clearing algorithms than their zonal counterpart, hindering the implementation of topological optimization. Moreover, it is claimed that the efficiency of zonal pricing can be significantly improved by considering this possibility. Our interest in this chapter is to provide a quantitative framework for substantiating this argument.

The chapter presents a modeling and algorithmic framework for analyzing zonal markets with switching, and also develops a policy analysis using a simulation model of the Central Western European (CWE) market. The chapter is organized as follows. Section 2.2 reviews the literature concerning the modeling of short-term electricity markets and the algorithmic approaches for solving the class of problems that emerge from zonal modeling. Section 2.3 presents the two-stage models of zonal electricity markets which are considered in the present analysis. In section 2.4, we present the algorithm that we have developed for solving the day-ahead market-clearing model with proactive switching. First, we show how this problem can be modeled as an adaptive robust optimization problem with mixed integer recourse (AROMIP). Then, we develop a new approach for solving the adversarial max-min subproblem corresponding to robustness to N-1 contingencies, and we show how this algorithm can be inserted into a column-and-constraint generation procedure for solving the AROMIP. Section 3.5 contains our case study of the CWE instance. Finally, section 2.6 concludes the chapter.

2.2 Literature review

2.2.1 Transmission switching

This early literature on transmission switching mentioned in the introduction considered a *dispatch* model. Therefore, the authors only focused on the benefits of switching as an option to be considered in *dispatch*, where commitment decisions are assumed to be fixed. These studies are not directly relevant in the context of our work which is focused on the interaction of switching with unit commitment.

It is only in subsequent work [HFO$^+$10] that the interactions of switching with day-ahead unit commitment were considered in the literature. However, these models are also not directly applicable to our context, since they represent a nodal transmission model. Instead, the novelty of our work is in considering a zonal transmission model, which is the predominant network model that is used in European day-ahead market clearing. The introduction of the zonal network representation, and its interaction with unit commitment, introduces a host of modeling and computational challenges, that are the focus of the present chapter.

As a combinatorial problem, OTS is associated to considerable computational challenges. Concretely, Hedman et al. [HFO$^+$10] point out that, due to big-M constraints, the Linear Programming (LP) relaxation of the problem is very weak. This implies that solvers struggle in providing tight lower bounds, thereby delaying the convergence of branch and bound algorithms significantly. Different approaches have been employed in the literature in order to cope with this computational complexity. Fisher et al. [FOF08] restrict the number of lines that can be switched off, and notice that most of the potential benefits of transmission switching can already be achieved with a limited switching budget, while the solving time is considerably reduced. Other authors have developed heuristic methods in order to obtain a high quality solution within a short amount of run time. In the work of Barrows et al. [BBB12], Fuller et al. [FRC12], and Wu and Cheung [WC13] the common idea is to resort to pre-processing in order to identify, a priori, the potential benefit of disconnecting each line in the network. This information is then used in order to solve transmission switching models where the switching actions are restricted to the most promising lines. More recently, polyhedral studies of the OTS problem have been developed. Here, a notable contribution is proposed by Kocuk et al. [KJD$^+$16], where the authors derive a cycle-based formulation of the OTS problem. The authors use this formulation in order to derive valid inequalities that are shown to strengthen the big-M formulation, thereby decreasing the computational time for solving the problem within a certain optimality gap.

2.2.2 Modeling of the European electricity market

In addition to the different representation of network constraints in the day-ahead market, the European design differs with respect to several other aspects from the US design. Accounting for these differences requires a different modeling set-up. This chapter build off of recent research on developing a precise model of the day-ahead and real-time operation of the European electricity market. Aravena and Papavasiliou [AP17] develop a hierarchy of European electricity market models that are targeted at accounting for unit commitment and the separation between energy and reserves. Based on this work, Han and Papavasiliou [HP16] develop a model of the European market that accounts for transmission switching. Although their model is highly simplified (the representation of day-ahead flow scheduling is inaccurate, security criteria are not

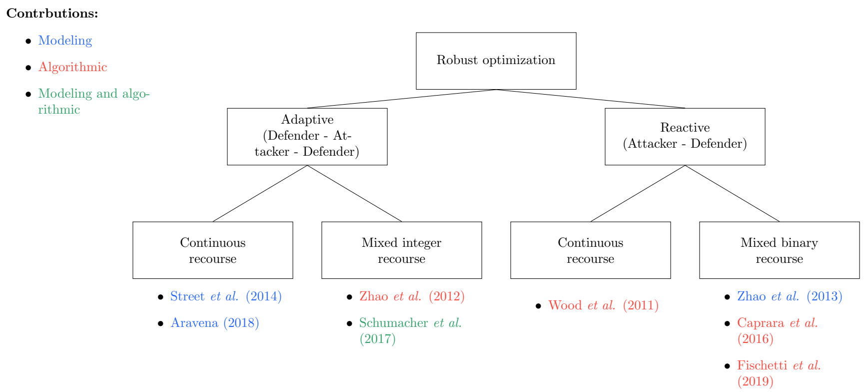


Figure 2.1: Decomposition of robust optimization into the different subproblems and the papers associated to each case. We also use a color code to distinguish the modeling and algorithmic contributions.

accounted for directly, the simultaneous optimization of unit commitment and switching is treated heuristically), this first analysis demonstrates encouraging results by demonstrating that transmission switching can lead to significant cost savings in a zonal market when re-dispatch and balancing are perfectly coordinated in real time.

2.2.3 Robust optimization

Accounting for the N-1 security criterion introduces significant complexity to the problem at hand. We review here some previous work that relates to our problem, either from a modeling or from an algorithmic point of view. The goal is not to be exhaustive. Sun and Lorca [SL17] provide an in-depth review of modeling and algorithmic approaches for robust optimization in power systems. A common attribute of the papers cited below is that they tackle a certain class of robust optimization problems. We distinguish models and algorithms for adaptive and reactive robust optimization. In both cases, the recourse problem can be continuous or include integer restrictions. This structure is represented in Figure 2.1.

Street et al. [SMA14] propose a formulation of the optimal power flow problem with N-k robustness as an Adaptive Robust Optimization (ARO) problem and present a cutting plane algorithm for solving it. The algorithm is inspired by dual methods for ARO that rely on Benders decomposition. Similarly, Aravena et al. [ALPS21] propose a cutting plane approach for solving the day-ahead market clearing problem of a zonal market clearing model, which respects the N-1 security criterion and the European rules for setting day-ahead prices. A new layer of complexity is added to these two models when we consider transmission switching as a recourse action. In this setting, the aforementioned approaches that rely on Benders decomposition cannot be used due to the presence of integer variables in the inner problem. An algorithm for attacking this class of problems has been proposed recently by Zhao and Zeng [ZZ12]. Their idea is to consider only a subset of the possible integer values in the inner problem, which results in an LP formulation that can be dualized. Promising candidate integer values are identified and added in a sequential manner. This approach has been applied to the optimal power flow problem with line interdiction [ZZ13b]. In a similar spirit, Schumacher et al. [SCC17] also generate switching variables as needed, in order to solve the N-2 security-constrained unit commitment problem with transmission switching.

In parallel to this work, new approaches have been developed in the literature on interdiction games for solving robust optimization problems with binary uncertainty and mixed integer recourse. Interdiction games are games between agents, grouped into a set of leaders and a set of followers, who have access to a set of common resources and where the leaders can deny access to certain resources to the followers. In a survey paper on the subject, Wood presents a cutting plane algorithm for solving interdiction games with continuous linear recourse [Kev11]. This approach has been employed by Caprara et al. [CCLW16] for solving the bi-level knapsack with interdiction constraints, which is a particular instance of an interdiction game with binary recourse. This algorithm has been extended by Fischetti et al. [FLMS19] to a particular class of interdiction games with mixed integer recourse, i.e. those with the property of monotonicity. Most network interdiction games, however, do not satisfy the property of monotonicity. The line interdiction game with transmission switching, which is the problem that we are interested in, is one of them. To the best of our knowledge, no efficient algorithm based on this approach has been proposed for solving the problem.

Our work combines the modeling and computational literature cited above by formulating the problem of day-ahead N-1 market clearing as an ARO with mixed integer recourse. We use the column-and-constraint generation algorithm of Zhao and Zeng for solving the outer loop. For the inner loop, that corresponds to an interdiction game with mixed-integer recourse, we present a new approach based on the cutting plane formulation used in interdiction games, that leverages its specific structure.

2.2.4 Contributions

The present chapter provides the following contributions to the literature:

1. We present a model for the organization of a zonal market that accounts for transmission switching at both the day-ahead and the real-time stages.

2. We propose a new approach for solving the adversarial subproblem (i.e. the max-min stage) of an adaptive robust optimization problem, which is a min-max-min problem in its entirety, when it has the structure of an interdiction game. The adversarial subproblem is a mixed integer

recourse problem with binary uncertainty. We integrate our approach for the resolution of the max-min adversarial subproblem into a known column-and-constraint generation algorithm for resolving AROMIP. In this way, we obtain a tractable procedure for solving the day-ahead market clearing problem with proactive switching.

3. We perform simulations of proactive and reactive transmission switching on a detailed instance of the CWE network and comment on the impacts of both proactive and reactive transmission switching on the operating costs of the system.

2.3 Models of transmission switching in zonal markets

In a zonal market, the nodes of the network are aggregated into a set of zones. Network constraints inside a zone are simplified, and there is a unique price for each zone. Market clearing takes place in two stages. The zonal market is cleared in the day-ahead stage and results in the commitment of slow units. Then, as network constraints have not been represented exactly, and as the state of the grid evolves between the day ahead and real time, the system operator conducts re-dispatch (also referred to as congestion management) and balancing close to real time, while respecting the commitment of units determined in the day-ahead stage. This process ensures the feasibility of the dispatch with respect to the actual state of the grid at the time of delivery, as well as the balancing of supply and demand.

In this work, we follow this two-settlement organization of the market by presenting a two-stage model with a zonal market clearing in the day ahead and a re-dispatch and balancing process in real time. Transmission switching can be used in both the first stage and the second stage.

Definition 2.1. *The term **reactive switching** is used when switching is employed only in real-time operations (where it affects balancing and congestion management).*

Definition 2.2. *The term **proactive switching** is used when switching is employed both in day-ahead market clearing (where it affects the commitment of slow units) as well as real-time operations.*

The two-stage structure of our model with the different inputs and outputs of each module is represented graphically in Figure 2.2.

2.3.1 Day-ahead market clearing with proactive switching

We define the net position of a node (resp. zone) as the difference between the power produced and consumed within that node (resp. zone). Whereas

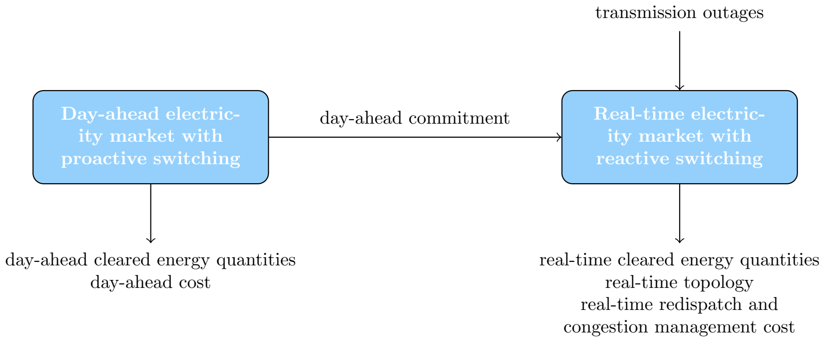


Figure 2.2: Block diagram of the two-settlement system used in this work.

a nodal market clearing model defines constraints on the net position of each node, a zonal market clearing model constrains the *zonal* net positions. The set of feasible zonal net positions depends on the chosen topology of the grid. We index the chosen topology by t in the sequel.

Extending the formulation in [ALPS21], we can write a zonal market clearing model with transmission switching in its simplest form as follows:

$$\min_{v\in[0,1],p,t} \sum_{g\in G} P_g Q_g v_g \tag{2.1}$$

$$\text{s.t.} \sum_{g\in G(z)} Q_g v_g - p_z = \sum_{n\in N(z)} Q_n \qquad \forall z \in Z \tag{2.2}$$

$$p \in \mathcal{P}_t \tag{2.3}$$

where Q_g and P_g correspond to the quantity and price bid by generator $g \in G$, $G(n)$ is the set of generators at node $n \in N$, Q_n is the forecast demand at node n, p_z corresponds to the net position of zone $z \in Z$, v_g is the acceptance/rejection decision for the bid placed by generator g, and $G(z), N(z)$ correspond to the set of generators and nodes within a zone z. The set $\mathcal{P}_t$ corresponds to the feasible set of net positions for a particular topology t. In practice, the day-ahead model (2.1)-(2.3) results also in the commitment of slow units that are allowed to submit block bids. We do not represent explicitly the binary commitment variables in this model for the sake of simplicity in the exposition.

Note that, in formulation (2.1)-(2.3), the control space is limited to a $|Z|$-dimensional space, which is much smaller than an $|N|$-dimensional space (the control space in nodal pricing). For instance, in the CWE instance that we use in our case study, there are 632 buses but only 5 zones.

We adopt the following assumption in the sequel:

Assumption 2.1. *We assume that $\mathcal{P}_t$ can be described as a set of linear inequalities which implicate a binary vector $t \in \{0,1\}^{|L|}$, where L is the set of lines in the network.*

This set $\mathcal{P}_t$, that imposes constraints on the net positions in the day-ahead market, has a central role in the zonal market model. It is the basic building block that must be defined in order to characterize the market and that should be modified in order to account for new features, such as transmission switching or the N-1 security criterion. We have discussed the different forms that the set $\mathcal{P}_t$ can take in the introduction to the thesis (section 1.2.2). For instance, the formulation currently used in practice in a large part of the EU is the one based on $\mathcal{P}^{\text{FBMC}}$, defined in equation (1.3). As we discussed in section 1.2.2, however, a certain number of challenges emerge when one tries to use $\mathcal{P}^{\text{FBMC}}$ for the purpose of policy analysis. Instead of building a model that attempts to replicate exactly the current practice, we follow the idea of [ALPS21] and use the set $\mathcal{P}^{\text{FBMC-EP}}$, defined in equation (1.3), as a starting point of our analysis of FBMC. We extend this definition in order to account for the possibility to use transmission switching while defining the acceptable set of net positions. This leads to the following definition for $\mathcal{P}_t$:

$$
\begin{aligned}
\mathcal{P}_t = \Big\{ & p \in \mathbb{R}^{|Z|} : \\
& \exists (\bar{v}, f, \theta, t) \in [0,1]^{|G|} \times \mathbb{R}^{|L|} \times \mathbb{R}^{|N|} \times \{0,1\}^{|L|} : \\
& \sum_{g \in G(z)} Q_g \bar{v}_g - p_z = \sum_{n \in N(z)} Q_n, \quad \forall z \in Z && \text{(2.4a)} \\
& \sum_{g \in G(n)} Q_g \bar{v}_g - \sum_{l \in L(n,\cdot)} f_l + \sum_{l \in L(\cdot,n)} f_l = Q_n, && \text{(2.4b)} \\
& \qquad\qquad\qquad\qquad\qquad\qquad \forall n \in N \\
& - t_l F_l \leq f_l \leq t_l F_l, \quad \forall l \in L && \text{(2.4c)} \\
& f_l \leq B_l(\theta_{m(l)} - \theta_{n(l)}) + M(1 - t_l), \quad \forall l \in L && \text{(2.4d)} \\
& f_l \geq B_l(\theta_{m(l)} - \theta_{n(l)}) - M(1 - t_l), \quad \forall l \in L \Big\} && \text{(2.4e)}
\end{aligned}
$$

Here, B_l is the susceptance of line $l \in L$, F_l is the thermal limit, and $m(l)$ and $n(l)$ are the adjacent nodes of line l (in the outgoing and incoming directions respectively), $L(m,n)$ is the set of lines directed from node m to node n, f_l is the flow through line l, θ_n is the voltage angle at node n, and M is a sufficiently large constant.

The interpretation of set $\mathcal{P}_t$ is as follows: for every acceptable vector of net positions p, there should exist a generator dispatch $\bar{v}$ that aggregates to this vector of net positions (equation (2.4a)) while respecting the nodal grid constraints (equations (2.4b)-(2.4e)) under transmission switching. Switching off a line is modeled here by binary variables t_l. If $t_l = 0$, then line l is disconnected. This implies that the flow on the line must be zero (equation (2.4c)) and the voltage angles at the two ends of the line must be independent ((2.4d) and (2.4e) become trivially satisfied if M is large enough).

In this way, the definition of $\mathcal{P}_t$ that we use circumvents the problem of circular definitions and discretionary values for the flow-based polytope and represents exactly the set of net positions that respects the two following principles: all trades that are feasible for the grid are included and all trades that provably lead to a violation of at least one transmission constraint are excluded.

2.3.2 Day-ahead market clearing with proactive switching and security criterion

In practice, the clearing of the European day-ahead market needs to respect the N-1 security criterion. We define N-1 robustness as the ability of a system to serve demand under any outage of a single transmission line in the system. The zonal market clearing model, as presented in (2.1) - (2.3), does not respect N-1 security. The modifications that are required in order to introduce N-1 security depend on whether the remedial actions (RA) (i.e. the actions that the TSOs resort to in response to a contingency) are preventive or curative.

Definition 2.3. *A **preventive** remedial action is an action (e.g. re-dispatching, topology measure, ...) taken by the TSO to respond to a potential contingency before the realization of that contingency.*

Definition 2.4. *A **curative** remedial action is an action taken by the TSO to react to the occurrence of a contingency.*

In theory, RAs can be preventive or curative [Eur18b]. In practice, however, re-dispatch is always used in a preventive way as curative re-dispatch is not considered to be safe enough by TSOs. Topology measures, in contrast, can be applied both preventively and curatively. In what follows, we modify model (2.1) - (2.3) in order to account for the N-1 security criterion with purely preventive dispatch (subsection 2.3.2), purely curative re-dispatch (subsection 2.3.2) and a hybrid preventive-curative re-dispatch (subsection 2.3.2).

Let $u \in \{0,1\}^{|L|}$ be the vector of contingencies. When one element of vector u is equal to 1, it means that the corresponding transmission line is out of service[1].

Preventive re-dispatch

The constraint on the acceptable net positions with preventive dispatch can be written as:

$$p \in \mathcal{P}_t^{\text{prev}}, \tag{2.5}$$

with

$$\mathcal{P}_t^{\text{prev}} = \Big\{ p \in \mathbb{R}^{|Z|} : \exists\ \bar{v} \in [0,1]^{|G|} :$$

[1]Generator contingencies are indirectly taken into account through reserve requirements, that are imposed in our case study but not represented here for clarity of the exposition. We therefore do not include generator contingencies in the N-1 security requirements directly, but we note that it would be straightforward to extend our model in order to take them into account.

$$\sum_{g \in G(z)} Q_g \bar{v}_g - p_z = \sum_{n \in N(z)} Q_n, \qquad \forall z \in Z$$
$$\bar{v} \in \bigcap_{\|u\|_1 \leq 1} \mathcal{V}_t(u) \Big\}$$

and

$$\begin{aligned}
\mathcal{V}_t(u) = \Big\{ & v \in [0,1]^{|G|} : \\
& \exists (f, \theta, t) \in \mathbb{R}^{|L|} \times \mathbb{R}^{|N|} \times \{0,1\}^{|L|} : \\
& \sum_{g \in G(n)} Q_g v_g - \sum_{l \in L(n,\cdot)} f_l + \sum_{l \in L(\cdot,n)} f_l = Q_n, \quad \forall n \in N \\
& - t_l F_l \leq f_l \leq t_l F_l, \qquad \forall l \in L \\
& f_l \leq (1 - u_l) B_l (\theta_{m(l)} - \theta_{n(l)}) + M(1 - t_l), \qquad \forall l \in L \\
& f_l \geq (1 - u_l) B_l (\theta_{m(l)} - \theta_{n(l)}) - M(1 - t_l), \qquad \forall l \in L \Big\}
\end{aligned}$$

The set $\mathcal{V}_t(u)$ corresponds to all dispatch decisions that respect power flow constraints and line limits under contingency u, when transmission switching is allowed. The interpretation of the set $\mathcal{P}_t^{\text{prev}}$ is thus the following: to every acceptable vector of net positions p, there should exist a generator dispatch $\bar{v}$ that aggregates to this vector of net positions, and that respects grid constraints for every contingency u.

The market clearing model under the N-1 security criterion with preventive dispatch can be represented through equations (2.1), (2.2) and (2.5).

Curative re-dispatch

The flow-based domain with curative re-dispatch can be described as follows:

$$p \in \bigcap_{\|u\|_1 \leq 1} \mathcal{P}_t^{\text{cur}}(u) \tag{2.6}$$

with

$$\begin{aligned}
\mathcal{P}_t^{\text{cur}}(u) = \Big\{ & p \in \mathbb{R}^{|Z|} : \\
& \exists (\bar{v}, f, \theta, t) \in [0,1]^{|G|} \times \mathbb{R}^{|L|} \times \mathbb{R}^{|N|} \times \{0,1\}^{|L|} : \\
& \sum_{g \in G(z)} Q_g \bar{v}_g - p_z = \sum_{n \in N(z)} Q_n, \ \ \forall z \in Z \\
& \sum_{g \in G(n)} Q_g \bar{v}_g - \sum_{l \in L(n,\cdot)} f_l + \sum_{l \in L(\cdot,n)} f_l = Q_n, \ \ \forall n \in N \\
& - t_l F_l \leq f_l \leq t_l F_l, \ \ \forall l \in L
\end{aligned}$$

$$f_l \leq (1-u_l)B_l(\theta_{m(l)} - \theta_{n(l)}) + M(1-t_l), \ \ \forall l \in L$$
$$f_l \geq (1-u_l)B_l(\theta_{m(l)} - \theta_{n(l)}) - M(1-t_l), \ \ \forall l \in L\Big\}$$

The market clearing model under the N-1 security criterion with curative re-dispatch can be represented through equations (2.1), (2.2) and (2.6).

The interpretation of equations (2.6) is as follows: a vector of net positions is acceptable if, for every contingency, there exists a dispatch that respects all grid constraints. Note the *fundamental difference* with the case of preventive re-dispatch: in *curative* re-dispatch, the dispatch can be *different* for every contingency, while in the case of *preventive* re-dispatch, the dispatch must be *the same* for every contingency. The reader is referred to Appendix 2.A for an illustrative example of the difference between a day-ahead model with preventive and curative re-dispatch.

Mathematically, the main difference between the flow-based domain described by equation (2.5) and the domain described by equation (2.6) is that, in equation (2.5), the intersection over all contingencies is over a set of dimension $|G|$; instead, in equation (2.6), the intersection is over a set of dimension $|Z|$, which is much smaller. This has computational implications, as we discuss later.

Hybrid preventive-curative re-dispatch

Combining the two previous models, we can easily extend them to the case of a re-dispatch that is neither purely preventive nor purely curative: the *hybrid* preventive-curative model. A hybrid model corresponds to a dispatch that is preventive for a subset of contingencies U_{prev}, and curative for all other contingencies $U_{\text{cur}} = U \backslash U_{\text{prev}}$. The flow-based domain for this hybrid model can be written as follows:

$$\mathcal{P}^{\text{FB}}_{\text{hyb}} = \Big\{ p \in \mathbb{R}^{|Z|} : \exists\ \bar{v} \in [0,1]^{|G|} :$$
$$\sum_{g \in G(z)} Q_g \bar{v}_g - p_z = \sum_{n \in N(z)} Q_n, \ \ \forall z \in Z \tag{2.7a}$$
$$\bar{v} \in \mathcal{V}_t(u) \ \ \forall u \in U_{\text{prev}}, \tag{2.7b}$$
$$p \in \mathcal{P}^{\text{cur}}_t(u) \ \ \forall u \in U_{\text{cur}} \Big\} \tag{2.7c}$$

Model (2.7) is a direct extension of the preventive and the curative cases. Equations (2.7a)-(2.7b) express the fact that the acceptable net positions should disaggregate in a dispatch that is feasible for all contingencies belonging to set U_{prev}, at the same time. Equation (2.7c) express the fact that for any contingency belonging to U_{cur}, the net positions disaggregate in a dispatch that is feasible for the grid, after the realization of that contingency.

2.3.3 Real-time re-dispatch and balancing

The goal of the re-dispatch and balancing process is to modify the day-ahead dispatch so as to balance the system at minimum cost in real time, while respecting network constraints. We assume that there is no uncertainty in demand and renewable production. Uncertainty can only materialize in a transmission line outage. Thus, we focus on the re-dispatch and balancing actions that are required (i) in case the zonal day-ahead auction violates any inter or intra-zonal transmission constraints and (ii) in case the operator is required to react to a transmission contingency that may occur between the day-ahead and real time.

We consider various real-time models in the case study, depending on the objective of the TSO and the degree of coordination among TSOs. The ideal standard in terms of coordination and cost minimization is the perfectly coordinated re-dispatch and balancing model which aims at minimizing real-time cost. We introduce reactive transmission switching to this model, and formulate it as an optimization problem as follows:

$$\begin{aligned}
\min_{\substack{v\in[0,1]\\ f,\theta,t}} \quad & \sum_{g\in G} P_g Q_g v_g \\
\text{s.t.} \quad & \sum_{g\in G(n)} Q_g v_g - \sum_{l\in L(n,\cdot)} f_l + \sum_{l\in L(\cdot,n)} f_l = Q_n, \\
& \qquad\qquad\qquad\qquad\qquad\qquad \forall n \in N \\
& -F_l t_l \le f_l \le F_l t_l, \qquad \forall l \in L \\
& f_l \le B_l(\theta_{m(l)} - \theta_{n(l)}) + M(1-t_l), \qquad \forall l \in L \\
& f_l \ge B_l(\theta_{m(l)} - \theta_{n(l)}) - M(1-t_l), \qquad \forall l \in L
\end{aligned}$$

We stress the fact that, although this second-stage real-time problem is nodal, the day-ahead zonal market clearing drives the real-time behavior because of the fixed unit commitment decisions that must be respected. We do not represent here the dependence of real-time dispatch on day-ahead unit commitment decisions, in order to simplify the exposition, but enforce this requirement for the simulations presented in the case study.

2.4 An algorithm for proactive transmission switching

In this section, we present an algorithm for solving the zonal day-ahead market clearing model under N-1 robustness with proactive transmission switching and a purely curative re-dispatch. We then discuss how the algorithm can be adapted to the case of a preventive re-dispatch. The zonal day-ahead market clearing problem can be written as:

$$\min_{\substack{v\in[0,1],p,\\ t\in\{0,1\}}} \sum_{g\in G} P_g Q_g v_g \tag{2.8}$$

$$\text{s.t.} \sum_{g\in G(z)} Q_g v_g - p_z = \sum_{n\in N(z)} Q_n, \qquad \forall z \in Z \tag{2.9}$$

$$p \in \bigcap_{\|u\|_1\leq 1} \mathcal{P}_t(u) \tag{2.10}$$

The difficulty in solving this problem lies in the fact that its equivalent monolithic formulation is too large to be solved directly, while the non-convexity of the set of feasible net positions prevents the use of a cutting plane approach similar to the one proposed by [ALPS21].

Our idea for solving this problem is to rewrite it as an adaptive robust optimization problem with mixed integer recourse (AROMIP), and to use a known column-and-constraint generation (C&CG) algorithm for this class of problems. The general AROMIP that we consider can be described as follows:

$$\min_{\mathbf{x}\in\mathbb{X}} \mathbf{cx} + \max_{\mathbf{u}\in\mathbb{U}} \min_{\mathbf{z},\mathbf{y}\in\mathbb{F}(\mathbf{u},\mathbf{x})} \mathbf{dy} + \mathbf{gz} \tag{2.11}$$

where $\mathbb{X} = \{\mathbf{x} \in \mathbb{R}^m_+ \times \mathbb{Z}^m_+ : A\mathbf{x} \geq b\}, \mathbb{F}(\mathbf{u},\mathbf{x}) = \{(\mathbf{z},\mathbf{y}) \in \mathbb{Z}^n_+ \times \mathbb{R}^p_+ : E(\mathbf{u})\mathbf{y} + G(\mathbf{u})\mathbf{z} \geq f(\mathbf{u}) - R\mathbf{u} - D(\mathbf{u})\mathbf{x}\}$, and the uncertainty set $\mathbb{U}$ is a bounded binary set in the form of $\mathbb{U} = \{u \in \mathbb{B}^q_+ : H\mathbf{u} \leq a\}$. This formulation is similar to that of [ZZ12]. The only difference is that we restrict ourselves to a pure binary uncertainty set $\mathbb{U}$, which corresponds in our case to the set of line contingencies. We also consider a more general form of $\mathbb{F}(\mathbf{u},\mathbf{x})$, where every parameter (E, G, f, D) can depend on the realization of uncertainty.

Let us now show how we can reformulate our problem in the general form (2.11). First, notice that in problem (2.8) - (2.10) it is equivalent to replace constraint (2.10) by

$$d(p, \bigcap_{\|u\|_1\leq 1} \mathcal{P}_t(u)) = 0 \tag{2.12}$$

where $d(p, \bigcap_{\|u\|_1\leq 1} \mathcal{P}_t(u))$ is the L_1 distance of injection p to the set of net positions.

We then penalize this function in the objective and show in Proposition 2.1 that there exists a penalizing factor $\overline{\lambda}$ such that problem (2.8) - (2.9), (2.12) is equivalent to

$$\min_{v\in[0,1],p,t} \sum_{g\in G} P_g Q_g v_g + \overline{\lambda}\Big(d(p, \bigcap_{\|u\|_1\leq 1} \mathcal{P}_t(u))\Big) \tag{2.13}$$

$$\text{s.t.} \sum_{g\in G(z)} Q_g v_g - p_z = \sum_{n\in N(z)} Q_n, \qquad \forall z \in Z \tag{2.14}$$

Proposition 2.1. *Consider the following two optimization problems:*

$$\begin{aligned}(P1): \min_{p,t} \ & cp \\ \text{s.t.} \ & Ap \le b \\ & d(p, \mathcal{P}_t) = 0\end{aligned}$$

and

$$\begin{aligned}(P2): \min_{p,t} \ & cp + \overline{\lambda}\, d(p, \mathcal{P}_t) \\ \text{s.t.} \ & Ap \le b\end{aligned}$$

where $\mathcal{P}_t$ is a general polyhedron described by a set of linear inequalities implicating an integer vector of variables t, and where $d(\cdot,\cdot)$ is the L1 distance function of a vector to a polyhedron. There exists a scalar $\overline{\lambda}$ such that (P1) and (P2) are equivalent[2]*.*

Proof. The proof is inspired by [SMBP19]. Suppose, without loss of generality, that $\mathcal{P}_t = \{p \in \mathbb{R}^Z : \exists t \in \{0,1\}^{|L|} : Vp + Ut \le W\}$, where V and U are matrices and W is a vector of appropriate dimensions. Then,

$$\begin{aligned}(P1) \Leftrightarrow \min_{t} \min_{p,\tilde{p},s_1,s_2} \ & cp \\ \text{s.t.} \ & Ap \le b \\ & \mathbb{1}^\top s_1 + \mathbb{1}^\top s_2 \le 0 \ \ [\lambda(t)] \\ & s_{1i} \ge p_i - \tilde{p}_i \quad \forall i = \{1,...,Z\} \\ & s_{2i} \ge \tilde{p}_i - p_i \quad \forall i = \{1,...,Z\} \\ & V\tilde{p} \le W - Ut \\ & s_1, s_2 \ge 0\end{aligned}$$

where the scalar $\lambda(t)$ is the inner-problem dual variable of the constraint on the distance between p and $\tilde{p}$, that depends on t. Let $\overline{\lambda}$ be an upper bound of $\lambda(t)$ for each $t \in \{0,1\}^{|L|}$. By using Lemma 1 of [SMBP19], we have that

$$\begin{aligned}(P1) \Leftrightarrow \min_{t} \min_{p,\tilde{p},s_1,s_2} \ & cp + \overline{\lambda}(\mathbb{1}^\top s_1 + \mathbb{1}^\top s_2)^+ \\ \text{s.t.} \ & Ap \le b \\ & s_{1i} \ge p_i - \tilde{p}_i \quad \forall i \in \{1,...,Z\} \\ & s_{2i} \ge \tilde{p}_i - p_i \quad \forall i \in \{1,...,Z\} \\ & V\tilde{p} \le W - Ut\end{aligned}$$

[2]Note that this proposition also holds for distances to more general forms of sets. The only property that we use in the proof is that the set $\mathcal{P}_t$ can be formulated as a set of linear inequalities involving a vector of integer variables. The proposition could thus also be applied to the L1 distance to the union of nonempty polyhedra that, as is well known, can be written in this form (see, for instance, [CSF19]).

$$s_1, s_2 \geq 0$$

By the non-negativity of s_1 and s_2, $(\mathbb{1}^\top s_1 + \mathbb{1}^\top s_2)^+ = \mathbb{1}^\top s_1 + \mathbb{1}^\top s_2$. Therefore, we conclude, by the definition of the L1 distance, that

$$(P1) \Leftrightarrow \begin{bmatrix} \min\limits_{p,t} \; cp + \overline{\lambda}\, d(p, \mathcal{P}_t) \\ \text{s.t. } \; Ap \leq b \end{bmatrix} \Leftrightarrow (P2)$$

□

We now define explicitly the distance function as the following max-min problem:

$$d(p, \bigcap_{\|u\|_1 \leq 1} \mathcal{P}_t(u)) = \max_{u \in \mathbb{U}} \min_{\tilde{p},t} \; \|p - \tilde{p}\|_1 \tag{2.15}$$

$$\text{s.t. } \; \tilde{p} \in \mathcal{P}_t(u) \tag{2.16}$$

With the formulation (2.13) - (2.14) which is justified by the result of Proposition 2.1, we are now in the framework of adaptive robust optimization with mixed integer recourse. The correspondence in notation between the generic AROMIP and our specific application is the following: $\mathbf{x} = (v, p)$[3], $\mathbf{z} = t$, $\mathbf{y} = (s_1, s_2, \tilde{p})$, $\mathbf{dy} + \mathbf{gz} = \sum_{g \in G} s_{1,g} + s_{2,g}$, $\mathbb{Y} = \{v, p : \sum_{g \in G(z)} Q_g v_g - p_z = \sum_{n \in N(z)} Q_n \; \forall z \in Z\}$, $\mathbb{U}$ is the set of all possible contingencies such that $\|u\|_1 \leq 1$, and

$$\mathbb{F}(\mathbf{u}, \mathbf{x}) = \Big\{ \tilde{p} \; : \tilde{p} \in \mathcal{P}_t(u) \\ s_{1,g} \geq p_g - \tilde{p}_g, \;\; \forall g \in G \\ s_{2,g} \geq \tilde{p}_g - p_g, \;\; \forall g \in G \Big\}$$

which can be written as a mixed integer linear feasibility set if Assumption 2.1 holds.

2.4.1 Outer-level column-and-constraint generation algorithm

Two different classes of methods have been proposed in the literature for solving two-stage robust optimization problems [ZZ13a]. *Benders dual methods*, as in Benders decomposition, use the dual information of the second-stage problem to sequentially approximate the first-stage value function. *Column-and-constraint generation methods* gradually include the variables and constraints

[3] Note that block bids were omitted in this section for simplicity of the exposition, but the equivalence with the AROMIP and correctness of the algorithm remain when these types of bids are considered. Indeed, in case of block bids, some elements of the vector v would have binary restrictions, which would imply that the corresponding vector $\mathbf{x}$ would have binary restrictions. This is in line with the definition of $\mathbf{x}$ in (2.11).

of the monolithic formulation. In [ALPS21], the first approach has been used, leading to a cutting plane based algorithm for solving the market clearing problem. However, as mentioned previously, the presence of binary variables in the second-stage problem, which correspond to transmission switching decisions, prevents the use of Benders dual methods in the context of our problem. In contrast, [ZZ12] describes how a column-and-constraint generation method can be used for solving adaptive robust optimization problems with integer variables in the recourse problem, provided we can solve exactly the second-stage max-min problem for a given choice of first-stage decisions.

Let us therefore assume that we can solve this second-stage problem (which corresponds, in our case, to computing the distance of a net position vector to the set of net positions). We will then explain how we can solve the second-stage problem in section 2.4.2. Algorithm 1 presents the column-and-constraint generation algorithm of [ZZ12] applied to our problem.

Algorithm 1 can be further simplified by noticing that, by definition of $\overline{\lambda}$, $\eta^* = 0$ at each iteration of the algorithm. This renders the algorithm totally independent of $\overline{\lambda}$, for $\overline{\lambda}$ sufficiently large. This also implies that $LB = \sum_g Q_g P_g v_g^*$ at each iteration, and thus that the algorithm will terminate when

$$d(p^*, \bigcap_{\|u\|_1 \leq 1} \mathcal{P}_t(u_i)) < \epsilon \quad \forall i \in \{1, ..., k\},$$

i.e. when the optimal net position obtained with only a subset of the possible contingencies is actually robust to all contingencies.

2.4.2 Inner level max-min problem

So far, we have assumed that we were able to solve exactly the second-stage problem. To have a complete algorithm, it remains to show how we can solve the second-stage problem. Recall that the second-stage problem can be written as follows:

$$\begin{aligned} d(p, \bigcap_{\|u\|_1 \leq 1} \mathcal{P}_t(u)) = \max_{u \in \mathbb{U}} \ & \min_{\tilde{p}, t} \ |p - \tilde{p}| \\ & \text{s.t. } \tilde{p} \in \mathcal{P}_t(u) \end{aligned} \tag{2.17}$$

Zhao and Zeng [ZZ12] propose solving this inner problem by using a column-and-constraint generation algorithm, exactly in the same fashion as the outer problem. With this method, the master problem first considers only a subset of topologies. The master problem clears with the best net position that corresponds to this subset of topologies. Then, the subproblem identifies the best topology than can react to the vector of net positions identified and this topology is added to the master problem. This procedure repeats until no more new topologies are added to the master. The drawback of this method is that, for each topology under consideration, the other variables (v, f, θ in our case) must be duplicated. This quickly introduces a bottleneck in terms of efficiency in the master problem, which increases the run time of the algorithm.

Algorithm 1 Column-And-Constraint Generation Algorithm

1. Set $LB = +\infty, UB = -\infty$ and $k = 0$

2. Solve the following master problem:

$$\begin{aligned} \textbf{MP:} \min_{\substack{v\in[0,1]\\ p,t^i,\eta}} & \sum_g Q_g P_g v_g + \overline{\lambda}\eta \\ \text{s.t.} & \sum_{g\in G(z)} Q_g v_g - p_z = \sum_{n\in N(z)} Q_n \\ & \eta \geq |p^i - p|, \qquad \forall i \in \{1, ..., k\} \\ & p^i \in \mathcal{P}_{t^i}(u^i), \qquad \forall i \in \{1, ..., k\} \end{aligned}$$

   Update $LB = \sum_g Q_g P_g v_g^* + \overline{\lambda}\eta^*$, where v^* and η^* are the optimal value of v and η in this subproblem. If $UB - LB < \epsilon$, terminate.

3. Call the oracle to solve subproblem $d(p^*, \bigcap_{\|u\|_1\leq 1} \mathcal{P}_t(u))$ and update UB as

$$\min\left(UB, \sum_g Q_g P_g v_g^* + d(p^*, \bigcap_{\|u\|_1\leq 1} \mathcal{P}_t(u))\right)$$

   If $UB - LB < \epsilon$, terminate.

4. Create variable p^i and add the following constraints

$$\begin{aligned} &\eta \geq |p^i - p| \\ &p^i \in \mathcal{P}_{t^i}(u_i^*) \end{aligned}$$

   where u^* is the optimal value of variable u in the subproblem of step 3.

In what follows, we propose a new approach for the inner max-min problem that avoids the bottleneck of the nested C&CG method. Our approach builds on the observation that this problem falls in the class of interdiction games, i.e. two-player Stackelberg games where the decision variables of the leader are binary. When set to 1, these variables force the corresponding follower variables to be 0, thereby interdicting the follower from choosing certain actions. In our context, the leader is Nature and is looking for the line to place out of service so as to maximize the disruption to the system operator. The system operator can react with switching.

Based on [FLMS19], we will first show how a cutting plane formulation of our problem can be obtained. Let $\mathcal{Q}$ be the set $\mathcal{P}_t(\mathbf{0})$ in the space of p and t, i.e. the feasible set of net positions and switching variables under no contingency. Then, problem (2.17) can be formulated equivalently as

$$\begin{aligned} \max_{u \in \mathbb{U}} \min_{\tilde{p},t} \ & |p - \tilde{p}| \\ \text{s.t. } & (\tilde{p}, t) \in \mathcal{Q} \\ & t_l u_l = 0 \qquad \forall l \in L \end{aligned} \tag{2.18}$$

We also have the following result.

Proposition 2.2. *Let us consider the following problem for a fixed vector u^*:*

$$\min_{\tilde{p},t} \ |p - \tilde{p}| \tag{2.19a}$$

$$s.t. \ (\tilde{p}, t) \in \mathcal{Q} \tag{2.19b}$$

$$t_l u_l^* = 0 \qquad \forall l \in L \tag{2.19c}$$

If λ_l is an optimal Lagrangian dual multiplier of constraint (2.19c)*, then problem* (2.19) *is equivalent to*

$$\begin{aligned} \min_{\tilde{p},t} \ & |p - \tilde{p}| + \sum_{l \in L} \lambda_l t_l u_l^* \\ s.t. \ & (\tilde{p}, t) \in \mathcal{Q}, \end{aligned} \tag{2.20}$$

i.e. there is no Lagrangian duality gap for constraint (2.19c)[4]*.*

Proof. First notice that, because constraint (2.19c) is an equality constraint, if the solution of problem (2.20) satisfies $t_l u_l^* = 0 \ \forall l \in L$, then the duality gap is zero. We define the following two quantities:

$$\begin{aligned} \alpha = \min_{\tilde{p},t} \ & |p - \tilde{p}| \\ \text{s.t. } & (\tilde{p}, t) \in \mathcal{Q} \end{aligned} \qquad \text{and} \qquad \begin{aligned} \beta_l = \min_{\tilde{p},t} \ & |p - \tilde{p}| \\ \text{s.t. } & (\tilde{p}, t) \in \mathcal{Q} \\ & t_l u_l^* = 0 \end{aligned}$$

[4]We use here the generalization of Lagrangian duality to mixed-integer programs that is defined in, e.g. [Tan05]. In general, the Lagrangian duality gap is nonzero for MIP.

If u^* is the zero vector, every real value is a dual optimal multiplier, the constraint $t_l u_l^* = 0$ is naturally satisfied $\forall l \in L$, and the duality gap is zero. Else, u^* is a vector of zeros with one entry set to one. Let m be the index of that entry. Note that the optimal objective value of problem (2.19) is β_m by definition. Let t_m^* be the optimal value of problem (2.20) with $\lambda_l = \beta_l - \alpha \ \forall l \in L$. If $t_m^* = 1$, then the optimal objective value of problem (2.20) is $\alpha + \beta_m - \alpha = \beta_m$. If $t_m^* = 0$, then its optimal objective is also β_m. We conclude that the duality gap is zero and that $\beta_l - \alpha$ is a dual optimal Lagrange multiplier. □

Using Proposition 2.2, we deduce that if λ_l is a dual optimal multiplier, problem (2.18) is equivalent to

$$\begin{aligned} \max_{u \in \mathbb{U}} \ \min_{\tilde{p}, t} \ & |p - \tilde{p}| + \sum_{l \in L} \lambda_l t_l u_l \\ \text{s.t.} \ & (\tilde{p}, t) \in \mathcal{Q} \end{aligned} \tag{2.21}$$

Note that the feasible set of the inner level of problem (2.21) does not depend on u anymore. Moreover, as its objective function is linear, at least one extreme point of $\operatorname{conv}(\mathcal{Q})$ is optimal, where conv denotes the convex hull. Let us denote by $\operatorname{ext}(\mathcal{Q})$ the set of extreme points of $\mathcal{Q}$. Then, problem (2.21) is consequently also equivalent to

$$\max_{u \in \mathbb{U}} \ \min \left\{ |p - \tilde{p}| + \sum_{l \in L} \lambda_l t_l u_l : (\tilde{p}, t) \in \operatorname{ext}(\mathcal{Q}) \right\}$$

and to

$$\begin{aligned} \max_{\substack{\zeta, p \\ u \in \mathbb{U}}} \ & \zeta \\ \text{s.t.} \ & \zeta \le |p - \tilde{p}| + \sum_{l \in L} \lambda_l t_l u_l, \quad \forall (\tilde{p}, t) \in \operatorname{ext}(\mathcal{Q}) \end{aligned} \tag{2.22}$$

where our problem has been rewritten in the form of a cutting plane formulation. Using this formulation, a cutting plane algorithm can be obtained by noticing the two following facts: (i) if problem (2.22) is solved with a subset of $\operatorname{ext}(\mathcal{Q})$, then we obtain an upper bound as well as an interdiction plan u; (ii) solving the inner level of problem (2.21) for a fixed u gives a lower bound as well as a new extreme point for the set $\operatorname{conv}(\mathcal{Q})$. Algorithm 2 formalizes the method that is suggested above.

If λ_l is a dual optimal multiplier, Algorithm 2 is guaranteed to converge in a finite number of iterations. Note, however, that if λ_l is a dual optimal multiplier, then every $\tilde{\lambda}_l$ such that $\tilde{\lambda}_l > \lambda_l$ is also a dual optimal multiplier. This method also results in the decomposition of the problem into two subproblems, as in the case of Zhao and Zeng [ZZ12]. The difference is that the Worst Uncertainty Oracle subproblem is solved much more efficiently than the master problem of [ZZ12]. Since the master problem is the bottleneck of [ZZ12], our approach achieves a material improvement over Zhao and Zeng.

Algorithm 2 Inner Level

1. Set $LB = +\infty, UB = -\infty$, $k = 0$ and $\text{ext}(\mathcal{Q})^0 = \emptyset$

2. Solve the following Worst Uncertainty Oracle :

$$\begin{aligned} \max_{\substack{\zeta, p \\ u \in \mathbb{U}}} \ & \zeta \\ \text{s.t.} \ & \zeta \leq |p - \tilde{p}| + \sum_{l \in L} \lambda_l t_l u_l, \ \ \forall (\tilde{p}, t) \in \text{ext}(\mathcal{Q})^k \end{aligned}$$

   Denote by u^* the optimal value of variable u and update UB to the optimal objective value.

3. Solve the following Best Reaction Oracle :

$$\begin{aligned} \min_{\tilde{p}, t} \ & |p - \tilde{p}| + \sum_{l \in L} \lambda_l t_l u_l^* \\ \text{s.t.} \ & (\tilde{p}, t) \in \mathcal{Q} \end{aligned}$$

   Denote by $\tilde{p}^*$ and t^* the optimal values of variables $\tilde{p}$ and t respectively.
   Let $\text{ext}(\mathcal{Q})^{k+1} \leftarrow \text{ext}(\mathcal{Q})^k \cup (\tilde{p}^*, t^*)$.
   Let $LB \leftarrow \max\{LB, |p - \tilde{p}^*| + \sum_{l \in L} \lambda_l t_l^* u_l\}$.

4. If $UB - LB < \epsilon$, terminate. Else, let $k \leftarrow k + 1$ and go back to step 2.

The speed of convergence is largely determined by the value of the dual multiplier. Notice that the smaller λ_l is, the tighter the formulation (2.22) is. Thus, the goal is to find the smallest possible dual optimal multiplier of constraint (2.19c). If λ_l is set to a large trivial value, the algorithm will have to generate almost all possible values of u before converging. In contrast, if the value chosen is close to its optimal value, the convergence can be much faster. In what follows, we present our idea for generating values for λ_l that yield fast convergence in the case of our problem.

We first mention that the proof of Proposition 2.2 highlights how we can obtain the best value of λ_l, which we denote by λ_l^*. Indeed, let

$$\begin{aligned} \gamma_l = & \min_{\tilde{p},t} \ |p - \tilde{p}| \\ & \text{s.t. } (\tilde{p}, t) \in \mathcal{Q} \\ & \quad t_l = 0 \end{aligned}$$

Then, $\lambda_l^* = \gamma_l - \alpha$. This value can be interpreted as *the cost of robustness*, i.e. the price to pay for being robust to the contingency of line l. It turns out that it can be easily upper bounded as follows: Let δ_l be defined as the objective value of the inner problem when we are robust to the contingency of line l without the possibility for switching as a recourse action:

$$\begin{aligned} \delta_l = & \min_{\substack{v \in [0,1] \\ \tilde{p}, f, \theta}} \ |p - \tilde{p}| \\ & \text{s.t.} \ \sum_{g \in G(z)} Q_g v_g - \tilde{p}_z = \sum_{n \in N(z)} Q_n, \ \ \forall z \in Z \\ & \quad \sum_{g \in G(n)} Q_g v_g - \sum_{j \in L(n,\cdot)} f_j + \sum_{j \in L(\cdot,n)} f_j = Q_n, \ \ \forall n \in N \\ & \quad - F_j \le f_j \le F_j, \ \ \forall j \in L \\ & \quad f_j = B_j(\theta_{m(j)} - \theta_{n(j)}), \ \ \forall j \in L \backslash \{l\} \\ & \quad f_l = 0 \end{aligned} \tag{2.23}$$

The value δ_l can be obtained much more efficiently than the value γ_l as it is a simple monolithic LP. Then, the following inequalities naturally hold:

$$\alpha \le \gamma_l \le \delta_l.$$

It follows that λ_l^* is bounded from above by $\delta_l - \alpha$. Algorithm 2 with $\lambda_l = \delta_l - \alpha$ is the approach that we use for solving (2.17).

2.4.3 Extension to preventive dispatch

It is straightforward to adapt the algorithm to the case of a purely preventive dispatch. In the master problem, instead of adding constraints

$$p \in \mathcal{P}_{t^i}(u^i), \qquad \forall i \in \{1, ..., k\}$$

we will add the following set of constraints, which describe the fact that the dispatch should be the same for every contingency:

$$\sum_{g \in G(z)} Q_g \bar{v}_g - p_z = \sum_{n \in N(z)} Q_n, \qquad \forall z \in Z$$
$$\bar{v} \in \mathcal{V}_{t^i}(u^i), \qquad \forall i \in \{1, ..., k\}$$

The inner problem, that identifies the next contingency to add in the master problem, now reads as follows:

$$\begin{aligned} d(\bar{v}, \bigcap_{\|u\|_1 \leq 1} \mathcal{V}_t(u)) = \max_{u \in \mathbb{U}} \ & \min_{\tilde{v}, t} \ |\bar{v} - \tilde{v}| \\ & \text{s.t. } \ \tilde{v} \in \mathcal{V}_t(u) \end{aligned} \tag{2.24}$$

The structure of problem (2.24) is exactly the same as that of problem (2.17). The same cutting plane algorithm can thus be used to solve the inner problem. Note, however, that although the algorithm can be adapted in a straightforward way, the problem with preventive re-dispatch and the problem with curative re-dispatch vary in terms of solution difficulty. This is due to the fact that, in the case of a preventive re-dispatch, the inner problem computes the distance to a set that is significantly higher dimensional than in the case of curative dispatch. Whereas the set $\mathcal{P}_t(u)$ has dimension $|Z|$, the number of zones in the system, the set $\mathcal{V}_t(u)$ is of dimension $|G|$, which corresponds to the number of generators. For instance, in our case study on a realistic instance of the CWE system, there are 5 zones, whereas the number of generators is almost 2000. This translates to an additional computational burden for achieving preventive N-1 robustness, compared to curative robustness.

2.4.4 Extension to hybrid dispatch

A simple combination of the algorithm for the preventive case and for the curative case can be used to clear the market with a hybrid N-1 robust dispatch. We note, however, that the important question of how to determine the set U_{prev} (i.e. the set of contingencies to be considered in a preventive way) remains. Our column-and-constraint generation algorithm for solving the day-ahead model suggests a way of defining this set. In this algorithm, each iteration generates a severe contingency that is added to the master problem, until convergence is reached. We propose that U_{prev} should consist of the n first contingencies produced by the algorithm, where n is an arbitrary parameter. By construction, these contingencies are selected among the most severe contingencies that the system operator is called to react to.

Note that this hybrid model, being a combination of the preventive and curative case, has a computational complexity that is intermediate between that of the curative and that of the preventive model. This complexity is increasing with n.

2.5 Case study on Central Western Europe

2.5.1 Simulation setup

In this case study, our goal is to quantify the impacts of different short-term market design options on the total day-ahead and real-time costs on a realistic instance of the CWE network. All simulations are performed on 32 different representative snapshots of system operation. Each snapshot corresponds to different demand forecasts, renewable forecasts and maintenance schedules for thermal generators.

Our analysis focuses on the impact of transmission switching on mitigating costs. As we discuss in the literature review, the transmission switching problem is computationally expensive, and one means of reducing its computational burden is by imposing a *switching budget*, i.e. a limit on the number of lines that can be switched. We use a switching budget of 6 lines in our analysis.

We use the same version of the CWE system as the one used in [ALPS21]. We present the topology of the system in Figure 2.3. The model consists of 6 countries: Belgium, the Netherlands, France, Austria, Luxembourg and Germany. These countries are grouped into 5 zones, with Luxembourg and Germany forming one single zone.

We separate the producing units into two sets, according to their flexibility and start-up capabilities. The on-off status of non-flexible units, which we refer to as *slow* units, is decided in the day-ahead, and must be respected during real-time operations. In contrast, the production of flexible units, which we refer to as *fast* units, is independent of the day-ahead schedule. This model of unit commitment with separation between slow and fast units follows the idea initially proposed by Ruiz *et. al.* [RPZ+09] and used in subsequent unit commitment models applied both to US markets [POO11] and European markets [AP17].

The system consists of: (i) 346 slow generators with a total capacity of 154 GW; (ii) 301 fast thermal generators with a total capacity of 89 GW; (iii) 1312 renewable generators with a total capacity of 149 GW; (iv) 632 buses; and (v) 945 branches. The average demand of the system amounts to 134 GW.

The formulations of the zonal market based on flow-based market coupling that we use are generalized versions of those presented in section 2.3, and are inspired by [ALPS21]. The day-ahead market clearing model considers commitment (on-off) decisions for slow generators[5], reserves and the N-1 security criterion[6].

All models and algorithms used in this study are implemented in Julia 1.0.1

[5] We allow slow generators to submit block bids (i.e. bids that are either entirely accepted or rejected) in the day-ahead auction.

[6] We assume that the commitment is determined along with the topology with the objective of maximizing welfare. Prices can be computed after the binary decisions have been fixed [OHK+10]. A full analysis of the implications of our zonal day-ahead model on pricing is however outside the scope of this work.

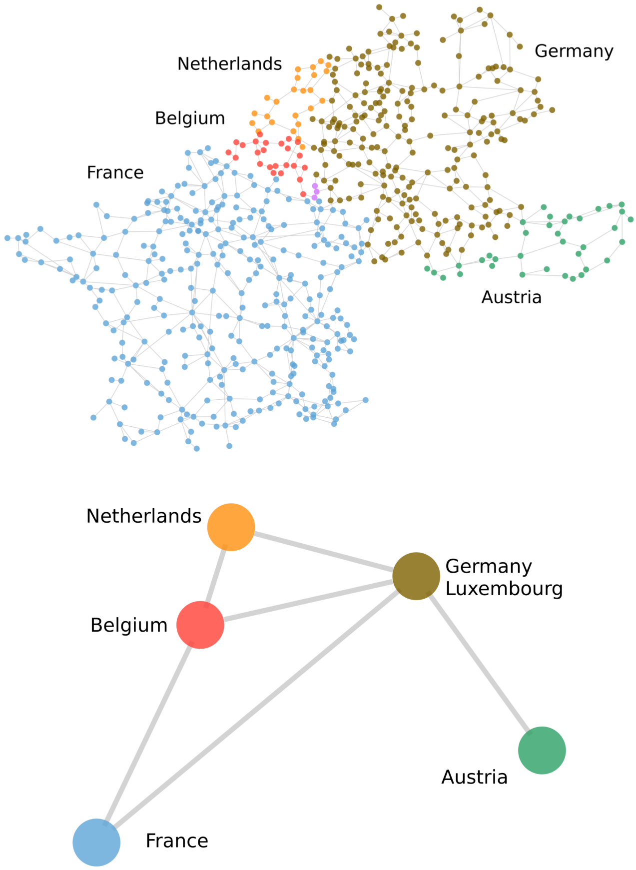


Figure 2.3: The CWE network model (top) and its zonal aggregation (bottom).

[BEKS17] using JuMP 0.18.4 [DHL17]. The models are solved with Gurobi 8.0. We parallelize the simulation over the different snapshots and we use the Lemaitre3 cluster, hosted at the Université catholique de Louvain, for the computations. The total cpu time for solving the day-ahead market clearing model with switching amounts to 10 hours and 36 minutes. We have also implemented the nested column-and-constraint approach of [ZZ12] for solving the max-min problem with MIP recourse, and record a 246% time increase compared to our proposed approach.

2.5.2 Benefits of switching in the benchmark flow-based market coupling model

In this section, we propose a benchmark model for FBMC with transmission switching and N-1 robustness and discuss the impacts of transmission switching on the total cost. As we explain in section 2.3, the way in which the N-1 security criterion is modeled in our day-ahead zonal market clearing model depends on whether costly remedial actions are preventive or curative. In theory, all remedial actions can be either preventive or curative [Eur18b]. In practice, however, TSOs consider that costly remedial actions (i.e. re-dispatching) should be fully preventive. As we discuss in section 2.3, simulating FBMC with switching and a purely preventive dispatch is computationally intractable for our instance. What we propose instead is to simulate the results for a hybrid preventive-curative model that is computationally manageable. The idea of the hybrid version of the model is that the dispatch should be robust in a preventive way to a subset of contingencies, U_{prev}, and robust in a curative way to all other contingencies, $U_{\text{cur}} = U \backslash U_{\text{prev}}$. We denote the number of contingencies considered in a preventive way by n. The computational complexity of solving the hybrid day-ahead model increases with n. This number should thus be selected as the highest number that keeps the model tractable for our instance. For FBMC with switching, we identify experimentally that n should be chosen equal to 5. Our benchmark FBMC model corresponds, therefore, to a hybrid preventive-curative model with $n = 5$.

Let us now analyse the impacts of transmission switching on this benchmark. Figure 2.4 presents the box plot of the hourly total cost of the flow-based market coupling benchmark under different assumptions about the timing of transmission switching (proactive or reactive switching).

We do not observe any significant difference between proactive and reactive switching. On the other hand, the introduction of switching improves substantially the efficiency of operations, as compared to FBMC without switching. The annual savings of using transmission switching are evaluated at 294 M€/year, which corresponds to a 3.0% reduction in total (day-ahead and real-time) costs.

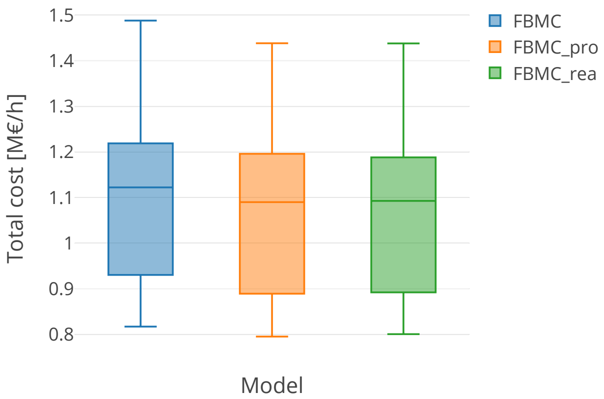


Figure 2.4: Hourly total cost of the different cases studied on 32 snapshots of CWE. "FBMC" refers to the case of a cost-minimizing real-time model without switching. The suffix "_rea" is for reactive (day-ahead) switching, "_pro" is for proactive (day-ahead and real-time) switching.

2.5.3 Sensitivity against the security criterion model

In this section, we discuss the sensitivity of the results with respect to our assumptions about N-1 robustness. As we discuss above, different assumptions on whether the TSO resorts to preventive or curative remedial actions lead to different FBMC models. Based on this distinction, in section 2.5.2, we propose a benchmark that corresponds to a hybrid preventive-curative model. The size of the preventive set U_{prev}, which we denote by n, needs to be determined so as to maintain a tractable day-ahead market clearing model when switching is considered. As we show in section 3, the difference between proactive and reactive switching is negligible. We therefore focus on the case of reactive switching, which is computationally less demanding, and analyse the evolution of the total cost with respect to n.

Figure 2.5 presents the evolution of the average hourly total cost on the 32 snapshots, as a function of n. We observe that the different choices of n exhibit similar performance, both for the situation with as well as without switching. The largest difference in total cost is observed for the case without switching, and amounts to less than 0.5%.

The results presented in Figure 2.5 correspond to the results for the so-called *N case*, i.e. the case where no contingency occurs in real time. As the models with different n differ in how they cope with N-1 robustness, we are also interested in analyzing an *N-1 case* in real time, i.e. a situation where a line fails between the day ahead and real time. We focus our analysis on a contingency that is hard for each model. For this purpose, we analyse the contingencies that are generated during the column-and-constraint generation

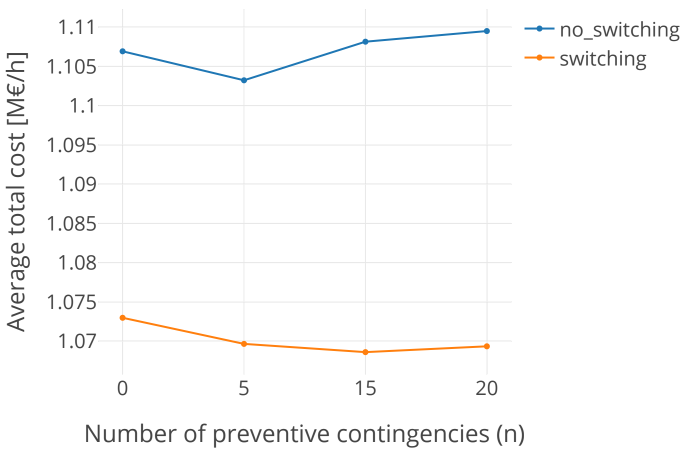


Figure 2.5: Evolution of the average hourly total cost on the 32 snapshots, as a function of the size of set U_{prev} in the N case. "no_switching" refers to the case of a cost-minimizing real-time model without switching, while "switching" corresponds to reactive switching.

algorithm that is described in section 2.4. We identify a specific contingency that appears to be consistently severe. This line corresponds to a cross-border line between Avelgem in Belgium and Avelin in France.

In Figure 2.6 we present the equivalent of Figure 2.5 for the N-1 case. The efficiency gap between purely curative N-1 robustness and the hybrid case with $n = 20$ now increases to 0.8% in the case without switching, and to 1.2% in the case with switching. We note that cases $n = 15$ and $n = 20$ are almost identical. This suggests that considering 20 lines in the definition of set U_{prev} may be sufficient, since the results appear to reach a stable behavior with 15 preventive contingencies. We further note that the benefits of reactive switching in FBMC are more important than in the N case, and amount to 3.5%.

2.5.4 Benchmarking of the results against a nodal market

In this section, we benchmark our results against a nodal market model with Locational Marginal Pricing (LMP). The model maintains the same two-stage structure as the zonal model that we have analyzed so far: unit commitment for slow units is determined in the day ahead, and re-dispatch and balancing are decided in real time. The difference with the zonal model is that the day-ahead unit commitment is now determined under a full nodal network model that includes all transmission constraints. Similarly to [ALPS21], we define N-1 security for LMP markets as the ability of a system to withstand any single-element transmission contingency, while maintaining its current nodal

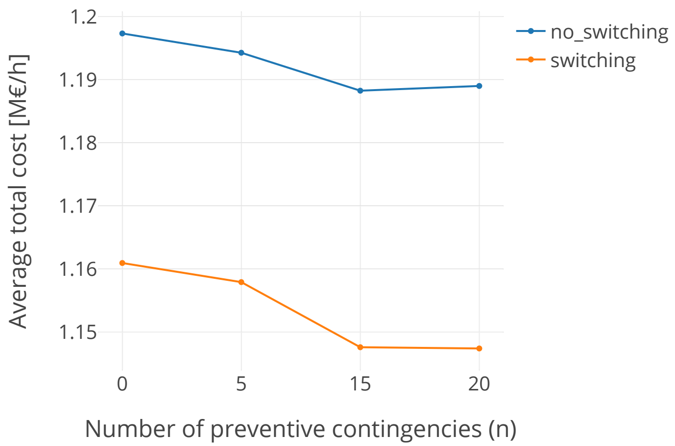


Figure 2.6: Evolution of the average hourly total cost on the 32 snapshots, as a function of the size of the set U_{prev} in the N-1 case. “no_switching” refers to the case of a cost-minimizing real-time model without switching, while “switching” corresponds to reactive switching.

injections and without violating any security constraints. Unlike in zonal markets, proactive transmission switching is currently not applied in practice in nodal electricity markets (e.g. the US market) to the best of our knowledge. Therefore, we simulate only reactive switching for the LMP benchmark. We use the same cutting-plane algorithm as the one developed in [ALPS21] for clearing the N-1 secure nodal unit commitment model.

Hereafter, we compare the results of LMP and the most efficient FBMC model (i.e. reactive switching with hybrid dispatch and $n = 20$) in the N state (i.e. no contingency). Figure 2.7 presents the box plot of the 32 snapshots for both LMP and FBMC, with and without reactive switching. As there is a significant difference between the mean and the median for the FBMC model, we also display the mean in a dashed line.

The first observation based on these results is that, when no contingency occurs, transmission switching contributes partially towards recovering the gap between nodal and zonal pricing. This gap is evaluated in our analysis at 2.1% (difference of “LMP” and “FBMC”) when there is no reactive switching, and decreases to 0.9% (difference of “LMP_rea” and “FBMC_rea”) with reactive switching. This translates to annual savings of 208M€ and 85M€ respectively. We thus observe that the benefits of switching are greater in the zonal setting than in the nodal setting, which is aligned with intuition.

The situation differs in the case where a contingency occurs in the system. Figure 2.8 is the equivalent of Figure 2.7 when a transmission line contingency has occurred. As in the case of section 2.5.3, we consider a failure of

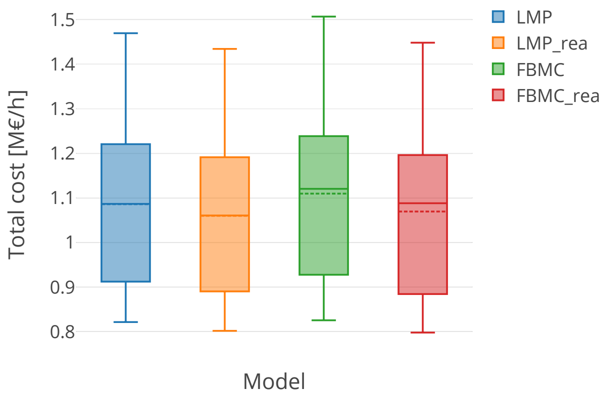


Figure 2.7: Hourly total cost of the different cases studied on 32 snapshots of CWE in the N case. "FBMC" refers to the case of a cost-minimizing real-time model without switching. The suffix "_rea" is for reactive switching.

the cross-border line between Avelgem (Belgium) and Avelin (France). Under this contingency, the gap between LMP and FBMC increases to 3.2% without switching, and to 2.2% with reactive switching. This suggests that the dispatch obtained with the nodal model is more robust to contingencies than that obtained by FBMC.

2.5.5 Additional sensitivities

Sensitivity on TSO coordination

In order to represent the possibility that the real-time operations of TSOs may not be perfectly coordinated, we fix day-ahead net positions of each zone to the result of the day-ahead market, and assume that each TSO is responsible for identifying re-dispatch and balancing actions that relieve congestion, *while maintaining the day-ahead net position* of its zone.

Note that this assumption is in line with the European viewpoint that considers the day-ahead market as the spot market, and relegates re-dispatch and balancing to a set of services that are deployed for supporting the day-ahead market positions. In recent years, this view has (fortunately) been relaxed with the emergence of a liquid intra-day market in Central Western Europe and with the move towards integrated pan-European platforms for balancing. As we demonstrate in the following paragraph, the view of treating the day-ahead market as the spot market for trading is detrimental towards efficiency, and there is therefore great value in coordinating inter-zonal dispatch closer to real time.

In Figure 2.9 we present the results of re-dispatch and balancing both for

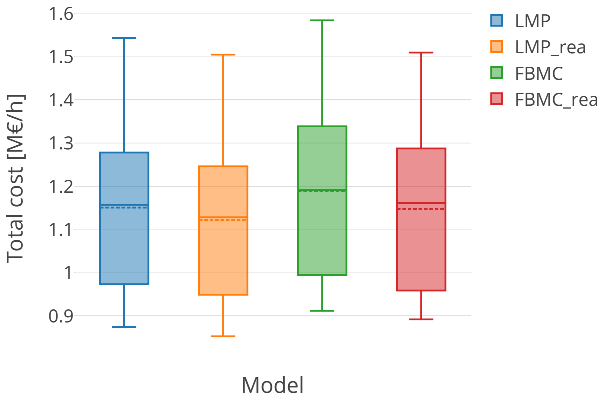


Figure 2.8: Hourly total cost of the different cases studied on 32 snapshots of CWE in the N-1 case. "FBMC" refers to the case of a cost-minimizing real-time model without switching. The suffix "_rea" is for reactive switching.

the case where the net positions are free to deviate, and also for the case where they are fixed to their day-ahead values. Two interesting observations can be made, based on Figure 2.9. The first one is that TSO coordination has considerable value. In the case without any switching action, the annual benefits of coordination are evaluated at 596M€. The second interesting observation is that the benefits of switching are significantly more important when the net positions are fixed. Whereas the cost decrease due to reactive switching with a budget of 6 lines amounts to 3.0%, this value increases to 6.6% if we assume that the net positions of the day-ahead must be maintained in real time. The intuition here is that when net positions are fixed, the TSO cannot take advantage of cross-border re-dispatch and has consequently less degrees of freedom for restoring a feasible dispatch. The additional degree of freedom that switching provides has, therefore, a greater relative importance.

Deviations from cost-minimization

In contrast to the transmission-constrained economic dispatch which takes place in real-time US operations, assuming that TSOs use a perfect cost-minimization in real time may not be the case in practice in European system operations, both due to the fact that (i) certain European TSOs do not use optimization algorithms in real time, but also because (ii) real time is not necessarily perceived as an "appropriate moment" for economic trade to take place. We thus simulate the two following variants. (i) In order to represent the view which supports that real time should be used for balancing the system, and not enhancing economic trade, we simulate a volume-based model, where the objective is to minimize the deviation with respect to the day-ahead schedule.

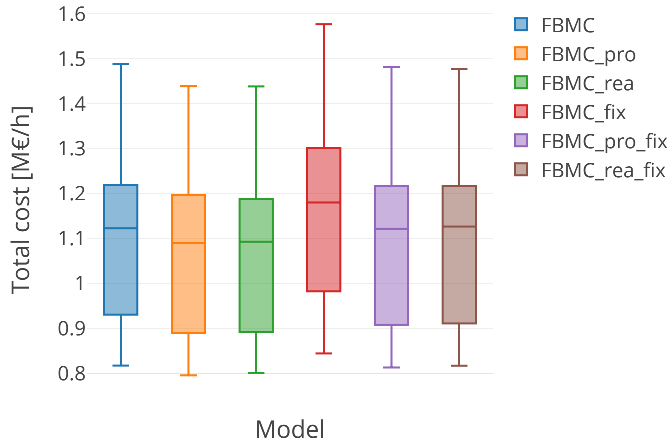


Figure 2.9: Hourly total cost of the different cases that are studied on 32 snapshots of CWE. The suffix "_fix" refers to the case where the net positions are fixed to their day-ahead values.

(ii) In order to represent the fact that certain TSOs do not use optimization algorithms, but rather heuristics, for determining real-time set-points, we simulate a PTDF-based heuristic for re-dispatching and balancing the system. Both models are presented in detail in appendix section 2.C.

Figure 2.10 presents the performance of a cost-minimizing real-time model with the performance of alternative methods. These results demonstrate that the assumptions about how re-dispatch and balancing are performed have a very significant impact on the analysis. Clearly, the perfectly coordinated cost-minimizing real-time model is the golden standard and outperforms the alternative methods by a large margin. It is worth pointing out that the relative advantage of using transmission switching is much less significant than the effect of using a real-time method that is aimed at operational efficiency. For instance, Figure 2.10 illustrates that the minimum volume model is almost insensitive to the method of switching that is used.

Switching more lines with a heuristic method

An important observation, that is discussed in section 2.5.4 and demonstrated in Figure 2.8, is that when a severe contingency occurs in the system, the gap between LMP and FBMC when reactive switching is allowed is still significant (more than 2%). As the benefits of switching have been found to be more important for FBMC, we might wonder whether this finding is sensitive to the number of lines that can be switched. Recall that we use a switching budget of 6 lines for this study. We solve the real-time models to a MIP

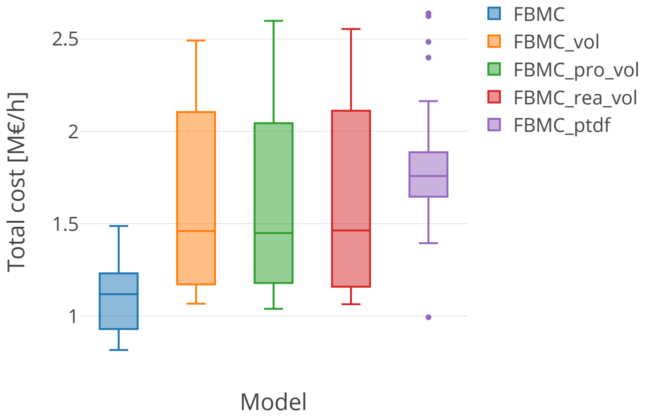


Figure 2.10: Hourly total cost of the different re-dispatch and balancing methods on 32 snapshots of CWE. In this figure, the "FBMC" series refers to the same series as the "FBMC" series of Figure 2.4, i.e. a cost-minimizing real-time model without switching. The suffix "_vol" is for the volume-minimizing real-time model, "_ptdf" is for the PTDF-based heuristic. Note that, for the latter, there is no switching action possible, as it is based on the PTDF obtained for a reference topology.

gap of 1% in order to keep the computation tractable. We now consider the LMP-based heuristic presented by Fuller et al. [FRC12] as an alternative real-time switching heuristic. We describe this heuristic in detail in section 2.B of the appendix. The number of lines that can be switched off is indicated by parameter Max_iter in Algorithm 3 of Appendix 2.B. We set this parameter to 40, thereby allowing up to 40 lines to be switched. This parameter is validated *a posteriori* by checking that it is never binding, i.e. that the best result is obtained at an iteration strictly less than 40.

Figure 2.11 demonstrates that it can indeed be beneficial to switch more than 6 lines. For FBMC, we evaluate this benefit at 50M€ annually, which corresponds to less than half a percent. However, the LMP-based market clearing still outperforms FBMC with reactive switching in the N-1 case when both models use Fuller's switching heuristic. The efficiency gain of LMP remains at 2.2%, the same as with the budget method.

2.6 Conclusion

In the first part of this chapter, we propose a two-stage model of a zonal electricity market with transmission switching at both the day-ahead and real-time stage. We cast the problem as a robust optimization problem (ARO) with mixed integer recourse, and we describe a novel algorithm for solving

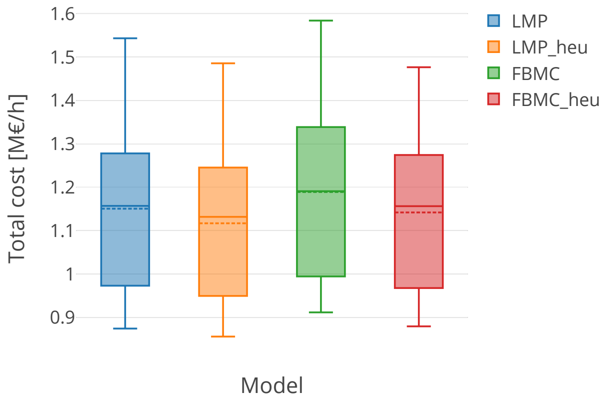


Figure 2.11: Hourly total cost of the different cases studied on 32 snapshots of CWE. The suffix "_heu" refers to the results when Fuller's heuristic is used, as described in Algorithm 3 in appendix.

ARO problems with mixed integer recourse that respect a certain structure. We apply the algorithm to the case of day-ahead market clearing with proactive line switching.

In the second part of the chapter, we propose a benchmark FBMC model, and we analyze the impacts of both proactive and reactive transmission switching on the operating costs of a realistic case study of the Central Western European system. We then perform a detailed sensitivity analysis in order to identify the sensitivity of our results on various assumptions related to short-term electricity operations. In particular, we consider (i) the influence of preventive versus curative security practices, (ii) the impact of contingencies, (iii) the level of TSO coordination and (iv) deviations from real-time cost minimization.

We summarize below the main observations that we can draw from our case study:

- The performance of proactive and reactive switching are similar.

- The number of contingencies that are considered in a preventive way in the day-ahead market clearing problem influences significantly the total cost when a contingency occurs in real time. This improvement is evaluated at 1.2% for reactive switching.

- Transmission switching is more beneficial for FBMC than for LMP. Considering transmission switching thus contributes towards recovering partially the efficiency gap between zonal and nodal market clearing. This gap is estimated at 1% in the N case, but increases to 2.2% in the N-1 case under the occurrence of a severe contingency.

- The impact of TSO coordination is significant, and is more important in the absence of transmission switching.
- Performing re-dispatch and balancing without aiming at operational efficiency may eclipse the potential efficiency gains of transmission switching.

We have not discussed the potential pricing and policy issues that can arise as a consequence of the additional non-convexities that are introduced to the market clearing procedure by transmission switching. These, however, are important questions that were already partially discussed in [OHK+10]. Further research is needed in order to develop a viable framework for quantifying the impact of transmission switching on market clearing prices.

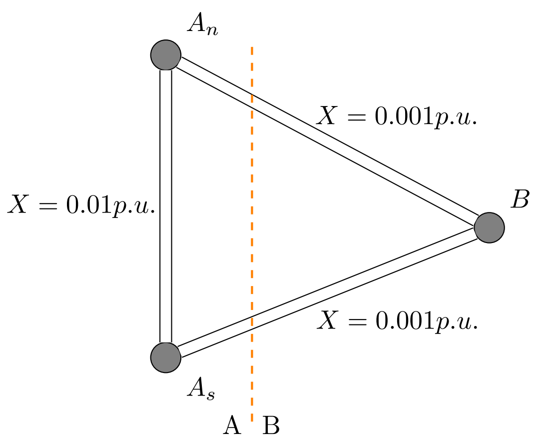


Figure 2.12: Network data of the three-node two-zone example that is presented in section 2.A.

2.A Illustration of the difference between preventive and curative re-dispatch

In this appendix, we describe an illustrative example that highlights the difference between a day-ahead model with preventive versus curative re-dispatch. We consider a three-node two-zone network, as shown in Figure 2.12. Zone A consists of two buses, An and As. Zone B consists of a single bus, bus B. Buses An and As are connected by two lines, each with a reactance of 0.01 per unit (p.u.). Bus An is connected to bus B by two lines, each with a reactance of 0.001 p.u. Bus As is connected to bus B by two lines, each with a reactance of 0.001 p.u. All lines obey a capacity limit of 1 GW. In order to simplify the exposition, we consider only contingencies that involve cross-zonal lines.

In the case of purely **curative re-dispatch**, the net position of zone A could reach up to 3 GW. Indeed, if a contingency occurs on a line between bus An and bus B, the transmission limits on the remaining elements can still be respected by a net injection of 1 GW in bus An and a net injection of 2 GW in bus As. If a contingency occurs on a line between bus As and bus B, the transmission limits on the remaining elements can still be respected by a net injection of 2 GW in bus An and a net injection of 1 GW in bus As.

However, in the case of purely **preventive re-dispatch**, the nodal net injections cannot be modified when the contingency occurs. In that case, it is not possible to inject more than 1083 MW in bus An and more than 1083 MW in bus As. In the day-ahead market clearing problem, the maximum net position of zone A is thus limited to 2.17 GW (preventive), and not to 3 GW (curative).

2.B Heuristic switching algorithm

We use the LMP-based heuristic presented by Fuller *et al.* [FRC12] as an alternative real-time switching heuristic. is summarized in Algorithm 3.

Algorithm 3 LMP-based switching heuristic

---

1. Set $k = 0$ and $\mathcal{O}_k = \emptyset$, where $\mathcal{O}_k$ is the set of lines that are disconnected, at iteration k.

2. Obtain ranking parameter α_l for all lines $l \in L$ as follows:

   - Fix the topology as $z_l = 0$ if $l \in \mathcal{O}_k$ and $z_l = 1$ otherwise and solve the corresponding real-time model, which is an LP.
   - Let π_n be the LMP of node n. Let $\alpha_l = \pi_{m(l)} - \pi_{n(l)}$

3. Solve the following switching problem, with a subset of the lines fixed:

$$
\begin{aligned}
\min_{\substack{v\in[0,1]\\ f,\theta,t}} \; & \sum_{g\in G} P_g Q_g v_g \\
\text{s.t.} \; & \sum_{g\in G(n)} Q_g v_g - \sum_{l\in L(n,\cdot)} f_l + \sum_{l\in L(\cdot,n)} f_l = Q_n, \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall n \in N \\
& -F_l t_l \le f_l \le F_l t_l, \quad \forall l \in L \\
& f_l \le B_l(\theta_{m(l)} - \theta_{n(l)}) + M(1 - t_l), \quad \forall l \in L \\
& f_l \ge B_l(\theta_{m(l)} - \theta_{n(l)}) - M(1 - t_l), \quad \forall l \in L \\
& z_l = 0, \quad \forall l \in \mathcal{O}_k \\
& z_l = 1, \quad \forall l \text{ s.t. } \alpha_l \ge 0 \\
& \sum_{l\in L} (1 - z_l) \le 1 + \text{length}(\mathcal{O}_k)
\end{aligned}
$$

4. Let $k \leftarrow k + 1$. If $k \ge Max\_iter$, stop. Else, go back to step 2.

2.C Non cost-based re-dispatch and balancing

2.C.1 Volume-based heuristic

The volume-minimizing model can be straightforwardly described with the same constraints as the cost-minimizing model that is presented in section 2.3. Instead, the objective function now corresponds to minimizing the total deviation from the day-ahead dispatch. The motivation for this model is the fact that European market operations prioritize the balancing of portfolios as real time approaches. The model can be described as follows:

$$
\begin{aligned}
\min_{\substack{v\in[0,1]\\ f,\theta,\\ t\in 0,1}} \; & \sum_{g\in G} Q_g \left|v_g - v_g^{\mathrm{DA}}\right| \\
\text{s.t.} \; & \sum_{g\in G(n)} Q_g v_g - \sum_{l\in L(n,\cdot)} f_l + \sum_{l\in L(\cdot,n)} f_l = Q_n, \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall n \in N \\
& -F_l t_l \le f_l \le F_l t_l, \quad \forall l \in L \\
& f_l \le B_l(\theta_{m(l)} - \theta_{n(l)}) + M(1 - t_l), \quad \forall l \in L \\
& f_l \ge B_l(\theta_{m(l)} - \theta_{n(l)}) - M(1 - t_l), \quad \forall l \in L
\end{aligned}
$$

Here, v_g^{DA} is the acceptance/rejection of generator g in the day-ahead process. It is fixed as a parameter for the real-time model.

2.C.2 PTDF-based heuristic

The idea of the PTDF-based heuristic that we propose for approximating the fact that certain TSOs do not employ optimization in real time is to use the PTDF matrix of the system in order to identify, for each overloaded line, the generators that contribute the most to the flow on this line. For these generators, we can compute the amount of reduction in production that is necessary in order to alleviate congestion on the line. By doing so, it is possible that other lines will be affected. We thus iterate over this process until no more lines are congested, as shown in Algorithm 4.

Algorithm 4: CRH (Congestion Removal Heuristic)

Input: initial dispatch v

Output: new dispatch that respects network constraints

let L_{cong} be the set of congested lines, sorted by congestion magnitude
while $L_{cong} \neq \emptyset$ **do**
    for *every* $l \in L_{cong}$ **do**
        let N_{sorted} be the set of nodes sorted w.r.t. $\mathrm{PTDF}_{l,n}$
        for $n \in N_{sorted}$ *until* $f_l \ge F_l$ **do**
            for $g \in G(n)$ *until* $f_l \ge F_l$ **do**
                $v_g = \max\{v_g - \frac{(f_l - F_l)}{\mathrm{PTDF}_{l,n}}, 0\}$
            restore power balance
    update L_{cong}

Part II

Long-term efficiency of zonal pricing

3 An analysis of zonal pricing from a long-term perspective

3.1 Introduction

The energy transition will require considerable investment in various technologies located throughout Europe. Except for remaining subsidies to particular technologies that are progressively dismantled, this investment process is meant to be driven by market forces. This means that investors will invest when and where their capacities are profitable. The condition for investment is nothing more than the standard principle that the present value of the cash flow accruing to a generation facility over its lifetime in a certain location should cover its overnight investment cost in that location. In this dissertation, we suppose that we have restated the investment criterion in its standard single-period expression that the annual cash flow accruing to the plant should cover the annualized investment cost.

The peculiar aspects of the power sector have required extensive discussions since the early days of the restructuring to competition. Some of these discussions are reflected in market designs and have implications on the cash flow generated by power plants. In other words, the choice of a market design influences the cash flows accruing to specific equipment and, because of the relation between investment cost and cash flows, the market design also influences the structure of the capital stock of the system and consequently the cost of the energy transition. The implication of the market design on the cash flows accruing to plants is thus an important question in the energy transition.

The relevance of the market design on investment is indeed well recognized in the notion of “missing money” that has now been discussed since more than a decade. Forward capacity markets, energy-only markets, and energy-only markets supplemented by operating reserve demand curves and strategic reserve (in some EU countries) are variations of market designs that are aimed at producing cash flows that are sufficient to cover investment cost. The underlying reasoning in these discussions is that prices based on short-run variable (essentially fuel) costs, which typically lack a scarcity premium, are not equal to short-run marginal costs (including the cost of unserved energy) and do not lead to cash flows that are sufficient for covering investment costs. This reasoning has been elaborated by various authors, with [Jos07] offering a particularly insightful presentation of the "missing money" problem. The paper also explains how this shortcoming stems directly from an early fundamental result of power system economics, that is attributed to [Boi60] and [Boi64], and was later quoted extensively in the literature. This result is of particular interest for our discussion, and can be stated very simply: short and long-run marginal costs (where short-run marginal costs include a scarcity premium and long-run marginal cost corresponds to the cost of investment) should be equal in an optimally designed system. The proposition implies that electricity prices based on marginal fuel costs are by construction unable to cover investment costs. Thus, prices based on marginal fuel costs distort investment, and therefore, require special measures in order to drive optimal investment.

The proposition is derived under standard convexity assumptions. This is the usual context in which capacity expansion problems are discussed. The result was developed by [Boi60] for a monopoly system where energy is, by regulation, priced at marginal cost. It directly applies to a restructured power system, where, because of market design, energy would be priced at marginal cost. Barring for indivisibilities related to unit commitment issues, this is the basic principle that underpins market restructuring. Summing up, cash flows based on marginal fuel cost, complemented by an appropriate instrument for removing the missing money, would provide the adequate cash flow for covering investment costs in the ideal world of perfect competition.

The special features of electricity systems make it difficult to submit the sector to standard competition. The market design is meant to achieve this task with some degree of approximation. Market designs can be different, and the question then arises whether they all reflect the same cash flows, and hence an economically viable environment for the same set of technologies. If this is not the case, then they will naturally not cover the same investment costs and hence will not lead to the same investment. The goal of this chapter is precisely to understand the impacts of the difference in market designs on investment, with a focus on market designs that implement nodal and zonal systems.

It is relatively straightforward to model capacity expansion based on nodal systems and analyse the relationship between cash flows and investment. Indeed, one notes that it is straightforward to show that Boiteux's result that relates long and short-run marginal cost in an optimized system extends using the same methodology to short and long-term nodal pricing. The statement is then that one can find short and long-run nodal prices that are equal in a system with a geographically optimal capacity mix. The same cannot be said, however, about zonal systems. Unlike in nodal systems, there is no unique way of implementing zonal pricing and different designs have been proposed. Although it is possible to generalize Boiteux's result on a very specific variation of zonal pricing, it does not hold on flow-based market coupling (FBMC), the methodology currently implemented in a large part of the European market. To state it differently, the classical result of equivalence between the central planning approach and the decentralized market ceases to hold. This is what we discuss in detail in the first part of this chapter.

Existing literature on missing money has, to a large extent, omitted considerations related to congestion. [CS05] does recognize that capacity markets, justified as a response to the missing money problem, need to have a locational component, but their treatment does not go beyond that. Past research has instead focused on proposing and assessing remedies to the missing money problem. One can identify from this literature two main classes of remedies: capacity remuneration mechanisms, advocated for instance in [CS05, FP08], and scarcity pricing based on operating reserve demand curves [Hog13]. There are, nowadays, still active discussions about these remedies and their ability to solve the missing money problem with sometimes contradictory results: [MT19] find that capacity markets can mitigate the missing money problem, while [New16] argues that they tend to exacerbate it.

The consideration of transmission constraints in the classical capacity expansion problem leads us to identify two new types of missing money problems:

1. The first missing money problem relates to zonal pricing in general and originates from the simplification of the transmission constraints. It is the effect that leads to a lack of investment in a system with zonal pricing compared to one with nodal pricing.

2. The second missing money problem is specific to zonal pricing with FBMC and leads to the breakdown of the equivalence between its centralized and decentralized formulations.

These two new types of missing money problems are different from the classical missing money problem that we mentioned above and that has been discussed extensively in the literature. In order to focus on these new inefficiencies, we abstract from the discussions on the classical missing money problem by assuming a perfect scarcity pricing mechanism based on the VOLL that entirely represents consumers' willingness to pay. This assumption leads to optimal investments when there is no congestion, or in the case of nodal pricing, and enables us to isolate the effects of the two new missing money problems that we identify in this work.

The chapter is organized as follows: we start by reviewing existing literature on the long-term impacts of a zonal design on investment in section 3.2. In section 3.3, we review in a uniform notation the long-term market equilibrium under nodal pricing and under the above-mentioned specific form of zonal pricing that allows its formulation as a Nash equilibrium. Then, in section 3.4 we describe how zonal pricing with FBMC can be modelled and we discuss the long-term equilibrium both from the perspective of a central planner and in a decentralized market. Finally, we perform a comparison of the different policies that we model in the chapter on a reduced version of our instance of the Central Western European (CWE) system. Section 3.6 provides a brief conclusion.

3.2 Literature review and contributions

3.2.1 Literature review

Recently, a stream of literature has emerged that studies the long-run effects of zonal pricing. A first series of papers is focused on transmission and generation investment in a zonal environment. The first paper of this series is [GMS+16]. In this paper, the authors propose a model of investment in the network by the TSO and in generation by private firms, by explicitly accounting for both the market interaction between unbundled transmission and generation companies and a zonal pricing model. The authors also analyse the impact of different network fee regimes for the recovery of network costs. In this paper, the focus is not on a careful modeling of zonal transmission constraints, instead a simplified zonal version of Kirchhoff's first law is used. It is assumed that inter-zonal lines can be used up to their full capacity and the model ignores intra-zonal lines. The model proposed in [GRSZ21] is similar: the structure remains the same, with a tri-level model that accounts for network investment by the TSO in the upper level, generation investment by private firms, and re-dispatch by the TSO at the lowest level. The main difference with [GMS+16] is the size and realism of the case study, which is now calibrated to the German electricity market. This allows the authors to draw conclusions on the effect of certain market improvements (market splitting, curtailment of renewable energy and redispatch-aware network investment) on the efficiency of operation. We note that the way in which zonal transmission constraints are represented in this second paper differs from [GMS+16]. Here, ATC market coupling is assumed with exogenous ATC values. A third paper in this stream of work is [EGK+21], in which the authors extend the models previously developed in order to model cross-zonal effects on the interaction between the regulator and private firms.

A second series of papers that considers both zonal pricing and long-term effects is targeted at studying the optimal delimitation of bidding zones. A notable contribution in this area is [GMWZ16], where the authors highlight the importance of accounting for long-term effects when considering the delimitation of bidding zones. The paper shows, using small illustrative examples, that more price zones might decrease welfare in the long run, which could seem counter-intuitive. The authors argue that more price zones could imply over-investment of generation capacity that would not be able to produce in real time, due to congestion that was omitted in the spot market. A subsequent paper [GKL+19] is focused on methods for solving the large tri-level mixed-integer mathematical program, which is how the studied problem is formulated by the authors. Two solution approaches are proposed: first, the reformulation of the problem as a single, but large, mixed-integer quadratic program; second, a tailored version of generalized Benders decomposition. The generalized Benders decomposition approach is then applied on a realistic but simplified representation of the German network in [AGK+20], in order to derive certain insights on the splitting of bidding zones in Germany. In this second series of papers, all contributions that we have mentioned so far are based on similar assumptions and structure: the authors employ multi-level models where a TSO or a regulator plays first, assuming perfect knowledge of the outcome of the capacity expansion by private firms. By contrast, in [FHK+21] the authors study the impact of a German zone split using an agent-based simulation model, where the regulator and market participants interact under imperfect information. The model is applied to a detailed instance of the German electricity grid in a multi-period setting that also considers auxiliary nodes in neighboring countries, in order to account for cross-border effects. The authors find that, under a split of the German bidding zone, congestion management costs would decrease by 2025 but slightly re-increase by 2035, due to the fact that the bidding zone delimitation would become outdated by then. This leads the authors to suggest that bidding zones should be adjusted regularly.

3.2.2 Contributions

In this section, we specify our contributions with the present chapter and describe how our models are positioned relative to the ones proposed in the existing literature.

In terms of the modeling of zonal pricing, our work contributes to the state of the art in the two following ways: First, we extend the model of FBMC proposed in [ALPS21] in order to account for generation investment by private firms. As we show in section 3.4, the specific methodology of FBMC introduces several challenges when viewed from a long-term point of view. Second, and in order to highlight the challenges that are associated with FBMC, we introduce a new model of zonal transmission constraints that is not subject to the same challenges. This model, that we refer to as zonal pricing with Price Aggregation (PA), is obtained by going back to the fundamental idea of zonal pricing which is that prices within the same zone should be the same. It is introduced in section 3.3.2.

In terms of modeling the long-term effects of the zonal design, our work differs from existing literature by modeling the interactions between investment by private firms and zonal transmission constraints. This is achieved in the present chapter by employing the model of FBMC with exact projection (FBMC-EP) which, as we discuss in section 1.2.2 of the introduction to the thesis, is independent of exogenous parameters. This enables us to identify a new inefficiency that occurs in FBMC when viewed from a long-term perspective, which is a key element of our work. Instead, existing papers on the subject either use simplified zonal transmission constraints or are based on exogenous parameters that prevent them from measuring the above-mentioned inefficiencies. Another key difference between existing literature and the present work relates to the structure of the model that we employ. As discussed above, existing papers either use multi-agent or multi-level models. In the latter case, the authors adopt the assumptions that some agent, in general the regulator or the TSO, will act as a leader of the game. We follow instead a formulation of capacity expansion models where all players act simultaneously, which is common in the literature on capacity expansion [ES11, Ozd13].

Finally, we mention here two features that we do not consider in this work: transmission line investment by the TSO, which is accounted for in the first stream of papers cited above, and endogenous bidding zone delimitation, which is the focus of the second stream.

3.3 Capacity expansion models in transmission-constrained electricity markets

Quoting Paul Joskow in [Jos06], “the goal of a well functioning market should be to reproduce the ideal central planning results”. More precisely, if we assume a perfectly competitive market, the key question in market design is whether there exists a set of prices that would lead price-taking profit-maximizing agents to reproduce the centralized solution in a decentralized way. In the context of capacity expansion in electricity markets, one can deduce the set of prices that reproduce the centralized results from the theory of marginal cost pricing, subject to a careful interpretation of this theory: the cost, here, has to include the long-term development cost [Boi60]. These pricing principles extend easily to transmission-constrained electricity markets. This is what we discuss in the present section, first in the case of nodal pricing, and then for zonal pricing.

3.3.1 Nodal pricing

Considering that the central planner accounts for all transmission constraints, in the form of the DC approximation, one can derive Locational Marginal Prices (LMP) that recover the optimal long-term solution in a decentralized way. In order to demonstrate this in a formal setting, let us recall the definition of the set of all net injections at the network buses that are feasible for the DC power flow equations, denoted by $\mathcal{R}$:[1]

$$\begin{aligned} \mathcal{R} = \Big\{ r \in \mathbb{R}^{|N|} \ \Big| \ & \exists f \in \mathbb{R}^{|K|} \ : \\ & f_k = \sum_{n \in N} PTDF_{kn} \cdot r_n, k \in K \\ & \sum_{n \in N} r_n = 0, -TC_k \leq f_k \leq TC_k, k \in K \Big\} \end{aligned} \tag{3.1}$$

The notation in this set of equations is as follows: f_k is the power flow on line $k \in K$, r_n is the net injection at node $n \in N$, $PTDF_{kn}$ is the power transfer distribution factor of line k and node n, and TC_k is the thermal limit

[1] For the sake of simplicity of the analysis, we only consider here pre-contingency transmission constraints. We note, however, that all our models can easily be extended to the case of N-1 robustness [ALPS21].

of line k. Note that this formulation of $\mathcal{R}$ is completely equivalent to the one that we introduced in section 1.2.2 of the introduction, in equation (1.1). The only difference is that we have projected the voltage angles out of the formulation and replaced it with a generic linear relation between the flows and the net injections, represented by the PTDF matrix. The reason why we use the PTDF formulation in this chapter is that, unlike in chapter 2, we consider here that the topology is fixed. Therefore, the voltage angle variables are not necessary to describe the transmission constraints and can be projected out of the formulation through the PTDF matrix. The set $\mathcal{R}$ describes completely the network constraints in the case of nodal pricing. Using this set, one can define the capacity expansion model from the central planner perspective as follows:[2]

$$\min_{x,y,s,r} \sum_{i \in I, n \in N} IC_i \cdot x_{in} + \sum_{i \in I, n \in N, t \in T} MC_i \cdot y_{int} + \sum_{n \in N, t \in T} VOLL \cdot s_{nt} \tag{3.2a}$$

$$(\mu_{int}) : y_{int} \leq x_{in} + X_{in}, i \in I, n \in N, t \in T \tag{3.2b}$$

$$(\rho_{nt}) : r_{nt} = \sum_{i \in I} y_{int} + s_{nt} - D_{nt}, n \in N, t \in T \tag{3.2c}$$

$$r_{:t} \in \mathcal{R}, t \in T \tag{3.2d}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{3.2e}$$

Here, IC_i is the annualized investment cost of technology $i \in I$, x_{in} is the investment in technology i at node n, MC_i is the marginal cost of technology $i \in I$, y_{int} is the power production by technology i in node n and period $t \in T$ where T is the set of hours in the year, $VOLL$ is the value of lost load, s_{nt} is the demand curtailment at node n in period t, X_{in} is the existing installed capacity of technology i at node n and D_{nt} is the demand at node n in period t. Dual variables are indicated between parentheses to the left of the associated constraints. The objective of the central planner in this optimization problem is to minimize total cost, which includes investment and operating costs, while respecting the operational constraints (3.2b), the network constraints (3.2d) and nodal balance (3.2c). In this model, the optimal values of ρ_{nt} correspond to the optimal LMPs that, as mentioned above, allow for a decentralized solution to the problem in a market context, as shown in [Ozd13]. The decentralization is obtained when assuming perfect competition in a market with 4 types of agents: producers, consumers, the TSO and an auctioneer that ensures market clearing. The market is modeled as a Nash equilibrium of the simultaneous game between the 4 groups of agents who are price takers and maximize their profit[3]. In particular, the TSO maximizes the value of its grid, i.e. the congestion rent, in line with the literature on markets

[2] Although the assumption of infinitesimal generation expansion is acceptable, it will likely not be a useful assumption for lumpy transmission expansion.

[3] We deal here with a special type of game that represents the equilibrium in a competitive market with utility functions that are quasilinear with respect to the payments. As a result,

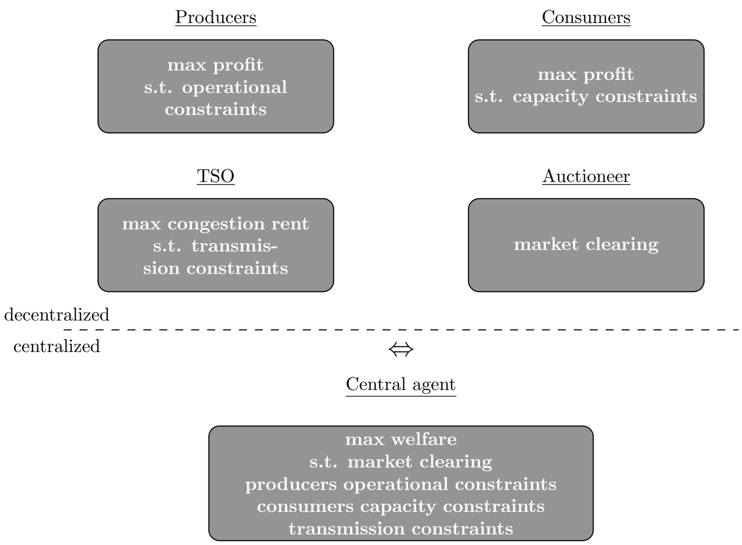


Figure 3.1: Equivalence between the decentralized game of the 4 groups of agents in the market and the welfare maximization problem of the central agent.

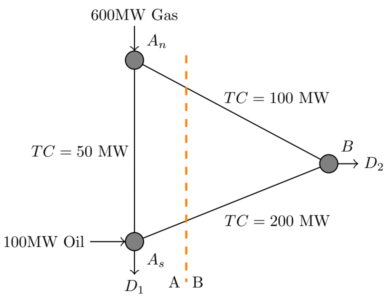
with transmission operations [Hog92, BS01, Ozd13]. Regarding consumers, we assume that electricity is priced at VOLL by the regulator in case of demand curtailment. Figure 3.1 represents the profit maximization problem of the 4 groups of agents and their relationship with the welfare maximization problem of the central planner. We do not describe the full set of conditions that characterizes the decentralized market-based model here, but we simply highlight one important complementarity condition from the set of KKT conditions of this problem:

$$0 \leq x_{in} \perp IC_i - \sum_{t \in T} \mu_{int} \geq 0 \;\; \forall i \in I, n \in N \tag{3.3}$$

This equation implies that an investment will be made in technology i at node n if the investment cost can be covered by scarcity rents μ_{int}, which are equal to the difference between the marginal cost and the price when the plant produces at its maximum capacity, and 0 otherwise. The formal decentralization interpretation is detailed in appendix 3.A.

Now that we have formally defined the capacity expansion problem under nodal pricing, let us examine its outcome on a small illustrative example. We will use this example throughout the chapter in order to illustrate the different models that we present. The data of the example is presented in Figure 3.2. The instance is a three-node, two-zone system with a load duration curve that

the Nash equilibrium is the solution to an LP and the solution set is convex. This is a specific situation that is in contrast with general games where there can be multiple disjoint Nash equilibria.



Technology	MC [€/MWh]	IC [€/MWh]
Coal	25	16
Gas	80	5
Nuclear	6.5	32
Oil	160	2

D_1 [MW]	D_2 [MW]	Duration [h]
0	7086	1760
0	9004	5500
300	10869	1500

VOLL = 3000€/MWh

Figure 3.2: Three-node two-zone network used in the illustrative example.

is aggregated into three demand blocks of different duration. The two nodes on the left belong to the same zone and contain existing capacity with 600 MW of gas in the upper node and 100 MW of oil in the lower node. Zone B on the right consists of a single node and hosts most of the demand, with no existing capacity.

The optimal solution in this example is to install 1918 MW of coal, 7086 MW of nuclear and 1715 MW of gas capacity in node B, and to install 300MW of gas capacity in the lower node of zone A. One observes that the optimal solution carries more capacity than the demand. The first reason is congestion. Although there is significant gas capacity in node A_n, not all this capacity can be used for serving demand due to the limited capacity of the lines. In the peak period, one observes that only 150 MW out of the 600 MW are used. The second reason is the large marginal cost of the oil generator which makes it more interesting to invest in gas capacity in node A_s instead of using the Oil unit to cover the peak demand at A_s. To summarize, the nodal pricing solution amounts to an investment cost of 267,515€ and an operating cost of 114,033€ which yields a total cost of 381,548€.

3.3.2 Zonal pricing

Under the zonal pricing paradigm, the nodes of the network are aggregated into a set of zones and electricity is priced at the zonal level. Unlike in nodal pricing, there is no unique and unambiguous way of representing the network constraints in a zonal market. However, there is a natural zonal pricing model that emerges if we go back to the fundamental property of zonal pricing which is that there should be a unique price per zone. This natural model can thus be obtained by taking the dual of the nodal market clearing problem, imposing that all nodal prices within the same zone are equal and going back to the primal space. The result of this manipulation is that the control variable in the balance constraint is now a zonal net position, that we denote by p_z, which is simply obtained as the projection of the nodal net injections into the space of zonal net positions. The reader is referred to appendix section 3.C for the details of this derivation. We now denote by $\mathcal{P}^{\text{PA}}$ the set of all network constraints under the zonal pricing paradigm, which can be seen as the equivalent of set $\mathcal{R}$ in nodal pricing. The exponent PA stands for Price Aggregation and is used for distinguishing the model from other variations of the set of zonal net positions that we have introduced in the introduction of the thesis (section 1.2.2). Mathematically, $\mathcal{P}^{\text{PA}}$ can be defined as follows:

$$\begin{aligned}\mathcal{P}^{\text{PA}} =\Bigg\{ p \in \mathbb{R}^{|Z|} \,\Bigg|\, & \exists (f,r) \in \mathbb{R}^{|K|} \times \mathbb{R}^{|N|} : p_z = \sum_{n \in N(z)} r_n \;\; \forall z \in Z, \\ & f_k = \sum_n PTDF_{kn} \cdot r_n \;\; \forall k \in K, \sum_n r_n = 0, \\ & -TC_k \leq f_k \leq TC_k \;\; \forall k \in K \Bigg\}\end{aligned} \tag{3.4}$$

Note that, although every line of the network is accounted for in equation (3.4), and could potentially be binding, the control variables are the zonal net positions p_z. This implies that the dispatch within each zone is based solely on the merit order and it will be, in general, infeasible regarding the complete set of grid constraints. In particular, market clearing based on equation (3.4) is not equivalent to the ideal zonal pricing model proposed in [BJ01]. In fact, as discussed in [Wei17], ideal zonal pricing is quite different than any other zonal pricing model as it is not a relaxation of nodal pricing but, instead, adds constraints to the nodal market clearing problem. A dispatch obtained with ideal zonal pricing is guaranteed to be feasible. This implies that, unlike in other zonal pricing models, no re-dispatch is needed. This comes with a major drawback, which is that ideal zonal pricing might be infeasible, as mentioned in [ES05]. This is in contrast with our zonal PA model, which is a more classical zonal pricing model: existence of a market-clearing solution is guaranteed, but re-dispatch will in general be needed.

For the specific case of the illustrative example of Figure 3.2, the set $\mathcal{P}^{\text{PA}}$ can be made explicit as follows:

$$\begin{aligned}\mathcal{P}^{\text{PA}} =\Bigg\{ p \in \mathbb{R}^{2} \,\Bigg|\, & \exists (r,f) \in \mathbb{R}^{3} \times \mathbb{R}^{3} : \\ & p_A = r_{A_n} + r_{A_s}, p_B = r_B \\ & f_1 = \frac{1}{3} r_{A_n} + \frac{2}{3} r_{A_s}, -200 \leq f_1 \leq 200 \\ & f_2 = \frac{2}{3} r_{A_n} + \frac{1}{3} r_{A_s}, -100 \leq f_2 \leq 100 \\ & f_3 = \frac{1}{3} r_{A_n} - \frac{1}{3} r_{A_s}, -50 \leq f_3 \leq 50 \\ & r_{A_n} + r_{A_s} + r_B = 0 \Bigg\}\end{aligned} \tag{3.5}$$

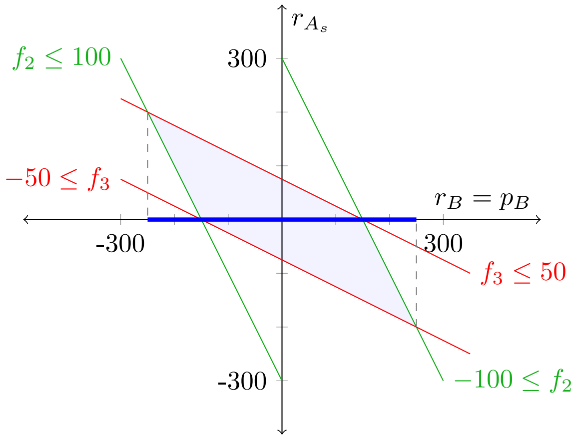


Figure 3.3: Illustration of the feasible set of net injections in the nodal model (light blue area) and the feasible net positions in the zonal model (thick blue line) for the 3-node 2-zone network. The feasible set of zonal net positions in the PA model is a projection of the set of feasible nodal net injections on the space of zonal net positions.

Note that, because $r_{A_n} + r_{A_s} + r_B = 0$, the feasible set of nodal net injections has only 2 independent dimensions. It can thus be represented in a 2D space. This is what we do in Figure 3.3, where the line capacity constraints and the feasible set of net injections are shown on the (r_{A_s}, r_B) space. Similarly, the feasible set of zonal net positions $\mathcal{P}^{\text{PA}}$ has 1 independent dimension and can be represented on a 1D space. It is represented by the thick blue line in Figure 3.3 on space $p_B = r_B$. As shown on the illustration (dashed grey lines), the feasible set of zonal net positions can be interpreted as the projection of the feasible set of nodal net injections on the space of zonal net positions.

Similarly to the case of nodal pricing, one can define the capacity expansion model from the central planner perspective, using the set $\mathcal{P}^{\text{PA}}$, as follows:

$$\min_{x,y,s,p} \sum_{i \in I, z \in Z} IC_i \cdot x_{iz} + \sum_{i \in I, z \in N, t \in T} MC_i \cdot y_{izt} + \sum_{n \in N, t \in T} VOLL \cdot s_{zt} \quad (3.6a)$$

$$(\mu_{izt}) : y_{izt} \leq x_{iz} + X_{iz}, i \in I, z \in Z, t \in T \quad (3.6b)$$

$$(\rho_{zt}) : p_{zt} = \sum_{i \in I} y_{izt} + s_{zt} - D_{zt}, z \in Z, t \in T \quad (3.6c)$$

$$p_{:t} \in \mathcal{P}^{\text{PA}}, t \in T \quad (3.6d)$$

$$x \geq 0, y \geq 0, s \geq 0 \quad (3.6e)$$

As we discuss in the introduction (section 1.2.3), bidding zone borders in Europe correspond mostly to the borders between Member States, with only a few exceptions. This indicates that the current delimitation of bidding zones is not the outcome of a technical analysis, but is rather the most acceptable solution from an institutional point of view[4]. This institutional decision has a simple consequence: there is a unique price per zone in the electricity market. This constitutes the fundamental property of zonal pricing, and this is the only thing that we impose in order to obtain model (3.6). We note that this does not mean that internal congestion is omitted. As one can observe in equation (3.4), every line in the network can influence the feasible set of net positions and thus the results of the zonal market.

The physical consequences in terms of congestion management of this political decision are, however, not trivial, and the simplicity of the economic interpretation of zonal pricing (i.e. unique price per zone) can quickly become confused with the physics. An evidence of this confusion can be found in the extensive use of the so-called "copper plate" assumption to describe the zonal market. It is often said in the literature that the zonal design relies on the copper plate assumption, but a precise definition of this terminology is rarely offered. One exception is [CRE17], which lists the two properties of a copper plate: unlimited internal transmission capacity and zero internal impedance. Under these conditions, the equivalence between nodal and the zonal model of equation 3.4 indeed holds. But one can also mention other definitions of the copper plate assumption that do not lead to the equivalence: [VD16] defines it as "ignoring transmission constraints within a zone", and [Har18] refers to a copper plate when transmission capacity is assumed to be unlimited within each bidding zone.

One should also note that the concept of copper plate is only an abstraction. In practice, a network can neither have unlimited capacity nor have zero internal impedance. In Proposition 3.1, we clarify the conditions for an equivalence between the two pricing models based on physical quantities.

Proposition 3.1. *Let us define the zonal network as the network obtained by aggregating the nodes of a zone into a single zone and by keeping only the cross-zonal lines. If,*

1. *the transmission capacity constraints of intra-zonal lines are never binding, and*

2. *the zonal network is radial (i.e. the graph associated to the network is a tree),*

then the nodal model (3.2) *and the zonal model* (3.6) *are equivalent.*

Proof. We prove this statement formally in the end of appendix section 3.C.

Note that the fact that unlimited transmission capacity within each zone is not a sufficient condition for equivalence between nodal and zonal pricing when the zonal network is not radial was already recognized at the time of the debates on nodal versus zonal pricing in the US [Hog98].

[4]We elaborate more on the links between the European institutional setting and market design choices in the conclusion of this dissertation, section 5.3.

The reasoning relating to the decentralization of the solution extends easily to the case of zonal pricing using model (3.6), where the dual variables ρ_{zt} are interpreted as zonal prices. Once again, under this design, investment costs are covered by zonal scarcity rents, i.e. the equation

$$0 \leq x_{iz} \perp IC_i - \sum_{t \in T} \mu_{izt} \geq 0 \;\; \forall i \in I, z \in Z \tag{3.7}$$

holds from the KKT conditions. The main difference between model (3.2) and model (3.6) is that decision variables x, y, s and p are now indexed on the zones. This difference raises two major questions concerning the implementation of the zonal market in real operations:

- How does the zonal dispatch y_{izt} translate into an implementable dispatch y_{int} in real time?
- How does the zonal investment x_{iz} translate into an actual nodal investment?

The first question is related to *re-dispatch*, that we have introduced in the previous chapter. Two main approaches regarding the implementation of re-dispatch currently co-exist in Europe. The traditional approach is a cost-based regulatory re-dispatch, whereby the TSO remunerates producers for being re-dispatched up or is remunerated in case of downward re-dispatch, in a pay-as-bid fashion, based on cost estimates derived from the competent regulatory authorities. The second approach that has started to gain importance more recently in Europe is a market-based re-dispatch [HS18]. Under this approach, the spot market is followed by a re-dispatch market where producers are allowed to bid freely and are remunerated based on a uniform price. In Europe, market-based re-dispatch is currently implemented in the UK, Italy, the Netherlands and in the Nordic market [GMS+18, HSMT19]. It should also be noted that the European Commission seems to favor market-based re-dispatch. It has made it the new default rule through Article 13 of the Electricity Regulation [Eur], although the article is subject to a list of strong exceptions. As the present work, in particular the large-scale case study presented in section 3.5, is focused on CWE in which most countries use cost-based re-dispatch, we assume cost-based re-dispatch for the entire CWE region. The case of the long-term equilibrium with zonal pricing followed by market-based re-dispatch is analyzed in the next chapter of this dissertation.

In theory, if generators are completely flexible and there are no unit commitment decisions made based on the zonal dispatch, zonal market-clearing followed by cost-based re-dispatch leads to the same welfare as nodal market clearing and only induces a welfare re-allocation[5]. In practice, however, a loss of welfare is associated to zonal unit commitment, as we discuss in chapter 2.

[5]This statement also assumes a unique TSO that manages the re-dispatch phase, a TSO with the goal of maximizing welfare (as opposed to minimizing deviations from day-ahead market clearing, which may sometimes be the case in practice), that there is no uncertainty

One can only be less affirmative in answering the second question, as it is not related to existing rules or procedures. The origin of the problem is that when one enforces a uniform price over all buses of a given zone, it becomes ambiguous where exactly a specific technology will choose to invest within that zone, and this despite exerting different levels of physical stress on the network of the zone. In this work, we adopt the optimistic assumption that the investment is made in the best possible location for the system. This enables us to compare nodal investment to a best-case version of zonal investment. This assumption is effectively equivalent to granting the TSO the power of deciding where the zonal investment will be located in the grid, with the objective of minimizing total re-dispatch costs.[6]

Putting everything together, we model the re-dispatch phase as a cost-based minimization problem with the full nodal network constraints available to the TSOs, and where the TSOs can choose the nodal disaggregation of zonal investment. We represent the re-dispatch phase as follows:

$$\min_{x,y,s,r,f} \sum_{i\in I, n\in N, t\in T} MC_i \cdot y_{int} + \sum_{n\in N, t\in T} VOLL \cdot s_{nt} \tag{3.8a}$$

$$\sum_{n\in N(z)} x_{in} = \bar{x}_{iz}, \ \ \forall z \in Z \tag{3.8b}$$

$$y_{int} \leq x_{in} + X_{in}, i \in I, n \in N, t \in T \tag{3.8c}$$

$$r_{nt} = \sum_{i\in I} y_{int} + s_{nt} - D_{nt}, n \in N, t \in T \tag{3.8d}$$

$$r_{:t} \in \mathcal{R}, t \in T \tag{3.8e}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{3.8f}$$

where $\bar{x}_{iz}$ is the solution from the zonal investment problem.

Let us now discuss the results of this zonal pricing model on our illustrative example. The optimal solution is to invest in 1918 MW of coal, in 7086 MW of nuclear and in 1615 MW of gas capacity. No investment is made in zone A, all in node B. One observes that the zonal solution leads to an under-investment of 400 MW of gas capacity in total (100 MW less than optimal in zone B and 300 MW less than optimal in zone A), compared to the nodal solution. This implies

---

in the system, no strategic behavior and that there are no irrevocable decisions taking place in the day ahead. Said differently, the conditions in the day ahead and in real time are identical. In this case, the dispatch solution found by re-dispatch and by the nodal pricing market are the same. Producers keep their infra-marginal rent, which induces a welfare re-allocation. This point of view is obviously not satisfied in practice, and it overlooks a number of important negative side-effects of a zonal price signal, such as inducing non-truthful bidding (inc-dec gaming) and a failure to provide an appropriate locational investment signal (which is the focus of the present chapter).

[6] Past experience suggests that real outcomes can violate our optimistic assumption. A case in point is the extensive development of wind capacity in the McCamey region in West Texas, despite the fact that the transmission export capabilities of the area were insufficient. The investments were based on the price of the entire Western Texas zone, which had insufficient granularity in order to guide optimal siting decisions [AZ06].

that producers cannot cover the full demand in the peak hour and there is a curtailment of 200 MW in node A_n and 100 MW in node B. In terms of cost, the zonal solution is significantly more expensive, with a total cost of 530,917€. This decomposes into an investment cost of 265,515€ and an operating cost (which includes re-dispatch costs) of 265,403€. One can observe that the zonal solution achieves minor savings in terms of investment cost, but faces a severe increase in operating cost, in part due to the demand curtailment that takes place in the peak hour.

Although interesting from a theoretical point of view, zonal pricing markets based on $\mathcal{P}^{\text{PA}}$ have not been implemented in practice. Instead, other methods have been proposed and used over the years to define the set of acceptable zonal net positions $\mathcal{P}$, as described in section 1.2.2 of the introduction to the dissertation. Currently, FBMC is the default approach for market coupling. From a long-term perspective, however, FBMC raises certain concerns that we shall discuss in detail in the next section.

3.4 Flow-based market coupling in the context of capacity expansion

FBMC deviates from the Price Aggregation (PA) model by introducing a set of rules in order to approximate the expected flows on inter-connectors. In this section, we will integrate Aravena's model (i.e. the set $\mathcal{P}^{\text{FBMC-EP}}$ defined in equation (1.4)) into a capacity expansion model in order to focus on the efficiency of FBMC from a long-term perspective.

We start by recalling Aravena's model of the network constraints in FBMC in section 3.4.1. We then present respectively the capacity expansion model of a central planner under FBMC and its decentralized version in sections 3.4.2 and 3.4.3, as well as their respective results on the small illustrative example introduced in the previous section.

3.4.1 Network constraints in FBMC

Using Aravena's model [ALPS21], the network constraints in FBMC can be written as follows:

$$\begin{aligned}\mathcal{P}^{\text{FBMC-EP}} = \Bigg\{ p \in \mathbb{R}^{|Z|} \,\Bigg|\, & \exists (f, r, \tilde{y}) \in \mathbb{R}^{|K|} \times \mathbb{R}^{|N|} \times \mathbb{R}^{|I||N|} : \\ & p_z = \sum_{n \in N(z)} r_n \;\; \forall z \in Z, r_n = \tilde{y}_{int} - D_{nt} \;\; \forall n \in N, \\ & 0 \le \tilde{y}_{int} \le X_{in} \;\; \forall i \in I, n \in N, \sum_n r_n = 0 \\ & f_k = \sum_n PTDF_{kn} \cdot r_n, -TC_k \le f_k \le TC_k, \;\; \forall k \in k \Bigg\}\end{aligned} \tag{3.9}$$

In this set of equations, variables $\tilde{y}_{int}$ can be understood as an auxiliary nodal dispatch. By introducing $\mathcal{P}^{\text{FBMC-EP}}$, TSOs ensure the existence of an auxiliary dispatch that respects the cleared zonal net positions and that can serve demand without curtailment.

The important thing to note here is that, in this setting, the TSOs do not only use grid quantities to provide network constraints to the market, but they also use quantities related to demand (D_{nt}) and installed capacity (X_{in}). The efficiency and practicability of this approach can be questioned from a short-term perspective. Indeed, it can be hard to forecast correctly D_{nt} and know exactly X_{in} for the system operator, and one can expect that this will be increasingly the case in the future as demand response and renewable integration will increase the uncertainty and variability in the grid. These difficulties, however, are not the subject of this chapter, where our focus is on the long-term efficiency of this design. In the long-term problem, the installed capacity is not known by the system operator but is rather a decision variable of the system. Therefore, one needs to include the capacity expansion variables x_{iz} into the set $\mathcal{P}^{\text{FBMC-EP}}$. We denote this extended set by $\mathcal{PX}^{\text{FBMC-EP}}$ that is now defined on the space of zonal net positions p *and* zonal investment x:

$$
\begin{aligned}
\mathcal{PX}^{\text{FBMC-EP}} = \Bigg\{ p \in \mathbb{R}^{|Z|}, x \in \mathbb{R}^{|I||Z|} \,\Bigg|\, & \exists (f, r, \tilde{y}, \tilde{x}) \in \mathbb{R}^{|L|} \times \mathbb{R}^{|N|} \times \mathbb{R}^{|I||N|} \times \mathbb{R}_+^{|I||N|} : \\
& p_z = \sum_{n \in N(z)} r_n \ \forall z \in Z, x_{iz} = \sum_{n \in N(z)} \tilde{x}_{in} \ \forall i \in I, z \in Z, \\
& r_n = \tilde{y}_{in} - D_n \ \forall n \in N, \ 0 \leq \tilde{y}_{in} \leq X_{in} + \tilde{x}_{in} \ \forall i \in I, n \in N, \\
& f_k = \sum_n PTDF_{kn} \cdot r_n \ \forall k \in K, \\
& \sum_n r_n = 0, -TC_k \leq f_k \leq TC_k, \ \ \forall k \in K \Bigg\}
\end{aligned}
\tag{3.10}
$$

Note that the set $\mathcal{PX}^{\text{FBMC-EP}}$ creates a dependency between the decisions of different agents in the model, hence the generalized Nash equilibirum (GNE) structure that is developed in further detail in section 3.4.3 and in appendix section 3.B. Finally, let us note that $\mathcal{PX}$ defines a polytope on the set of zonal net positions and zonal investment. That is, it can be expressed as a set of M linear inequalities on p and x, and there exists $V \in \mathbb{R}^{M \times |Z|}$, $U \in \mathbb{R}^{M \times |I||Z|}$ and $W \in \mathbb{R}^M$ such that

$$
(p, x) \in \mathcal{PX}^{\text{FBMC-EP}} \Leftrightarrow \sum_{z \in Z} V_{mz} p_z + \sum_{i \in I, z \in Z} U_{miz} x_{iz} + W_m \geq 0 \ \forall m \in \{1, ..., M\}
\tag{3.11}
$$

3.4.2 Centralized capacity expansion under FBMC

Using set $\mathcal{PX}^{\text{FBMC-EP}}$ and its expression as linear constraints defined in (3.11), one can easily define the capacity expansion problem from the central planner's perspective:

$$\min_{x,y,s,p} \sum_{i\in I,z\in Z} IC_i \cdot x_{iz} + \sum_{i\in I,z\in N,t\in T} MC_i \cdot y_{izt} + \sum_{n\in N,t\in T} VOLL \cdot s_{zt} \tag{3.12a}$$

$$(\mu_{izt}) : y_{izt} \leq x_{iz} + X_{iz}, i \in I, z \in Z, t \in T \tag{3.12b}$$

$$(\rho_{zt}) : p_{zt} = \sum_{i\in I} y_{izt} + s_{zt} - D_{zt}, z \in Z, t \in T \tag{3.12c}$$

$$(\gamma_m) : \sum_{z\in Z} V_{mz} p_z + \sum_{i\in I,z\in Z} U_{miz} x_{iz} + W_m \geq 0, \;\; \forall m \in \{1, ..., M\}, t \in T \tag{3.12d}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{3.12e}$$

Problem (3.12) is an optimization model of investment in the zonal system. It is similar to the nodal investment problem except for the zonal representation of the grid, represented by constraints (3.12d). Very much like the nodal model and the zonal model with price aggregation, its KKT conditions define prices. Because the optimization problem contains the zonal network constraints, that depend on both the zonal net position and the investment, these constraints are priced in the KKT conditions. This takes place through dual variables γ_m that modify the investment criterion of the generators by imposing revenue that induces them to modify their investment so that the FBMC network constraints are respected. This variable results in truly internalizing the dependence of investment on the network constraints. It can be interpreted as a zonal subsidy that internalizes this dependence. It has the same role as in environmental policy: when imposed at the right value (like the right value of a CO_2 tax) it guarantees that the externality caused by the investment in capacity in a particular location is internalized.

The KKT condition associated to the investment variables can now be written as follows:

$$0 \leq x_{iz} \perp IC_i - \sum_{t\in T} \mu_{izt} - \sum_{m\in\{1,...,M\}} U_{miz}\gamma_m \geq 0 \;\; \forall i \in I, z \in Z \tag{3.13}$$

One observes an important difference between condition (3.13) and the condition under the PA zonal pricing model (3.7): the investment cost in the centralized FBMC model is not covered solely by the scarcity rents obtained from selling electricity. Revenues associated to the network constraints must be added to cover it. This implies that model (3.12) cannot be readily decentralized using zonal prices ρ_{zt} associated to constraint (3.12c). Energy-only markets under FBMC are thus imperfect and network constraints must be

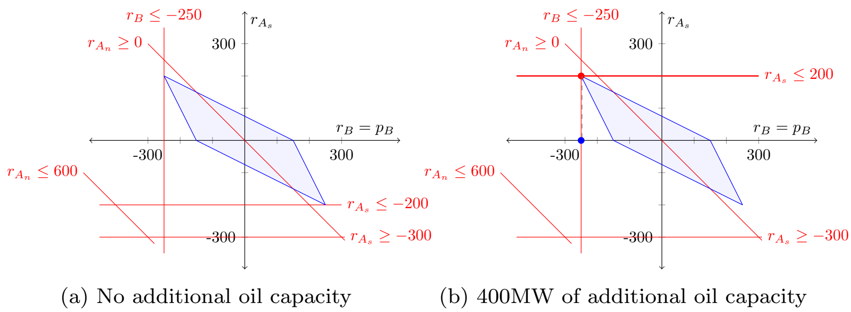


Figure 3.4: Representation of the flow-based constraints on the space of r_{A_s} and r_B. The PA polytope is shown in blue. The additional constraints imposed in FBMC are shown in red with no oil capacity invested (left) and 400MW of oil capacity invested in node A_s (right). One observes that the problem is infeasible in the first case. The investment of 400MW of capacity makes the problem feasible in the second case, as shown with the red dot and its projection on the space of net positions (blue dot).

priced if we want to restore the link between the problem of the central planner and the decentralized problem. The decentralization would not only be based on energy prices ρ_{zt} but also on these network prices γ_m.

Returning to our illustrative example, the results of the FBMC model of the central planner are the same as the PA model in node B, i.e. 1918 MW of coal, 7086 MW of nuclear and 1615 MW of gas capacity. However, the solution differs in zone A, with the investment of 400 MW of additional Oil capacity in node A_s. In terms of cost, this yields 120,882€ for operating cost, 266,315€ for investment cost, and a total cost of 387,197€. The total cost is higher than in nodal pricing, but is considerably reduced compared to the PA model. The reason is that, in the FBMC model, re-dispatch is ensured to be feasible without load shedding. This results in significant operating cost savings. One should note that, as we mention earlier, investment costs are not covered by the sole sale of electricity in this case. Indeed, the price in zone B amounts to 12.56€/MWh, 27.2€/MWh and 97.52€/MWh in the first, second and third period respectively. If we focus on the specific case of gas capacity in node B, we observe that it will only produce in the peak period. It will thus achieve a net profit of $97.52 - 80 = 17.52$€/MWh in the 1500 hours of the peak period, which gives $\frac{17.52 \cdot 1500}{8760} = 3$€/MWh of net profit and is below the investment cost of 5€/MWh.

In order to further illustrate the FBMC model and highlight its differences with the PA model, we represent the constraints of set $\mathcal{PX}^{\text{FBMC-EP}}$ on the space (r_{A_s}, r_B) in Figure 3.4. Unlike in the case of zonal PA, the flow-based polytope depends on the capacity invested in every node. On the left panel, we present the flow-based constraints with the generation capacity corresponding to the results of the PA model. On the right panel, the capacity corresponds to the results of the centralized FBMC model. The FBMC model imposes additional constraints on the nodal net injection variables, compared to the PA model. These additional constraints are presented in red. The important thing to observe is that with the capacity of the PA model (Figure 3.4a), the flow-based polytope is empty and the dispatch problem is thus infeasible. The centralized FBMC model will invest in capacity in node A_s until the polytope becomes non-empty, which is represented by the red dot in the nodal space and the blue dot, its projection, in the zonal space (Figure 3.4b). The result is that there is an additional 400MW of oil capacity that is invested in node A_s.

Regarding the precise value of the term γ_m, one should note that the market clearing price in zone A in the peak period does not change compared to the PA model. It remains at 80€/MWh, the marginal cost of the gas capacity that is in excess in zone A. This implies that the oil capacity built in the south is not cleared and its scarcity rent $\sum_{t \in T} \mu_{izt}$ is zero. By equation (3.13), $\sum_{m \in \{1,\ldots,M\}} U_{miz}\gamma_m = IC_i = 2$€/MWh. The same value of 2€/MWh also holds for all technologies in zone B, which confirms what we have just discussed in the case of gas above: the investment cost of 5€/MWh from which we subtract the net profit of 3€/MWh equals the term in γ_m of 2€/MWh.

Finally, we note that the set $\mathcal{PX}^{\text{FBMC-EP}}$ of feasible net positions defined by the TSOs depends on the decision variables of the producers only through the investment x_{iz}. Therefore, the term in $\sum_{m \in \{1,\ldots,M\}} U_{miz}\gamma_m$ is a capacity-based term that does not depend on the time period. Moreover, additional capacity can only expand the set $\mathcal{PX}^{\text{FBMC-EP}}$, not restrict it. For this reason, the term in γ_m is always positive and can thus be interpreted as a subsidy.

3.4.3 Decentralized capacity expansion under FBMC

Let us now consider the case where one does not complete the market by the incentive represented by γ_m. This is the situation corresponding to the current FBMC market design, in which the network constraints (3.12d) are not priced and there is no revenue associated with γ_m, while the market is decentralized by definition in a liberalized electricity market. One then drops this variable from the KKT conditions of problem (3.12). This corresponds to replacing equation (3.13) by (3.7) in its KKT conditions. One obtains a new complementarity problem and the question is to understand what it represents. Given that we no longer have an optimization problem, one may wonder whether one ends up with a Nash equilibrium problem. This would be quite compatible with an unpriced externality: a set of agents, each maximizing its profit in a world with a non-internalized externality, is typically a Nash equilibrium. But the generation/transmission problem raises an issue that is rooted in the separation of these functions. While the separation was motivated by the competitive nature of generation and the monopoly of the grid, it raises a difficulty in the zonal system that is absent from the nodal system. All lines are priced in the nodal system and the separation between the two functions can then be decentralized by prices. This is not as straightforward in a zonal system, as information about the generation and load are usually used in the determination of the network

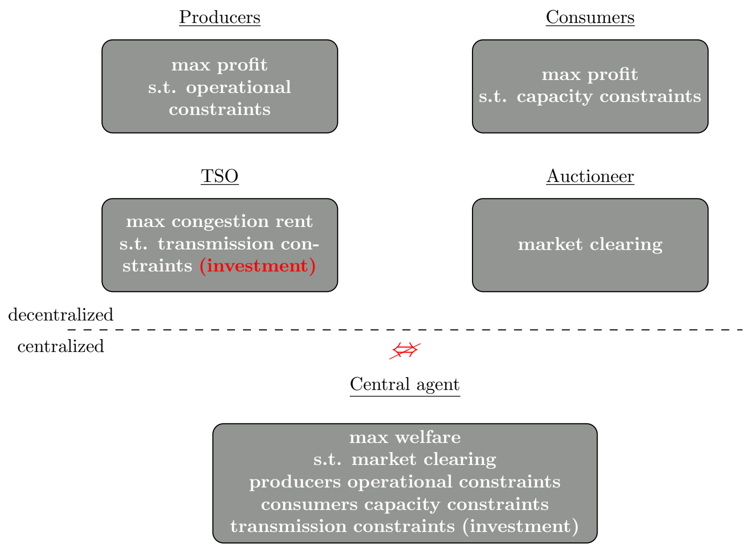


Figure 3.5: Schematic representation of the decentralized game and the problem of the central planner. In the case of FBMC, the feasible set of the TSO depends on generation investment, which corresponds to decision variables of the producers. The appropriate solution concept of the game is thus the GNE and the equivalence between the centralized and decentralized formulations is broken.

constraints, as it is the case in flow-based market coupling, for example. One could also decentralize by prices in the zonal system if one would modify the design so that it does not use information about generation and load in the network constraints (as in the PA model) or if one prices the network constraints through the γ_m variable, as discussed above. A difficulty with the nature of the equilibrium arises when one does not resort to this and one requires the decentralization of the imperfect market. The coordination between generation and transmission achieved by pricing the network constraints (3.12d) must now be achieved by quantities when this price is absent. The technical consequence is that the expected Nash equilibrium (NE) among generators becomes a GNE between generators and the TSO because of the integration of investment variables in the TSO network constraints. Figure 3.5 illustrates the difference in the structure of the game in the case of FBMC compared to that of nodal pricing and zonal with PA that was shown in Figure 3.1. The overall problem can be described as a linear complementarity problem which characterizes the KKT conditions of the profit maximization problems of the four market agents of Figure 3.5. We insist on the fact that the difficulty discussed here regarding the transition from a NE to a GNE is not a fundamental property of the zonal system itself, but a consequence of how transmission capacity is currently computed in the flow-based market coupling methodology implemented in Europe. The reader is referred to appendix section 3.B for the full developments of this decentralization, including the justification of the GNE nature of the problem and the complete set of equations that defines it.

The GNE structure also raises questions regarding existence and uniqueness. As it turns out, the problem might be infeasible in general because of the condition that each accepted zonal net position should disaggregate into at least one nodal dispatch that meets the forecast demand. One can already understand this from a very simple example of a one-zone two-node network. Assume that there is an existing capacity of 2 MW in one node, a demand of 1 MW in the other node, and that the two nodes are connected by a fictitious line of 0 MW of capacity. Clearly, because of congestion on the line, capacity must be built on the second node to cover the entire demand without curtailment. However, the zonal market does not see this internal congestion and will understand the demand and capacity as zonal quantities. The zonal price will thus be bounded by the marginal cost of the existing unit and this revenue will in general be insufficient for investing in new capacity.

In our model, we therefore assume that TSOs can also invest in capacity in order to ensure the feasibility of re-dispatch at any time. This capacity can be assimilated to a network reserve, such as the one that has been implemented in Germany in order to support generators in Southern Germany. These generators were necessary for the security of supply, but were not financially viable without this aid [BC19]. Existence of solutions on this modified model is guaranteed and, in general, there will be disjoint sets of solutions, as we prove formally in appendix section 3.D.

Let us now examine the results of this new model on the illustrative example. In the case without network reserve investment from the TSOs, the model is infeasible. This can be understood from the results of the problem of the central planner: if one wishes to impose feasibility of re-dispatch, some investment will be made that leads to revenues from the short-term market that do not cover the investment costs. For the case with network reserve, let us assume that this network reserve has an investment cost of 200€/MWh, i.e. higher than all other market-based technologies, but that its marginal cost is 0€/MWh. The following solution is found: 1918 MW of coal, 7086 MW of nuclear and 2015 MW of gas capacity in node B and 50 MW of network reserve in the lower node of zone A. The operating costs amount to 113,868€, the investment costs to 277,515€, thus resulting in a total cost of 391,383€. In terms of efficiency, this model is more costly than the one of the central planner, due to the need of investment in network reserve, but is less expensive than the PA model.

Finally, one should note that the results of the PA model are obtained under the assumption of no network reserve. However, it could also be profitable for the TSOs in that model to invest in network reserve in order to mitigate the costs of demand curtailment at the re-dispatch phase. Therefore, in order to provide a fair comparison between zonal with PA and zonal with FBMC, we propose another model which is the same as the zonal PA, but where the TSO is allowed to invest in network reserve, in order to improve the efficiency of the re-dispatch, even if this is not strictly necessary to make the problem feasible.

Policy	Inv. in node B	Inv. in node A_s	Op. costs	Inv. costs	Total costs
Nodal	Coal: 1918MW Nuclear: 7086MW Gas: 1715MW	Gas: 300MW	267,515€	114,033€	381,548€
PA	Coal: 1918MW Nuclear: 7086MW Gas: 1615MW		265,403€	265,515€	530,917€
PA-NR	Coal: 1918MW Nuclear: 7086MW Gas: 1615MW NR: 100MW	NR: 200MW	107,042€	325,515€	432,557€
FBMC-C	Coal: 1918MW Nuclear: 7086MW Gas: 1615MW	Oil: 400MW	120,882€	266,315€	387,197€
FBMC-D	Coal: 1918MW Nuclear: 7086MW Gas: 2015MW	NR: 50MW	113,868€	277,515€	391,383€

Table 3.1: Summary of the results of the different policies on the illustrative example. PA-NR is the PA model presented in section 3.3.2, with network reserve. FBMC-C is the centralized FBMC model of section 3.4.2 and FBMC-D is its decentralized version (section 3.4.3).

We refer to this model as PA-NR. Taking this possibility into account, we do observe an investment of 100 MW of network reserve in node B and 200 MW in node A_s, which reduces the total cost of the PA solution to 432,557€. Even in this case, the PA-NR solution is still significantly more costly than the decentralized FBMC solution. This shows that the equivalence between centralized and decentralized solutions is not sufficient to arrive to an efficient zonal model.

A summary of the results of all the different policies on the illustrative example is provided in Table 3.1.

3.5 Results on the CWE case study

The goal of this section is to present the results of the different policies on a realistic instance of the CWE area. The dataset, models and algorithms used for this case study are provided in an online Git repository: `https://github.com/qlete/ZonalLongterm`.

The capacity expansion problems and re-dispatch problems for the Nodal, centralized FBMC and Price Aggregation policies correspond to single linear optimization problems that can be readily solved. The models are implemented in Julia [BEKS17] using JuMP v0.21.5 [DHL17] and solved with Gurobi 9.1. The decentralized capacity expansion FBMC model is solved using a linear splitting-based method that we regularize in order to improve convergence. The method essentially corresponds to iteratively solving modified versions of the centralized FBMC model until a fixed point is reached. We refer the reader to Appendix section B for the details of our solution methodology.

Type	Number of units	Total installed capacity [GW]
Nuclear	73	77.67
Natural gas	403	56.38
Coal	93	30.7
Lignite	59	20.82
Oil	75	6.37
Other	189	6.08

Table 3.2: Total installed capacity of conventional units in the database per type of fuel.

3.5.1 Dataset

Our starting point is the dataset used in previous work by the authors [ALPS21]. The network data is an updated version of the European grid model of [HB13]. The generation data is obtained from [Ope20]. Table 3.2 presents the total installed capacity per generator type for the entire CWE region.

The network model of [HB13] does not contain the latitude and longitude of buses, but is accompanied by coordinates on an internal coordinate system that is employed in the PowerWorld software, which we assume to correspond to a linear transformation of true geographical coordinates. In order to assign generators to network buses, we first perform a geo-referencing of the network data by obtaining the locations of known substations and use a linear regression in order to extrapolate the remaining locations. We then collect approximate locations of the generators and assign them to the closest network bus[7]. Our time series data (hourly demand, solar and wind production in each country) are obtained from the ENSTO-E Transparency Platform for the year 2018. Because of the complexity of solving the GNE corresponding to the decentralized FBMC capacity expansion problem, we perform a dimensionality reduction on the dataset. Using clustering techniques, the 8760 hours of the year are reduced to 20 representative time periods and the network is reduced from 632 buses to 100 buses. More details on this dimensionality reduction method are provided in Appendix A.

The models that we use for the case study are generalized versions of the models presented in sections 3.3 and 3.4. In the models of the CWE case study, we also consider revenues from reserve provision, where reserve is assumed to be cleared simultaneously with energy. We also consider fixed operating and maintenance costs. Existing units that cannot cover their fixed costs are decommissioned. We assume that investment is possible in 3 different technologies, similarly to [AGK$^+$20]: CCGT units, OCGT units and Combined Heat and Power CCGT units. We use the same cost data as [AGK$^+$20], which are

[7] In most cases, the locations of generators are given in our dataset. When this is not the case, the precise location or approximate location of the municipality is collected individually for each unit.

Type	IC [k€/MW yr]	FC [k€/MW yr]	MC [€/MWh]
CCGT	80.1	16.5	61.29
OCGT	56.33	9.33	100.4
CCGT&CHP	94.39	16.5	41.37

Table 3.3: Annualized investment cost, annualized fixed operating and maintenance cost and marginal cost for the three investment technologies considered for investment in the CWE case study.

Type	FC [k€/MW yr]	MC [€/MWh]
Nuclear	92	9.1
Natural gas	9.33	93.42-121.37
Coal	46.29	44.5-58
Lignite	101.5	36.7-42.12
Oil	9.33	116-210
Other	113.16	38.64

Table 3.4: Annualized fixed cost and marginal cost range of existing open-cycle generators per type of fuel.

presented in Table 3.3.

Wind and solar expansion are accounted for in an exogenous way. The fixed and marginal costs of existing capacity are also sourced from [AGK+20] and completed from [Ope21] when missing. We distinguish between CHP and non-CHP generators. The marginal cost of CHP generators is reduced by 20€/MWh in order to represent the additional revenues from the sale of heat. Finally, we assume a capacity expansion horizon of 2035. Consequently, we remove from the dataset the generators that will be shut down by then, based on the information available in the OPSD dataset [Ope20]. We also remove all nuclear units from Belgium and Germany and integrate the planned closure of 14 nuclear reactors by EDF by 2035 for France [Int20]. A common concern with nodal pricing that is sometimes raised in the literature through non-quantitative arguments is that it is expected to increase price volatility and decrease liquidity in local trading hubs [AHT19, AVFM20]. This can be related to the proposal of Hogan for contract networks [Hog92, Hog99], which are designed to address the problem in a hierarchical fashion. Hogan anticipates that, in markets with nodal pricing, zonal hubs with high liquidity can be identified [Hog99]. These zonal hubs form a contract network, different from real network, on which transmission congestion contracts can be traded. The contracts obtained at these zonal hubs are indeed imperfect hedges for market participants that are located at the local buses. These imperfect hedges could in turn have a negative effect on the risk of investment and its cost. In order to test the robustness of our findings and understand the impact of such

Policy	Op. costs [M€/yr]	Inv. costs [M€/yr]	Total costs [M€/yr]	Efficiency losses [%]
Nodal	15,855	10,432	26,287	-
Nodal risky	15,858	10,529	26,387	0.38
FBMC-C	16,314	10,221	26,535	0.94
FBMC-D	16,368	10,700	27,068	3.0
PA-NR	16,835	10,909	27,744	5.5

Table 3.5: Performance comparison of the different policies.

increase in investment costs under nodal pricing, we consider an additional simulation where the investment costs in the nodal design are increased by 5%. One should note, however, that to the best of our knowledge, an increase of investment costs due to limited liquidity in nodal pricing compared to zonal pricing has not been formally proven in the literature and is, at this stage, speculative. In particular, a decrease in liquidity has not been observed in US markets when they transitioned to a nodal design, and US nodal markets are considered to be sufficiently liquid nowadays [NB11, Dua19].

3.5.2 Efficiency comparison of the different policies

We start by presenting a comparison between Nodal, Nodal risky (5% increase in investment costs), centralized FBMC, decentralized FBMC and zonal with price aggregation in terms of their investment and operating cost performance. FBMC-D, as a design that is both decentralized and based on FBMC, is our closest proxy to the design of the current market. The other policies that we model are benchmarks against which we evaluate the efficiency of the existing design. In particular, FBMC-C enables us to quantify the inefficiencies that are due to the break of equivalence between the centralized and decentralized versions. Table 3.5 presents the investment cost, operating cost and total cost of each of the 5 policies. The efficiency ranking that we observe in the illustrative example also holds for the CWE case study: the nodal policy is the one that achieves the lowest total cost. Notably, this result still holds when the investment costs in nodal are increased by 5%, although the difference with the centralized FBMC policy is reduced. The centralized FBMC policy outperforms significantly its decentralized version, which leads to two notable conclusions: (i) the inefficiencies introduced by the interaction between zonal transmission constraints and investment in the long run are important and (ii) completing the market with network subsidies associated to the dual variables γ_m discussed in section 3.4.2 could, in theory, result in significant benefits. Regarding the zonal price aggregation policy, although its centralized and decentralized formulations are equivalent, one observes that it is the most expensive. This is important: it demonstrates that the PA zonal design is not a simple remedy to the inefficiencies that we have described in this chapter. Comparing the nodal policy with FBMC-C, one observes that nodal pricing

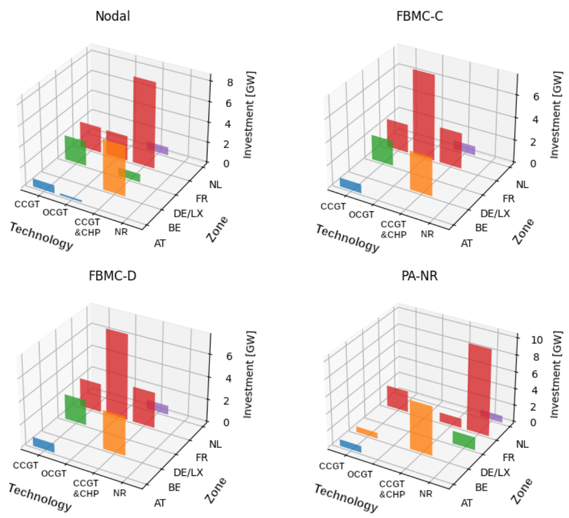


Figure 3.6: Total investment per zone and per technology for each policy.

exhibits higher investment costs, but these are more than compensated by an improvement in operating efficiency. This difference stems mostly from a better locational allocation of revenues within each zone, which allows the nodal policy to better identify profitable decommissioning than the zonal policies (24.7 GW for nodal compared to 21.9 GW for FBMC-C, dropping to 12.2 GW for zonal pricing with price aggregation). The efficiency gap between FBMC-C and FBMC-D is mostly due to the cost of network reserve investment. While there is no investment in network reserve in nodal and the centralized FBMC, both FBMC-D and PA exhibit significant out-of-market network reserve investment (7 GW and 12.6 GW respectively).

3.5.3 Qualitative difference between policies

In this section, we analyse in additional detail the qualitative differences in the solutions of the different policies. We focus on three specific aspects: i) the difference in the type of technologies of the final capacity mix, ii) the decommissioning behavior of the different policies, and iii) welfare re-allocation.

Capacity mix

In Figure 3.6 and Figure 3.7, we present respectively the total investment and decommissioning per zone and per technology for each policy. We have already

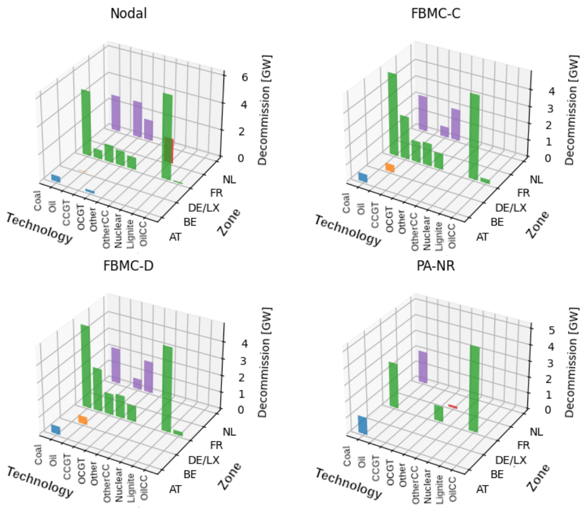


Figure 3.7: Total decommissioning per zone and per technology for each policy.

observed that the nodal policy leads to more decommissioning. In Figure 3.7, we can observe that this is particularly marked for coal and lignite plants in Germany. This more important coal and lignite decommissioning under the nodal policy can be explained by the lower nodal price that some of these units face. Figure 3.8 presents the locational distribution of the average price under each policy as well as the amount of existing coal and lignite capacity in each bus, represented by the relative size of the node. One observes that the second largest nodal coal and lignite capacity is located in a bus in eastern Germany that faces a large decrease in price under the nodal policy, which implies that these plants do not achieve the profits that are required for covering their fixed operating and maintenance costs. This result can also be related to the observation made in [CRE19] that lignite and hard coal take advantage of structural downward re-dispatch in Germany. The fact that these production units are often re-dispatched down indicates that they benefit from an infra-marginal rent from the day-ahead market that would not exist with nodal pricing. Our results suggest that this also has an impact in the long term, with some of these coal and lignite units being kept profitable only with these rents.

Decommissioning behavior

Another interesting observation that can be based on Figures 3.6 and 3.7 is that both the nodal and the centralized FBMC policies sometimes invest in and decommission the same technology in the same zone. This is clearly observed,

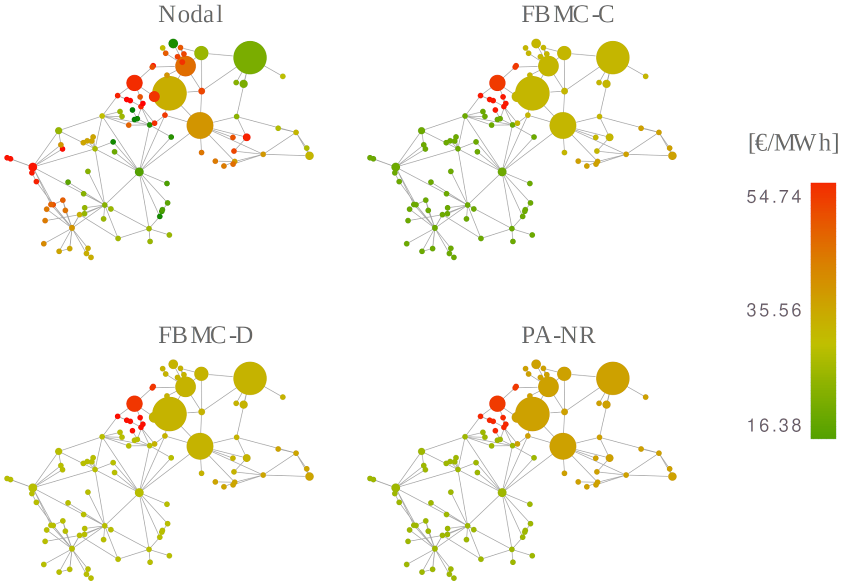


Figure 3.8: Locational distribution of the average price under the four different policies. The size of the nodes is proportional to the amount of existing coal and lignite at these locations.

Figure 3.9: Total surplus difference with the nodal solution for the different types of agents. Negative surplus means that the agent earns more surplus in the nodal solution.

for instance, for CCGT in Germany. The reason why this can occur under the nodal policy is straightforward: because of the high locational variability in the nodal prices, one CCGT unit could be profitable in a node with a high price while this same unit would be incapable of recovering its fixed operating and maintenance cost in a low-price node. The same phenomenon for the centralized FBMC policy could seem counter-intuitive at first glance because of the zonal nature of the price. One should, however, recall that the nodal location of capacity does influence the price under centralized FBMC, as it is directly used when defining the transmission constraints, as one can observe in equation (3.10). The zonal pricing policy with price aggregation, in contrast, would never lead to such a situation as it is incapable of differentiating capacities of the same technology within the same zone.

Welfare re-allocation

Finally, we wish to stress that, while each policy leads to a different total welfare, as discussed in section 3.5.2, the allocation of welfare is also not homogeneous between the different agents. Figure 3.8 already highlights the important welfare re-allocation between the different locations. In Figure 3.9, we display the difference between the total economic surplus of the different agents in each zonal policy and in the nodal policy. We split the welfare between producer surplus, consumer surplus, TSO congestion rents, and TSO surplus due to re-dispatch. A negative value corresponds to a higher surplus in the nodal policy, whereas a positive value corresponds to a higher surplus in the corresponding zonal policy. Clearly, ignoring internal congestion improves price convergence, which benefits mostly consumers (in terms of surplus, but not necessarily in terms of final electricity price paid, as we explain below). The revenues of the TSO under the nodal policy result from the auctioning of valuable transmission capacity. If we limit these revenues solely to cross-zonal congestion, they decrease drastically. Moreover, the congestion will have to be handled out of the market, which results in a cost to the TSO.

Regarding the difference between the zonal policies themselves, we observe that FBMC-C is the zonal policy that imposes the largest price differences among zones. When cross-zonal transmission constraints are further simplified, the price in France increases, which leads to a transfer of surplus from consumers to producers.

One should finally note that the increased economic surplus for consumers is counter-balanced by the fact that the loss of revenues and increased costs for the TSO will be compensated by increased network tariffs. Different rules coexist in Europe for designing transmission tariffs. One thing that is safe to say is that the costs are not allocated homogeneously among market participants. Some important differences in tariffs exist, for instance, between producers and consumers [ENT18] and among industrial and local consumers [Män15]. Consumers, and especially local consumers, are thus expected to bear the greater part of this loss of TSO revenues.

Finally, one may wonder whether network tariffs could play a role in mitigating the inefficiencies in investment caused by zonal pricing by introducing a locational component to these tariffs. It is important to stress that tariffs alone cannot solve the issues that we identified in this chapter, as these issues are related to missing money problems associated to zonal pricing. Therefore, additional revenues to producers are required in order to obtain a closer to optimal investment plan, not additional charges. Locational network tariffs could however play a role in combination with other zonal missing money remedies in order to steer investment to the right location. This is the subject of the next chapter of this dissertation.

3.6 Conclusion

The capacity expansion problem is a key analytical tool in an era of energy transition. In this chapter, we have revisited this problem in light of the ongoing discussion regarding capacity allocation in European zonal markets. We propose a model of capacity expansion in zonal pricing markets based on FBMC and we show that the equivalence between the formulation of the central planner and the decentralized formulation ceases to hold. The decentralized problem is thus formulated as a GNE. We then perform a case study on a realistic instance of CWE and we provide a comparison of the four designs that we discussed: nodal pricing, centralized FBMC, decentralized FBMC and zonal pricing with price aggregation.

We find large efficiency gaps between the four designs, with nodal pricing significantly outperforming the different zonal variations. In particular, we evaluate the efficiency losses of the current decentralized FBMC design at around 3%. According to our simulations, about two thirds of these losses are due to the break of equivalence between the centralized and the decentralized versions of capacity expansion. The efficiency losses are even greater for the zonal PA policy, which shows that equivalence between centralized and decentralized formulations is not a sufficient condition for a zonal design to be efficient.

From a qualitative point of view, we discuss some specific differences between the solutions of the considered designs. One first important difference relates to the final capacity mix of the solutions. We observe that the higher granularity in nodal prices leads to more decommissioning of coal and lignite power plants in Germany and their replacement by gas-fired units. We also provide observations on the welfare re-allocation. We observe that the zonal policies lead to a significant increase in consumer surplus at the expense of decreased TSO revenues and increased TSO costs. We remark, however, that the net effect of this phenomenon will be decreased consumer surplus, the amount of which depends on the tariff design policies of individual Member States.

3.A Decentralized capacity expansion in nodal pricing

In this section, we formally describe the decentralized capacity expansion problem and show its equivalence with the central planning formulation. The decentralization of electricity markets with transmission constraints was first suggested in [Hog92] and formally described in [BS01]. We repeat these results here using the same notation as in the present chapter.

We distinguish four types of agents: producers, consumers, the TSO and a Walrasian auctioneer that clears the market at each location and determines the market clearing price. We describe sequentially the profit maximizing problems of each type of agent as well as the corresponding necessary and sufficient KKT conditions.

Producers. For each $i \in I, n \in N$

$$\begin{aligned} \max_{x_{in}} & \sum_{t \in T} \Big((\rho_{nt} - MC_i) y_{int} \Big) - IC_i x_{in} \\ (\mu_{int}) : & X_{in} + x_{in} - y_{int} \geq 0 \\ & x_{in} \geq 0, y_{int} \geq 0 \end{aligned} \tag{3.14}$$

KKT system:

$$\begin{aligned} & 0 \leq x_{in} \perp IC_i - \sum_{t \in T} \mu_{int} \geq 0, i \in I, n \in N \\ & 0 \leq y_{int} \perp MC_i + \mu_{int} - \rho_{nt} \geq 0, i \in I, n \in N, t \in T \\ & 0 \leq \mu_{int} \perp X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \end{aligned} \tag{3.15}$$

Consumers. For each $n \in N$

$$\begin{aligned} \max_{s_{nt}} & \sum_{t \in T} (VOLL - \rho_{nt})(D_{nt} - s_{nt}) \\ \text{s.t. } (\delta_{nt}) : & D_{nt} - s_{nt} \geq 0, t \in T \\ & s_{nt} \geq 0 \end{aligned} \tag{3.16}$$

KKT system:

$$\begin{aligned} & 0 \leq s_{nt} \perp VOLL - \rho_{nt} + \delta_{nt} \geq 0, n \in N, t \in T \\ & 0 \leq \delta_{nt} \perp D_{nt} - s_{nt} \geq 0, n \in N, t \in T \end{aligned} \tag{3.17}$$

TSO. The TSO maximizes its congestion rent while respecting network constraints, i.e.

$$
\begin{aligned}
&\max_{r_{nt}} - \sum_{n \in N, t \in T} r_{nt}\rho_{nt} \\
(\psi_{kt}) &: f_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
(\phi_t) &: \sum_{n \in N} r_{nt} = 0, t \in T \\
(\lambda^-_{kt}, \lambda^+_{kt}) &: -TC_k \leq f_{kt} \leq TC_k, k \in K, t \in T
\end{aligned}
\tag{3.18}
$$

KKT system:

$$
\begin{aligned}
r_{nt} \text{ free } &\perp \rho_{nt} + \phi_t - \sum_{k \in K} PTDF_{kn} \cdot \psi_{kt} = 0 \\
f_{kt} \text{ free } &\perp \psi_{kt} + \lambda^+_{kt} - \lambda^-_{kt} = 0 \\
\psi_{kt} \text{ free } &\perp f_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
\phi_t \text{ free } &\perp \sum_{n \in N} r_{nt} = 0, t \in T \\
0 \leq \lambda^+_{kt} &\perp TC_k - f_{kt} \geq 0, k \in K, t \in T \\
0 \leq \lambda^-_{kt} &\perp TC_k + f_{kt} \geq 0, k \in K, t \in T
\end{aligned}
\tag{3.19}
$$

Auctioneer. The Walrassian auctioneer maximizes the excess demand at each node and at each period, i.e. for each $n \in N, t \in T$

$$\max_{\rho_{nt}} \ \rho_{nt}(r_{nt} + D_{nt} - \sum_i y_{int} - s_{zt}) \tag{3.20}$$

KKT system:

$$\rho_{nt} \text{ free } \perp r_{nt} + D_{nt} - \sum_i y_{int} - s_{nt} = 0, n \in N, t \in T \tag{3.21}$$

A pure strategy Nash equilibrium is obtained when all agents solve simultaneously their profit-maximization problem, that is when the KKT conditions of each agent hold simultaneously. It is thus equivalent to the following linear

Mixed Complementarity Problem (MCP):

$$
\begin{aligned}
0 \leq x_{in} &\perp IC_i - \sum_{t \in T} \mu_{int} \geq 0, i \in I, n \in N \\
0 \leq y_{int} &\perp MC_i + \mu_{int} - \rho_{nt} \geq 0, i \in I, n \in N, t \in T \\
0 \leq \mu_{int} &\perp X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\
0 \leq s_{nt} &\perp VOLL - \rho_{nt} + \delta_{nt} \geq 0, n \in N, t \in T \\
0 \leq \delta_{nt} &\perp D_{nt} - s_{nt} \geq 0, n \in N, t \in T \\
r_{nt} \text{ free } &\perp \rho_{nt} + \phi_t - \sum_{k \in K} PTDF_{kn} \cdot \psi_{kt} = 0 \\
f_{kt} \text{ free } &\perp \psi_{kt} + \lambda^+_{kt} - \lambda^-_{kt} = 0 \\
\psi_{kt} \text{ free } &\perp f_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
\phi_t \text{ free } &\perp \sum_{n \in N} r_{nt} = 0, t \in T \\
0 \leq \lambda^+_{kt} &\perp TC_k - f_{kt} \geq 0, k \in K, t \in T \\
0 \leq \lambda^-_{kt} &\perp TC_k + f_{kt} \geq 0, k \in K, t \in T \\
\rho_{nt} \text{ free } &\perp r_{nt} + D_{nt} - \sum_i y_{int} - s_{nt} = 0, n \in N, t \in T
\end{aligned}
\tag{3.22}
$$

One observes that this MCP is exactly equivalent to the set of KKT conditions of the central planner's capacity expansion problem in nodal pricing (3.2). This shows the equivalence between the centralized and decentralized capacity expansion problem in nodal pricing.

3.B Decentralization of capacity expansion under FBMC

In order to describe the decentralization of the capacity expansion problem under FBMC, we proceed similarly to the nodal pricing case of appendix section 3.A. We first formulate the profit maximization problem of the four types of agents (consumers, producers, TSO and auctioneer). The problem of the producers, the consumers and the auctioneer are essentially the same, except that the variables are zonal. The problem of the TSO is modified to take into account the FBMC constraints.

Producers. For each $i \in I, z \in Z$

$$\begin{aligned} &\max_{x_{iz}} \sum_{t \in T} \Big((\rho_{zt} - MC_i) y_{izt} \Big) - IC_i x_{iz} \\ &(\mu_{izt}) : X_{iz} + x_{iz} - y_{izt} \geq 0 \\ &\quad x_{iz} \geq 0, y_{izt} \geq 0 \end{aligned} \tag{3.23}$$

where $X_{iz} = \sum_{n \in Z(n)} X_{in}$.

Consumers. For each $z \in Z$

$$\begin{aligned} &\max_{s_{zt}} \sum_{t \in T} (VOLL - \rho_{zt})(D_{zt} - s_{zt}) \\ &\text{s.t. } (\delta_{zt}) : D_{zt} - s_{zt} \geq 0, t \in T \\ &\quad s_{zt} \geq 0 \end{aligned} \tag{3.24}$$

where $D_{zt} = \sum_{n \in Z(n)} D_{nt}$.

TSO. The TSO maximizes its congestion rent while respecting the FBMC network constraints, i.e.

$$\begin{aligned} &\max_{p_{zt}} - \sum_{z \in Z, t \in T} p_{zt} \rho_{zt} \\ &(\nu_z) : \sum_{i \in I} x_{iz} - \sum_{n \in N(z)} \tilde{x}_n = 0, z \in Z \\ &(\tilde{\rho}_{zt}) : p_{zt} - \sum_{n \in N(z)} \tilde{r}_{nt} = 0, z \in Z, t \in T \\ &(\tilde{\rho}_{nt}) : \tilde{r}_{nt} - \tilde{y}_{nt} + D_{nt} = 0, n \in N, t \in T \\ &(\tilde{\mu}_{nt}) : \sum_{i \in I} X_{in} + \tilde{x}_n - \tilde{y}_{nt} \geq 0, n \in N, t \in T \\ &(\psi_{kt}) : \tilde{f}_{kt} - \sum_{n} PTDF_{nk} \cdot \tilde{r}_{nt} = 0, k \in K, t \in T \\ &(\phi_t) : \sum_{n} \tilde{r}_{nt} = 0, t \in T \\ &(\lambda^-_{kt}, \lambda^+_{kt}) : -TC_k \leq \tilde{f}_{kt} \leq TC_k, k \in K, t \in T \end{aligned} \tag{3.25}$$

The tilde notation is used here in order to represent variables that do not correspond directly to physical quantities, but are used as extended variables for representing the feasible set of net positions in FBMC.

Auctioneer. The Walrasian auctioneer maximizes the excess demand at each node and at each period, i.e. for each $n \in N, t \in T$

$$\max_{\rho_{nt}} \ \rho_{nt}(r_{nt} + D_{nt} - \sum_i y_{int} - s_{zt}) \tag{3.26}$$

Here, we are not in the classical setting of a Nash equilibrium because the set of possible decisions of the TSO depends on the decisions of other agents, namely the investment x_{iz} of the producers. We are thus in the setting of a GNE [Har91], which corresponds to the joint solution of the KKT conditions of all agents. It can thus be formulated as the following MLCP:

$$\begin{aligned}
0 \le x_{iz} &\perp IC_i - \sum_{t \in T} \mu_{izt} \ge 0, i \in I, z \in Z \\
0 \le \mu_{izt} &\perp X_{iz} + x_{iz} - y_{izt} \ge 0, i \in I, z \in Z, t \in T \\
0 \le y_{izt} &\perp MC_i + \mu_{izt} - \rho_{zt} \ge 0, i \in I, z \in Z, t \in T \\
0 \le s_{zt} &\perp VOLL - \rho_{zt} + \delta_{zt} \ge 0, z \in Z, t \in T \\
0 \le \delta_{zt} &\perp D_{zt} - s_{zt} \ge 0, z \in Z, t \in T \\
\nu_z \text{ free } &\perp \sum_{i \in I} x_{iz} - \sum_{n \in N(z)} \tilde{x}_n = 0, z \in Z \\
\rho_{zt} \text{ free } &\perp -p_{zt} + \sum_{i \in I} y_{izt} + s_{zt} - D_{zt} = 0, z \in Z, t \in T \\
\tilde{\rho}_{zt} \text{ free } &\perp -p_{zt} + \sum_{n \in N(z)} \tilde{y}_{nt} - D_{zt} = 0, z \in Z, t \in T \\
0 \le \tilde{y}_{nt} &\perp \tilde{\mu}_{nt} - \tilde{\rho}_{Z(n)t} - \tilde{\rho}_{nt} \ge 0, i \in I, n \in N, t \in T \\
0 \le \tilde{x}_n &\perp -\sum_{t \in T} \tilde{\mu}_{nt} + \nu_{Z(n)} \ge 0, n \in N \\
p_{zt} \text{ free } &\perp \rho_{zt} + \tilde{\rho}_{zt} = 0, z \in Z, t \in T \\
r_{nt} \text{ free } &\perp \tilde{\rho}_{nt} - \phi_t + \sum_{k \in K} PTDF_{kn} \cdot \psi_{kt} = 0 \\
f_{kt} \text{ free } &\perp -\psi_{kt} + \lambda^+_{kt} - \lambda^-_{kt} = 0 \\
0 \le \tilde{\mu}_{int} &\perp X_{in} - \tilde{y}_{int} \ge 0, i \in I, n \in N, t \in T \\
\tilde{\rho}_{nt} \text{ free } &\perp -r_{nt} + \tilde{y}_{nt} - D_{nt} = 0, n \in N, t \in T \\
\phi_t \text{ free } &\perp \sum_{n \in N} r_{nt} = 0, t \in T \\
\psi_{kt} \text{ free } &\perp f_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
0 \le \lambda^+_{kt} &\perp TC_k - f_{kt} \ge 0, k \in K, t \in T \\
0 \le \lambda^-_{kt} &\perp TC_k + f_{kt} \ge 0, k \in K, t \in T
\end{aligned} \tag{3.27}$$

3.C Zonal model with price aggregation

Let us first write the economic dispatch problem with nodal transmission constraints and fixed nodal capacity X_{in}:

$$\begin{aligned}
&\min_{y,r,f} \sum_{i\in I, n\in N} MC_i y_{in} \\
(\mu_{in}) &: X_{in} - y_{in} \geq 0, i \in I, n \in N \\
(\rho_n) &: -r_n + \sum_{i\in I} y_{in} - D_n = 0, n \in N \\
(\psi_k) &: f_k - \sum_{n\in N} PTDF_{kn} \cdot r_n = 0, k \in K \\
(\phi) &: \sum_{n\in N} r_n = 0 \\
(\lambda_k^-, \lambda_k^+) &: -TC_k \leq f_k \leq TC_k, k \in K \\
&y \geq 0
\end{aligned} \tag{3.28}$$

The nodal market clearing prices can be obtained by solving the dual of this problem:

$$\begin{aligned}
&\max_{\rho,\mu,\psi,\phi,\lambda} \sum_n D_n \rho_n - \sum_{in} X_{in}\mu_{in} - \sum_k TC_k(\lambda_k^+ + \lambda_k^-) \\
(r_n) &: \rho_n + \sum_k PTDF_{kn}\psi_k - \phi = 0, n \in N \\
(y_{in}) &: MC_i - \rho_n + \mu_{in} \geq 0, i \in I, n \in N \\
(f_k) &: -\psi_k - \lambda_k^- + \lambda_k^+ = 0, k \in K \\
&\mu, \lambda^+, \lambda^- \geq 0
\end{aligned} \tag{3.29}$$

Applying directly the fundamental property of zonal pricing, which is that the prices within the same zone should be equal, we obtain a natural zonal extension of nodal pricing by adding to model (3.29) the following constraints:

$$\rho_{n_1} = \rho_{n_2}, \ \ \forall (n_1, n_2) \in z, \forall z \in Z \tag{3.30}$$

This is equivalent to introducing a new variable ρ_z for each zone z, corresponding to the price of the zone, and imposing the following constraints:

$$\rho_n = \rho_z, \ \ \forall n \in N(z), \forall z \in Z \tag{3.31}$$

This in turn is equivalent to replacing every occurrence of ρ_n by $\rho_{Z(n)}$ in (3.29), yielding:

$$
\begin{aligned}
\max_{\rho,\mu,\psi,\phi,\lambda} & \sum_n D_n \rho_{Z(n)} - \sum_{in} X_{in}\mu_{in} - \sum_k TC_k(\lambda_k^+ + \lambda_k^-) \\
(r_n) : & \rho_{Z(n)} + \sum_k PTDF_{kn}\psi_k - \phi = 0, n \in N \\
(y_{in}) : & MC_i - \rho_{Z(n)} + \mu_{in} \geq 0, i \in I, n \in N \\
(f_k) : & -\psi_k - \lambda_k^- + \lambda_k^+ = 0, k \in K \\
& \mu, \lambda^+, \lambda^- \geq 0
\end{aligned}
\tag{3.32}
$$

Model (3.32) will thus produce zonal market clearing prices and is thus a model for clearing the market under the zonal pricing paradigm. As every zonal pricing model, it involves a simplification of the transmission constraints, that will become clear when moving back to the primal space:

$$
\begin{aligned}
\min_{y,r,f} & \sum_{i\in I, n\in N} MC_i y_{in} \\
(\mu_{in}) : & X_{in} - y_{in} \geq 0, i \in I, n \in N \\
(\rho_z) : & -\sum_{n\in N(z)} r_n + \sum_{i\in I, n\in N(z)} y_{in} - \sum_{n\in N(z)} D_n = 0, z \in Z \\
(\psi_k) : & f_k - \sum_{n\in N} PTDF_{kn} \cdot r_n = 0, k \in K \\
(\phi) : & \sum_{n\in N} r_n = 0 \\
(\lambda_k^-, \lambda_k^+) : & -TC_k \leq f_k \leq TC_k, k \in K \\
& y \geq 0, s \geq 0
\end{aligned}
\tag{3.33}
$$

We observe that nodal variables y_{in} of the same zone are not distinguished in the transmission constraints. They can therefore be aggregated into y_{iz}. Finally, we can introduce a variable corresponding to the zonal net injection p_z defined by

$$p_z = \sum_{n\in N(z)} r_n$$

Problem (3.32) is thus found to be equivalent to the following zonal problem:

$$
\begin{aligned}
&\min_{y,r,f} \sum_{i\in I, z\in Z} MC_i y_{iz} \\
&X_{iz} - y_{iz} \geq 0, i \in I, z \in Z \\
&-p_z + \sum_{i\in I} y_{iz} - D_z = 0, z \in Z \\
&p_z - \sum_{n\in N(z)} r_n = 0, z \in Z \\
&f_k - \sum_{n\in N} PTDF_{kn} \cdot r_n = 0, k \in K \\
&\sum_{n\in N} r_n = 0 \\
&-TC_k \leq f_k \leq TC_k, k \in K \\
&y \geq 0, s \geq 0
\end{aligned}
\tag{3.34}
$$

We now go on to prove the conditions of equivalence between the nodal and zonal models stated in section 3.3.

Proof of Proposition 3.1. We start by clarifying that the zonal graph is the pair (Z, L^{inter}) where Z is the set of zones and L^{inter} is the set of inter-zonal lines, i.e.

$$
L^{\text{inter}} = \left\{ \Big(Z(m(k)), Z(n(k)) \Big) \ \forall k \in L \text{ s.t. } Z(m(k)) \neq Z(n(k)) \right\}, \tag{3.35}
$$

where $m(k)$ and $n(k)$ are respectively the source and destination nodes of line k. Similarly, one can define the set of intra-zonal lines: $L^{\text{intra}} = L \backslash L^{\text{inter}}$. The first hypothesis, stating that the intra-zonal transmission constraints are never binding, implies that $\lambda_k^+ = \lambda_k^- = 0 \ \forall k \in L^{\text{intra}}$ in the dual of the nodal model (3.29). This in turns implies that $\psi_k = 0 \ \forall k \in L^{\text{intra}}$. Using the first equation of model (3.29), we get

$$
\rho_n = - \sum_{k\in L^{\text{inter}}} PTDF_{kn} \psi_k + \phi, n \in N
$$

We show that the right-hand side of this equation is the same for nodes in the same zone by looking specifically at the PTDF. Let us consider a given zone z. For this zone, one must distinguish two cases: (i) the hub node used to construct the PTDF belongs to zone z and (ii) the hub node does not belong to the zone. If the hub node belongs to zone z, then, by the second hypothesis stating that the zonal network is radial, there is no path of power from a node of that zone to the hub node that passes through an inter-zonal line, which implies that $PTDF_{kn} = 0$ for all $k \in L^{\text{inter}}$ and for all $n \in z$ and, in turn, that $\rho_n = \phi,\ \forall n \in z$. If the hub node does not belong to zone z, then, by the second hypothesis, there is a unique path in the zonal graph between zone z and the zone to which the hub node belongs. Let us denote by L_z^{inter} the set of lines (that are all inter-zonal) that belong to this unique path. Clearly, every path from any node in zone z to the hub node passes through every line of L_z^{inter} and there is no path of power that goes through $L^{\text{inter}} \backslash L_z^{\text{inter}}$. This means that, for every node in zone z, $PTDF_{kn}, k \in L^{\text{inter}}$ will be either 1 if $k \in L_z^{\text{inter}}$ or 0 if $k \in L^{\text{inter}} \backslash L_z^{\text{inter}}$. This yields the following equation:

$$\rho_n = - \sum_{k \in L_z^{\text{inter}}} \psi_k + \phi,\ \ \forall n \in z$$

Once again, the right-hand side of this equation does not depend on n. One can thus conclude that, under the two hypotheses of Proposition 3.1, the nodal market will clear with equal nodal prices for nodes in the same zone. As this is the only thing that we impose for deriving (3.32) from (3.29), (3.29) and (3.32) are equivalent, which in turn implies that (3.28) and (3.33) are equivalent. Note that the result is proven for the short-term market, but the proof can be adapted using the same arguments for capacity expansion models. □

3.D Existence and uniqueness of solutions to the decentralized problem

In this section, we discuss the properties of the decentralized FBMC problem regarding existence and uniqueness of solutions.

3.D.1 Existence

We have already mentioned that the decentralized investment problem with FBMC might not have a solution if the TSO does not invest. When the TSO is allowed to invest in strategic reserve, a solution can be recovered. We discuss here whether the existence of a solution is guaranteed in this second case. It turns out that it is indeed the case, as we now show. To prove existence, we rely on the theory of LCP and in particular on the properties of LCP with copositive matrices and their homogeneous counterpart. We start by noting that, using the transformation of MLCPs to LCPs introduced in section 1.3, the decentralized FBMC capacity expansion problem with strategic reserve can be formulated as the following LCP:

$$0 \leq x_{iz} \perp IC_i - \sum_{t \in T} \mu_{izt} \geq 0, i \in I, z \in Z \tag{3.36a}$$

$$0 \leq \mu_{izt} \perp X_{iz} + x_{iz} - y_{izt} \geq 0, i \in I, z \in Z, t \in T \tag{3.36b}$$

$$0 \leq y_{izt} \perp MC_i + \mu_{izt} - \rho_{zt}^{+} + \rho_{zt}^{-} \geq 0, i \in I, z \in Z, t \in T \tag{3.36c}$$

$$0 \leq s_{zt} \perp VOLL - \rho^+_{zt} + \rho^-_{zt} + \delta_{zt} \geq 0, z \in Z, t \in T \tag{3.36d}$$
$$0 \leq \delta_{zt} \perp D_{zt} - s_{zt} \geq 0, z \in Z, t \in T \tag{3.36e}$$
$$0 \leq \nu_z \perp \sum_{i \in I} x_{iz} - \sum_{n \in N(z)} \tilde{x}_n \geq 0, z \in Z \tag{3.36f}$$
$$0 \leq \rho^+_{zt} \perp -p^+_{zt} + p^-_{zt} + \sum_{i \in I} y_{izt} + s_{zt} - D_{zt} \geq 0, z \in Z, t \in T \tag{3.36g}$$
$$0 \leq \rho^-_{zt} \perp p^+_{zt} - p^-_{zt} - \sum_{i \in I} y_{izt} - s_{zt} + D_{zt} \geq 0, z \in Z, t \in T \tag{3.36h}$$
$$0 \leq \tilde{\rho}^+_{zt} \perp -p^+_{zt} + p^-_{zt} + \sum_{n \in N(z)} \tilde{y}_{nt} + \sum_{n \in N(z)} z_{nt} - D_{zt} = 0, z \in Z, t \in T \tag{3.36i}$$
$$0 \leq \tilde{\rho}^-_{zt} \perp p^+_{zt} - p^-_{zt} - \sum_{n \in N(z)} \tilde{y}_{nt} - \sum_{n \in N(z)} z_{nt} + D_{zt} = 0, z \in Z, t \in T \tag{3.36j}$$
$$0 \leq \tilde{y}_{nt} \perp \tilde{\mu}_{nt} - \tilde{\rho}^+_{Z(n)t} + \tilde{\rho}^-_{Z(n)t} - \tilde{\rho}^+_{nt} + \tilde{\rho}^-_{nt} \geq 0, n \in N, t \in T \tag{3.36k}$$
$$0 \leq \tilde{x}_n \perp -\sum_{t \in T} \tilde{\mu}_{nt} + \nu_{Z(n)} \geq 0, n \in N \tag{3.36l}$$
$$0 \leq p^+_{zt} \perp \rho^+_{zt} - \rho^-_{zt} + \tilde{\rho}^+_{zt} - \tilde{\rho}^-_{zt} \geq 0, z \in Z, t \in T \tag{3.36m}$$
$$0 \leq p^-_{zt} \perp -\rho^+_{zt} + \rho^-_{zt} - \tilde{\rho}^+_{zt} + \tilde{\rho}^-_{zt} \geq 0, z \in Z, t \in T \tag{3.36n}$$
$$0 \leq \tilde{r}^+_{nt} \perp \tilde{\rho}^+_{nt} - \tilde{\rho}^-_{nt} - \phi^+_t + \phi^-_t + \sum_{k \in K} PTDF_{kn} \cdot (\psi^+_{kt} - \psi^-_{kt}) \geq 0 \tag{3.36o}$$
$$0 \leq \tilde{r}^-_{nt} \perp -\tilde{\rho}^+_{nt} + \tilde{\rho}^-_{nt} + \phi^+_t - \phi^-_t - \sum_{k \in K} PTDF_{kn} \cdot (\psi^+_{kt} - \psi^-_{kt}) \geq 0 \tag{3.36p}$$
$$0 \leq f^+_{kt} \perp -\psi^+_{kt} + \psi^-_{kt} + \lambda^+_{kt} - \lambda^-_{kt} \geq 0 \tag{3.36q}$$
$$0 \leq f^-_{kt} \perp \psi^+_{kt} - \psi^-_{kt} - \lambda^+_{kt} + \lambda^-_{kt} \geq 0 \tag{3.36r}$$
$$0 \leq \tilde{\mu}_{nt} \perp \sum_{i \in I} X_{in} - \tilde{y}_{nt} \geq 0, n \in N, t \in T \tag{3.36s}$$
$$0 \leq \tilde{\rho}^+_{nt} \perp -\tilde{r}^+_{nt} + \tilde{r}^-_{nt} + \tilde{y}_{nt} + z_{nt} - D_{nt} \geq 0, n \in N, t \in T \tag{3.36t}$$
$$0 \leq \tilde{\rho}^-_{nt} \perp \tilde{r}^+_{nt} - \tilde{r}^-_{nt} - \tilde{y}_{nt} - z_{nt} + D_{nt} \geq 0, n \in N, t \in T \tag{3.36u}$$

$$0 \leq \phi^+_t \perp \sum_{n \in N} \tilde{r}^+_{nt} - \tilde{r}^-_{nt} \geq 0, t \in T \tag{3.1v}$$
$$0 \leq \phi^-_t \perp \sum_{n \in N} \tilde{r}^-_{nt} - \tilde{r}^+_{nt} \geq 0, t \in T \tag{3.1w}$$
$$0 \leq \psi^+_{kt} \perp f^+_{kt} - f^-_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} \geq 0, k \in K, t \in T \tag{3.1x}$$
$$0 \leq \psi^-_{kt} \perp -f^+_{kt} + f^-_{kt} + \sum_{n \in N} PTDF_{kn} \cdot r_{nt} \geq 0, k \in K, t \in T \tag{3.1y}$$

$$0 \leq \lambda^+_{kt} \perp TC_k - f^+_{kt} + f^-_{kt} \geq 0, k \in K, t \in T \tag{3.1z}$$

$$0 \leq \lambda^-_{kt} \perp TC_k + f^+_{kt} - f^-_{kt} \geq 0, k \in K, t \in T \tag{3.1aa}$$

$$0 \leq Z_n \perp \tilde{IC} - \sum_{t \in T} \gamma_{nt} \geq 0, n \in N \tag{3.1ab}$$

$$0 \leq z_{nt} \perp \tilde{MC} + \gamma_{nt} - \tilde{\rho}^+_{nt} + \tilde{\rho}^-_{nt} - \tilde{\rho}^+_{Z(n)t} + \tilde{\rho}^-_{Z(n)t} \geq 0, n \in N, t \in T \tag{3.1ac}$$

$$0 \leq \gamma_{nt} \perp Z_n - z_{nt} \geq 0, n \in N, t \in T \tag{3.1ad}$$

This is essentially problem (3.27) transformed into an LCP, and where we have introduced two primal variables: Z_n, which corresponds to the capacity invested by the TSO in strategic reserve at node n, and z_{nt}, which corresponds to the production by strategic reserve units at node n and in period t. The TSO can participate to the re-dispatch with the strategic reserve and variables γ_{nt} are thus added in the auxiliary dispatch. There is also a new dual variable γ_{nt} which is associated to the capacity constraint on the strategic reserve production. The parameters $\tilde{IC}$ and $\tilde{MC}$ are respectively the investment and marginal cost of the strategic reserve. All free variables have been decomposed into two positive variables to write the problem in the form of an LCP. Note that this problem is still exactly equivalent to the KKT conditions of the centralized FBMC problem with strategic reserve, except for the first equation, which represents the condition for investment in which variable ν_z has been removed.

To prove existence of solutions, we apply Theorem 1.1. We first show that the matrix M of our LCP (3.36) is copositive. To do this, we note that M is the sum of two matrices:

$$M = \tilde{M} + \tilde{N}$$

where $\tilde{M}$ is skew-symmetric (it is the matrix associated to the equivalent LCP of the centralized problem) and where $\tilde{N}$ is of the following form (in block formulation):

$$\tag{3.2}$$

$$\tilde{N} = \begin{array}{c c} & \begin{array}{ccccc} & & \nu_z & & \end{array} \\ \begin{array}{c} x_{iz} \\ \\ \\ \\ \end{array} & \begin{pmatrix} 0 & \dots & I & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \end{array} \tag{3.3}$$

Here, I is the rectangular identity matrix, i.e. a matrix with 1 in the entries associated to line x_{iz} and column ν_z, and 0 otherwise. This implies that

$$v^\top M v = v^\top \tilde{M} v + v^\top \tilde{N} v = v^\top \tilde{N} v = \sum_{iz} x_{iz} \nu_z$$

where v is the full vector of variables, which is indeed nonnegative if each x_{iz} and ν_z are nonnegative.

Now, let $v^* = (x^*_{iz}, y^*_{izt}, \dots, \gamma^*_{nt})$ be a solution to the homogeneous version of LCP (3.36). Using (3.1ab) and (3.1ad), we have that $\gamma^*_{nt} = 0$. From (3.1ac), we deduce that $\tilde{\rho}^{*+}_{nt} - \tilde{\rho}^{*-}_{nt} \leq -\tilde{\rho}^{*+}_{Z(n)t} + \tilde{\rho}^{*-}_{Z(n)t}$. (3.36a), (3.36b) and (3.36c) imply that $\rho^{*+}_{zt} - \rho^{*-}_{zt} \leq 0$ which in turn leads to $\tilde{\rho}^{*+}_{zt} - \tilde{\rho}^{*-}_{zt} \geq 0$ by (3.36m) and (3.36n) and thus to $\tilde{\rho}^{*+}_{nt} - \tilde{\rho}^{*-}_{nt} \leq 0$. Finally, we have that

$$
\begin{aligned}
q^\top v^* = &\sum_{i,z} IC_i x^*_{iz} + \sum_{i,z,t} MC_i y^*_{izt} + \sum_{z,t} V s^*_{z,t} + \sum_{kt} TC_k(\lambda^{+*}_k + \lambda^{-*}_k) + \\
&\sum_{int} \tilde{X}_{in} \tilde{\mu}^*_{int} + \sum_n \tilde{IC}_n Z^*_n - \\
&\sum_{zt} D_{zt}(\rho^{*+}_{zt} - \rho^{*-}_{zt} + \tilde{\rho}^{*+}_{zt} - \tilde{\rho}^{*-}_{zt}) - \sum_{nt} D_{nt}(\tilde{\rho}^{*+}_{nt} - \tilde{\rho}^{*-}_{nt})
\end{aligned}
$$

All the terms of the first and second lines are nonnegative, as they correspond to the product of nonnegative quantities. In the third line, the first term is zero by equations (3.36m) and (3.36n) and the second term is nonnegative by the non-positivity of $\tilde{\rho}^{*+}_{nt} - \tilde{\rho}^{*-}_{nt}$ that we have shown above. By Theorem 1.1, we can conclude that the LCP (3.36) has a solution. We note that the non-positivity of $\tilde{\rho}^{*+}_{nt} - \tilde{\rho}^{*-}_{nt}$ has been obtained using the equations of the strategic reserve investment and thus that this argument does not hold without the strategic reserve, which is consistent with our observation that, in that case, the problem is in general infeasible.

3.D.2 Uniqueness

In order to investigate the uniqueness of solutions to the decentralized FBMC, we return to the illustrative two-zone, three-node example. Recall that we have found the following solution to the problem:

Policy	Inv. in node B	Inv. in node A_s
FBMC-D	Coal: 1918MW Nuclear: 7086MW Gas: 2015MW	NR: 50MW

Initializing the solver with a large number of starting points, we identify a second solution to the LCP (3.36), different from the one reported in section 3.4. This solution is the following:

Policy	Inv. in node B	Inv. in node A_s
FBMC-D	Coal: 1918MW Nuclear: 7086MW Gas: 1715MW	NR: 200MW

Using these two solutions, we can check whether their convex combination is also part of the solution set, as would be the case in an LCP derived from a linear program. It turns out that this is not the case[8] and thus that the solution set of our problem is non-convex but is instead the union of finitely many polytopes.

Moreover, the determinant of the Jacobian matrix at both solutions is zero, which suggests that the solutions are not isolated. We confirm this numerically by solving the problem of the maximization of the distance to a solution in its implied polytope, which is unbounded. In the illustrative example, one extreme ray is given by $t\nu_1$ for $t > 0$.

We note that this potential multiplicity of equilibria can be considered as an additional inefficiency related to the current methodology of FBMC as it relates to investment. This source of inefficiency is distinct from the one that is discussed in the main body of the chapter, and relates to the fact that the total cost can be degraded in one of these equilibria compared to nodal and centralized FBMC. In the main body of the chapter, we compute and discuss only one of these potentially multiple equilibria. Computing several of them is an interesting question. However, it is computationally intractable to do so in a systematic way for the instance of the size that we consider in this chapter. A more formal treatment of this multiplicity of equilibria and their implication is thus relegated to future research.

[8]It can be checked that the mid point between the two identified solutions is not itself a solution.

4 Locational instruments for efficient power generation investment under zonal pricing

4.1 Introduction

4.1.1 Motivation

The debate between nodal and zonal pricing is by all means an old story in the US evolution of the power system. In their three-phase review of the history of the process, [HO19] note, almost in passing, that the subject had been settled in the first phase, that is after the Californian experience: the zonal system had fundamental flaws that the nodal system avoided. The Texas experience came a bit later but confirmed the observation [TW21]. The situation is quite different in Europe: a recent report [ACE22] issued in response to criticisms against the high prices observed in the power market in the end of 2021 (that is well before the perturbations on the gas market induced by the events in Ukraine) argues that the current European (zonal) system is functioning properly but that one might consider some improvements, among which a reduction of the size of the zones. In particular, the nodal system is mentioned as an example of this reduction (page 27 of [ACE22]). This difference between the US and EU points of view after so many years of theory and experience on a subject that looks purely technical is strange and must have some fundamental explanation. It cannot be a misunderstanding of the nodal system, which had been described in the French literature as early as [BS52]. But institutions suggest a possible reason for divergent points of view, as we discuss in detail in the conclusion of this dissertation (section 5.3). The restructuring of the power system in Europe was part of the integration of sectoral markets that created the internal market in 1992, with delays expected in “difficult sectors”, electricity being one of them. It suffices to note, here, that a nodal model in an integrated European market would have merged national systems and interconnections into a new supranational model where all components of national networks would be placed on the same footing. Conversely, the coupling of the zonal national systems maintained their individuality and developed a framework facilitating exchanges between them through interconnections. In other words, the zonal system in the internal EU market would have had a supranational flavor that the zonal system was thought to be able to bypass.

In contrast with what is sometimes common (both in public perception in Europe and the US), Europe is a very weak supranational entity with resistance against any further integration beyond what was acquired in the early days. The supranational dimension implicit in nodal pricing certainly did not help in accelerating its adoption in Europe. This chapter is an attempt to deal with this issue: it takes the existing day-ahead zonal system as given but attempts to complete it by addressing the needed integration at another level. This is done by inserting features with a nodal flavor in market processes on which stakeholders are still working. Needless to say, the economic and physical requirements remain and one could argue that our proposals look like a covering up of the nodal system (referring to Molière's Tartuffe, "Cachez ce sein que je ne saurais voir" adapted here into "Cover up that nodal system, which I can't endure to look upon"). Instead, in our work, we attempt the contrary: we do not try to cover up the remedy, but to make it clearly visible. The requirements of basic physics and economics embedded in the nodal system, namely the need to internalize the externalities created by Kirchhoff's laws, come up in our analysis depending on the remedies considered and they are always clearly referred to. Our attempt in this work is to fully insert them in the zonal system.

Our main interest is to understand the theoretical conditions under which additional market-based instruments could restore the efficiency of nodal pricing in the long run. In order to identify candidate market-based instruments for restoring the efficiency of the zonal design, we base ourselves on the thorough review of existing locational instruments provided in [EKH20]. To the list of [EKH20], we add market-based re-dispatch as a market-based way to introduce a locational component in the price. In total, we consider three main classes of candidates: capacity-based instruments, energy-based instruments and market-based re-dispatch. The present chapter analyses the combination of zonal pricing and these three classes of locational instruments under a unifying modeling framework in order to study their efficiency, first from a theoretical point of view and then, based on simulations, in a realistic case study.

4.1.2 Related literature

There is an emerging stream of literature on the study of the impact of locational instruments in zonal markets from a quantitative perspective. The first paper that tackles this problem quantitatively is [GRSZ19]. The authors propose a tri-level model of the long-run equilibrium of zonal pricing with cost-based re-dispatch. The tri-level structure aims at representing the sequential nature of the problem: grid investment by the TSO, capacity investment by private firms and re-dispatching in the short term by the TSO. A simplified spatially differentiated capacity signal based on average nodal prices is then added to the model. The authors find that, although it influences the location of investment, the introduction of a well-calibrated capacity signal only slightly improves welfare, as the signal fails to also impact the operational efficiency of the system. [SZ20] study the impact of uniform pricing in Germany on the siting of wind generation. They find that nodal pricing, by incentivizing the siting of investment closer to load centers and thereby reducing wind curtailment, has a positive effect on welfare. The paper also investigates the restoration of a locational component to the uniform price through capacity-based latitude-dependent connection charges. However, the authors find that these simplified locational instruments are not adequate for mitigating the inefficiencies associated to uniform pricing. Finally, [Eic21] proposes a model for quantifying the optimal capacity-based locational signal for restoring the efficiency of investment in zonal pricing with cost-based re-dispatch. Unlike the two papers previously cited, this paper is the first to consider a technology-differentiated signal. The author shows that differentiating by technology is a necessary condition for restoring the efficiency of nodal pricing. The model of [Eic21] differs in its structure from the model that we propose in the present chapter: it is a two-stage model where the regulator first decides on the capacity signal and then, private firms react to the regulator's decision by investing in generation capacities taking into account the capacity signal. The author finds that, although locational capacity signals have a significant cost-saving potential, they do not lead to a full recovery of the efficiency of nodal pricing.

Our work also relates to the literature on the quantitative analysis of zonal pricing with market-based re-dispatch. In the early days of the discussions on market-based congestion management in Europe, [DH02] analyzed five different designs among which uniform pricing followed by market-based re-dispatch, which was thus referred to as *countertrading*. The authors proposed a stylized two-node one-zone example and concluded that, in theory, countertrading is efficient in the short term. The model does not account for inc-dec gaming, which is an important aspect in market-based re-dispatch. Potential inefficiencies stemming from deviations from this idealized setting analyzed in the paper are discussed qualitatively in the work of [DH02].

One situation that could be associated to inefficiencies in market-based re-dispatch in the short term takes place when the TSO does not act as a price-taker in the re-dispatch market. This situation is analyzed by [GMS$^+$18] under the assumption of absence of inc-dec gaming. The authors show, on an illustrative example, that, in this case, the outcome could be inefficient.

An important drawback of re-dispatch markets is that they offer the possibility to market participants to engage in arbitrage between the zonal and re-dispatch price, the so-called inc-dec game, leading to a distortion of long-run investment incentives as well as the extraction of short-term rents by those exercising the gaming strategy for essentially offering nothing to the system. This situation is analyzed quantitatively in [HL15]. The authors show that inc-dec gaming is an arbitrage strategy that also occurs under perfect competition. They conclude that, under certain restrictive assumptions, zonal pricing followed by market-based re-dispatch is efficient in the short term, and the outcomes only differ by a redistribution of the welfare.

More recently, [HS20] proposed a simple model on a two-node one-zone example in order to analyze inc-dec gaming. The authors emphasize that the design leads to undue arbitrage opportunities for market participants, even when they cannot exercise market power. The authors further discuss the conditions under which this arbitrage is exacerbated or can be mitigated.

4.1.3 Contributions

This chapter contributes to the existing literature in the following ways:

- Our main contribution is to propose a unifying modeling framework for analyzing the long-run investment equilibrium resulting from different market-based congestion management designs, including capacity and energy-based locational instruments, market-based re-dispatch, and several of their variations.

- Using this common framework, we establish theoretical conditions under which the long-run efficiency of nodal pricing can be recovered in a zonal pricing market.

- Regarding market-based re-dispatch, we propose a model that internalizes the investment decisions of private firms in generating capacity while accounting for inc-dec gaming. From a mathematical point of view, we show how the loss of efficiency of this design originates from the property of generalized Nash equilibirum of the associated game and formally establish the existence of such an equilibrium.

- We provide a comparison of the different designs on a large-scale instance of the Central Western European (CWE) network area and show how the generalized Nash equilibrium problem associated to market-based redispatch can be solved by a splitting-based algorithm that leverages its specific structure.

4.1.4 Organization of the chapter

The chapter is organized as follows: we start, in section 4.2, by providing a roadmap to the study. The terminology used in the chapter is described, the different designs that we analyze are listed and the main policy messages of the chapter are summarized. Section 4.3 is the main section of the chapter: it presents our models and our theoretical results. Then, in section 4.4, we propose a case study using a realistic instance of the CWE network area, on which we compare the efficiency of the different designs considered in the chapter. Section 4.5 concludes the chapter.

4.2 Roadmap to the study

4.2.1 Terminology

Markets. Our focus in this work is on the restoration of the efficiency of nodal pricing using additional market-based instruments. For this reason, to any locational instrument considered in the chapter, a market will be associated. We distinguish in total four different markets: (i) the *electricity market* which is the basic market, nodal or zonal, on which electricity is traded, (ii) a *capacity market*, which is the market associated to capacity-based locational instruments, (iii) an *energy market*, associated to energy-based locational instruments and (iv) a *re-dispatch market* which is a nodal market subsequent to the zonal electricity market that arises under market-based re-dispatch.

Prices. To every market, a price is associated. The price in the electricity market will be referred to as the electricity price, the price in the capacity market as the capacity price and so on.

Two-sided vs one-sided markets The capacity and energy markets can either be two-sided or one-sided. The two-sided case is the general case where the TSO and producers can be both buyers and sellers on these markets. The case of one-sided capacity or energy markets refers to the situation in which the TSO is only a buyer or a seller on these markets. For instance, if the TSO is only a buyer on the capacity market, it will only be willing to pay candidate investors but will not be paid through the capacity market.

Positive or negative prices. In the case of one-sided capacity or energy markets, there are two situations: either the TSO is a buyer and the producers are sellers or the other way around. In the former situation, the price is positive and the situation will thus be referred to as *one-sided capacity (energy) market with positive capacity (energy) price.* In the latter situation, we have a one-sided capacity (energy) market with negative capacity (energy) price.

4.2.2 Studied designs

In this section, we enumerate the main classes of designs that we analyze in this paper as well as their variations. Table 4.1 summarizes the list and states the main properties of each design.

We study three main classes of locational instruments combined with zonal pricing: (i) locational capacity markets, (ii) locational energy markets and (iii) market-based re-dispatch. In the two former classes, we assume that the re-dispatch is cost-based. These designs are compared against two benchmarks: (i) nodal pricing, which corresponds to the most efficient design possible under our assumptions and (ii) pure zonal pricing with cost-based re-dispatch, which

Design abbreviation	Class	Sense	Locational differentiation	Technological differentiation	Temporal differentiation	Literature
TSCM	capacity	+/-	✓	✓		[Eic21]
POSCM	capacity	+	✓	✓		
NOSCM	capacity	-	✓	✓		
TSCMNTD	capacity	+/-	✓			[GRSZ19, SZ20]
TSEM	energy	+/-	✓		✓	
OSEM	energy	+	✓		✓	
MBR	re-dispatch		✓		✓	[HL15, HS20]
MBRTSCM	re-dispatch & capacity	+/-	✓	✓	✓	
MBRPCM	re-dispatch & capacity	+	✓	✓	✓	

Table 4.1: Summary of the studied design and their main properties. The literature column cites existing papers in which the corresponding design has been studied.

is the closest to the currently implemented design in Europe. For each of the three main classes of instruments, we study a certain number of variations.

For locational capacity markets, we consider four variations: (a) two-sided capacity markets, (b) one-sided markets with a negative capacity price, (c) one-sided markets with a positive capacity price and (d) two-sided markets without technological differentiation. For the three former variations, we assume that the capacity price is differentiated among both locations and technologies, whereas the latter is only differentiated among locations.

For locational energy markets, we consider two variations: (a) two-sided and (b) one-sided energy markets. In both of these markets, the price is differentiated among locations and time periods.

For market-based re-dispatch, we consider three variations: (a) pure zonal pricing with market-based re-dispatch, where there is no additional market, (b) market-based re-dispatch with a two-sided capacity market, and (c) market-based re-dispatch with a one-sided capacity market. Note that these two latter variations therefore correspond to the only two designs among all that are studied in this chapter that are comprised of three markets: a zonal electricity market, a re-dispatch market and a capacity market. All other designs are comprised either of one single market (the nodal and zonal pricing benchmarks) or two markets.

4.2.3 Link with existing initiatives in Europe

We provide here a list of some of the closest examples of existing markets that implement certain locational instruments in Europe.

Locational capacity signal

Great Britain, Ireland and Sweden all implement a capacity-based instrument that is differentiated among locations. The main difference between these initiatives and our modeling framework is that, in these markets, the price signals are computed ex ante and, therefore, do not correspond to a market mechanism. Moreover, we note that none of these instruments are differentiated among technology. We describe below the specificities of the initiatives of each country.

Great Britain In Great Britain, generators are charged based on their installed capacity, and there is an explicit locational component based on a splitting of the network into 27 zones [THE19, EKH20]. The signal can be either a charge or a premium.

Ireland In Ireland, there is a capacity charge with a locational component, which is computed with the so-called reverse MW-mile approach. Quoting [Com03], "This approach allocates a share of the fixed costs of the network to the generator based on its usage of the transmission system, reflecting the fact that cost depends on the distance and direction that power is being transmitted as well as the level of power being transmitted." Generators can be credited for reducing overall transmission flows.

Sweden There is a nodal annual capacity charge in Sweden based on 51 nodes [THE19]. The locational differentiation is based solely on the latitude: the charge is higher in the North for generation and higher in the South for consumption, so that generators pay more in the North whereas consumers pay more in the South.

Locational capacity market

France There is a capacity market in France, but it has no spatial differentiation. However, as discussed in [EKH20], France has recently implemented a specific tender to build new capacity in Brittany that was facing adequacy issues. This particular capacity tender thus had explicit spatial differentiation.

Locational energy-based signal

Norway and Sweden are indicative examples of European markets that implement an energy-based signal with spatial differentiation. In both countries, the signal corresponds to a multiplier on the zonal price [EKH20].

Norway The Norwegian system has an energy component to the tariffs, the so-called "energiledd" (energy part). It depends on a "marginal loss rate" that is different for every location in the system [Sta19]. The loss factors are

Instrument	Example country	Specificity
Locational capacity signal		
	Great Britain	Differentiation based on 27 zones.
	Ireland	Specific methodology called the reverse MW-mile approach.
		Based on the distance, direction and level of transmitted power.
	Sweden	Depends only on the latitude.
Locational capacity market		
	France	Specific tender on top of capacity market for Brittany.
Locational energy signal		
	Norway	Marginal loss rate multiplier on energy price.
		Updated on a weekly basis.
	Sweden	Loss coefficient multiplier on the energy price.
		Linearly dependent on latitude.
		Updated on a yearly basis and fixed for the whole year.
Market-based re-dispatch		
		Mandatory for units > 60MW.
		Pay-as-bid.

Table 4.2: Summary of the different examples of existing implementations of locational instruments in Europe

updated every week based on a given load, and differentiated for day/night and weekday/weekend. They are administratively limited at ±15%.

Sweden In Sweden, energy charges are based on a loss coefficient that depends linearly on the latitude of the node [THE19]. The coefficients are updated on a yearly basis.

Market-based re-dispatch

The Netherlands The Dutch market is an indicative example of a market where re-dispatch is market-based, as documented in [Ten19]. Participation to re-dispatch is mandatory for units with a capacity that exceeds 60MW and voluntary for units with a lower capacity. The market is pay-as-bid.

Summary

Table 4.2 provides a summary of the different examples of existing implementations of locational market instruments in Europe.

4.2.4 Preview of the policy messages

From our analysis, we conclude that seven of the studied designs are theoretically able to recover the efficiency of the nodal benchmark: two-sided capacity markets with technology differentiation, one-sided capacity markets with a negative capacity price and technology differentiation, all three energy markets (whether two-sided or one-sided), market-based re-dispatch with a two-sided capacity market and market-based re-dispatch with a one-sided capacity market and a positive capacity price. These conclusions only hold under a set of strict assumptions. In particular, the efficiency of capacity and energy markets only holds under specific conditions on the feasible set of zonal net positions. These assumptions are not respected in the current European methodology of capacity calculation. We find that, although it restores a locational component in the electricity price, the least efficient solution is obtained in the case of zonal pricing with market-based re-dispatch. This is due to the distortion of long-run incentives due to arbitrage when the two markets co-exist, and confirms results of the literature that are obtained with short-term models.

These results are revisited from a theoretical point of view in section 4.3, and from an empirical point of view in section 4.4.

4.3 Analysis

4.3.1 Modeling framework and main assumptions

The goal of this chapter is to develop a modeling framework for comparing the different designs of market-based congestion management in the long run. We model the long-run economic equilibrium on the electricity market as a Nash equilibrium between three types of agents: a single TSO, the producers, and a Walrasian auctioneer that enforces market clearing conditions (i.e. as the power exchange). In order to focus on the difference between the designs regarding congestion management, we adopt a set of simplifying assumptions that allow us to isolate the effects related to congestion management. In particular, we assume that the market is perfectly competitive, and we model the agents as price takers. This implies that we ignore market power. Although real markets deviate from the situation of perfect competition, our view is that it remains essential to understand the performance of market designs under perfect competition, as it is unlikely that identified inefficiencies will disappear in imperfectly competitive markets. The assumption of perfect competition allows us to keep the models transparent and tractable.

We assume that the profit-maximizing problems of all agents are convex and that there are no inter-temporal operating constraints: in the short run, all periods are independent. We consider that producers can invest in generation capacity in a continuous way, and we ignore transmission capacity expansion, for which the continuity assumption would be unrealistic. Electricity demand is assumed to be known and inelastic[1].

The investment problem is a two-stage problem in which producers first decide on their investment and then decide on their production at each period

[1] It is worth emphasizing that this is a strong assumption in the current and future electricity system where demand response is expected to play an increasingly important role. More importantly, in a situation where it becomes very difficult to monitor the opportunity costs of consumers, the importance of the assumption of perfect cost-based re-dispatch, which is also a condition for the restoration of nodal efficiency, should not be underestimated.

in the short-run market given their investment decision. Under the assumption of perfect competition, however, the two-stage formulation is equivalent to a single-stage formulation where the agents decide at the same time their investment and their production in the short-run market [GOS13]. For this reason, throughout the chapter we consider only single-stage formulations.

In the simplest situation, i.e. when there is no additional market instrument, agents compete on a single market, the electricity market, with locationally differentiated prices (with nodal granularity for nodal pricing and zonal granularity for zonal pricing). Throughout the chapter, we add different market instruments (capacity or energy-based locational instruments, as well as market-based re-dispatch). These market instruments are represented as additional markets in which the agents compete. For the same reason as the one mentioned in the previous paragraph, we use a single-stage formulation in which agents compete at the same time in the electricity market and in markets for additional instruments.

For zonal pricing models, we consider both models with cost-based re-dispatch and models with market-based re-dispatch. In the case of cost-based re-dispatch, we assume that the variable cost of production is perfectly known to the TSO when compensating the producers. Moreover, all resources are available for re-dispatch, and there is no irrevocable decision based on the outcome of the zonal pricing auction. In particular, there is no unit commitment decision made based on the zonal auction, and we assume that the zonal net positions cleared in the zonal auction are *not* firm and can be freely modified during the re-dispatch phase.

We now describe the models for the different market designs that we consider based on this common modeling framework and set of assumptions.

4.3.2 Nodal and zonal pricing benchmarks

Nodal pricing

As we describe in the previous chapter (in section 3.3), the nodal capacity expansion problem can be written as follows:

$$
\begin{aligned}
\min_{x_{in}, y_{int}} \ & \sum_{int} MC_i y_{int} + \sum_{in} IC_{in} x_{in} \\
\text{s.t. } & X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\
& r_{nt} - \sum_{in} y_{int} + D_{nt} = 0, n \in N, t \in T \\
& r_{:t} \in \mathcal{R}, t \in T
\end{aligned}
\tag{4.1}
$$

Note that we have omitted, here, the possibility for load shedding in Problem (4.1) to lighten the formulations, but the analysis would be similar if it would be accounted for. Problem (4.1) describes both the results of a centralized capacity expansion problem and of its decentralized counterpart, where the three types of agents (TSO, producers and auctioneer) compete on an electricity market with nodal pricing.

Within our set of assumptions, the nodal pricing market design achieves the best efficiency possible while managing congestion in a market-based way, in the sense that it achieves the lowest possible total operating and investment cost in the long run. Therefore, the results of this nodal pricing problem are used as a benchmark throughout the chapter. In particular, we say that a market design is efficient if it achieves the same total cost as the nodal pricing design.

Zonal pricing

The zonal pricing market design is a building block for the different designs that we compare in the chapter, which correspond to market-based instruments added on top of the pure zonal pricing design. For this reason, pure zonal pricing is also an important benchmark in order to evaluate the impact of each instrument on efficiency. The formulation of the equilibrium in pure zonal pricing is a direct extension of its nodal pricing version, as we discuss in chapter 3, and can be obtained by solving the following optimization problem:

$$\begin{aligned} \min_{x_{in}, y_{int}} \; & \sum_{int} MC_i y_{int} + \sum_{in} IC_{in} x_{in} \\ \text{s.t.} \; & X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\ & p_{zt} - \sum_{i, n \in N(z)} y_{int} + D_{zt} = 0, z \in Z, t \in T \\ & p_{:t} \in \mathcal{P}, t \in T \end{aligned} \tag{4.2}$$

Note that we have made an additional modification to the model of the previous chapter (model (3.6)): we assume, here, that the investment cost depends on the location and, as a consequence, that the investment decision has nodal granularity. We make this assumption in order to obtain more general results for the instruments that can restore the efficiency of zonal pricing.

We keep the definition of $\mathcal{P}$ general at this stage, and only assume that it includes the power balance constraint ($\sum_{zt} p_{zt} = 0 \quad \forall t \in T$) and that it depends solely on grid quantities. This is a necessary condition for the equivalence between centralized and decentralized formulations to hold. The reader is referred to section 4.3.6 for a discussion of the impact of different formulations of $\mathcal{P}$ on the analysis of this chapter.

Cost-based re-dispatch Problem (4.2) correctly represents the investment and production decisions in the zonal auction that would result from zonal pricing followed by cost-based re-dispatch within our set of assumptions. Indeed, as cost-based re-dispatch does not modify the revenues of the producers, their investment decision is based solely on the results of the zonal auction itself. Problem (4.2), however, is not sufficient to understand the efficiency of zonal pricing in the long run as the obtained dispatch y_{int} does not respect the real nodal transmission constraints of the grid. The TSO must resort to re-dispatch and its cost must be accounted for when computing efficiency. As stated in section 4.3.1, we assume in this chapter that the TSO has full flexibility when performing re-dispatch. The goal of the TSO is to minimize re-dispatch cost, which is equivalent to solving a nodal economic dispatch problem at each period given capacity investments $\bar{x}_{in}$:

$$\begin{aligned} \min_{y_{int} \geq 0} \; & \sum_{in} MC_i y_{int} \\ \text{s.t. } & X_{in} + \bar{x}_{in} - y_{int} \geq 0, i \in I, n \in N \\ & r_n - \sum_i y_{int} + D_n = 0, n \in N \\ & r_{:t} \in \mathcal{R} \end{aligned} \tag{4.3}$$

TSO coordination for re-dispatch As we mention in section 4.3.1, this model of re-dispatch assumes that there is no constraint on the final net positions obtained after re-dispatch, in line with the assumption made in the analysis of the short-term efficiency of zonal pricing in chapter 2 (see, for instance, the discussion in section 2.5.5). This implies that the TSOs coordinate perfectly including in resorting to cross-border re-dispatch. This assumption is important, as it implies that zonal pricing is efficient in the short run. Indeed, in this case, the re-dispatch problem of the TSO is equivalent to the nodal economic dispatch problem, and the final dispatch obtained is thus the same as the dispatch obtained under nodal pricing. The two designs would thus differ only by the allocation of revenues between the different agents. For this reason, as we try to understand the designs that can restore the efficiency of zonal pricing, this assumption can also be viewed in this work as a necessary condition without which the recovery of efficiency is not guaranteed to take place in zonal pricing.

Although this condition is not fully respected in the current European practices, there is ongoing effort towards improved cross-border coordination. Indeed, the Guideline on Capacity Allocation and Congestion Management (CACM) of the European Commission states that TSOs should develop a common methodology for coordinated re-dispatch and countertrading ([Eur15], Art. 35, §1). Moreover, it is now stated explicitly that cross-border re-dispatch must be considered for the calculation of available cross-border capacity in the recent recast of the electricity regulation ([Eur19], Art.16, §4). This is the reason why we do not penalize the deviation from day-ahead net positions in the re-dispatch problem[2].

[2]Recall that the sensitivity of the short-term market results with respect to this assumption is analyzed in section 2.5.5

4.3.3 Locational capacity markets

In this section, we investigate the potential of locational capacity markets for improving the efficiency of investment under zonal pricing. In particular, we would like to understand the conditions under which such an instrument could lead to an equivalent efficiency between nodal and zonal pricing. We assume that the producers and the TSO interact in a capacity market in addition to the electricity market. The existing initiative in Europe that is the closest to the situation that we model here is the case of France, with the spatially constrained tender on top of the capacity market that was organized in Brittany, as we discuss in further detail in section 4.2.3.

We assume that the TSO has a certain inflexible demand of capacity in the capacity market that is based on the solution of a prospective resource adequacy study. This corresponds to current practices, such as ENTSO-E's Mid-term Adequacy Forecast [ENT20].

Resource adequacy. We assume that the resource adequacy problem solved by the TSO is a nodal capacity expansion problem with perfect information on the state of the system, including on marginal and investment costs of the producers. This is a strong assumption that can also be viewed in the context of this work as a condition without which efficiency of nodal pricing is not guaranteed be recovered. The TSO therefore solves a problem that is exactly equivalent to problem (4.1) prior to its participation to the capacity market. We denote by $\bar{x}_{in}$ the value of the optimal investment in the solution of problem (4.1).

Two-sided capacity market with locational and technological differentiation

We start by considering the case of a capacity market that has full flexibility: it is two-sided and depends on the location n and technology i of the investment. We describe successively the profit-maximizing problem of each agent in the electricity and capacity markets in this case.

Producers.

$$\begin{aligned} \max_{y_{int}, x_{in}} & \sum_t \left(\rho_{Z(n)t} y_{int} - MC_i y_{int}\right) - IC_{in} x_{in} - \pi_{in} x_{in} \\ \text{s.t. } & X_{in} + x_{in} - y_{int} \geq 0 \\ & x_{in}, y_{int} \geq 0 \end{aligned} \tag{4.4}$$

where π_{in} is the capacity price. If it is positive (negative), it corresponds to an additional cost (revenue) to the producers.

TSO. The TSO maximizes its profit from the transmission of electricity, given the zonal prices, by controlling the injection and withdrawal of power in every zone while ensuring that the resulting zonal net positions belong to $\mathcal{P}$, the set of feasible zonal net positions:

$$\begin{aligned} \max_{p_{zt}} \; & -\sum_{zt} p_{zt}\rho_{zt} \\ \text{s.t.} \; & p_{:t} \in \mathcal{P} \end{aligned} \tag{4.5}$$

Auctioneer of the electricity market. The auctioneer determines zonal prices while ensuring that the market clears in every zone of the network. This function can be represented as a profit-maximizing problem for zone z and period t as:

$$\max_{\rho_{zt}} \; \rho_{zt}(p_{zt} + D_{zt} - \sum_{i,n \in N(z)} y_{int}) \tag{4.6}$$

Auctioneer of the capacity market. In the capacity market, the auctioneer determines the price that leads to a matching between the capacity invested by the producers and the inflexible demand of the TSO:

$$\max_{\pi_{in}} \; \pi_{in}(\bar{x}_{in} - x_{in}) \tag{4.7}$$

Note that this way of modeling capacity markets is similar to the one proposed in [GOS13]. However, compared to the formulation of [GOS13], we add technological and locational differentiation in the capacity market.

Equivalent optimization problem. It can be easily checked (by comparing KKT conditions) that, in this case, the Nash equilibrium can be obtained by solving the following equivalent optimization problem:

$$\begin{aligned} \min_{x_{in}, y_{int}} \; & \sum_{int} MC_i y_{int} + \sum_{in} IC_{in} x_{in} \\ \text{s.t.} \; & X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\ & p_{zt} - \sum_{i,n \in N(z)} y_{int} + D_{zt} = 0, z \in Z, t \in T \\ & p_{:t} \in \mathcal{P}, t \in T \\ & \bar{x}_{in} - x_{in} = 0, i \in I, n \in N \end{aligned} \tag{4.8}$$

We are interested in the conditions under which Problem (4.8), which corresponds to the equilibrium in zonal pricing with a two-sided capacity market, is efficient in the long run. As stated in section 4.3.2, we already know that zonal pricing followed by cost-based re-dispatch is efficient in the short run under our assumptions. Moreover, any solution to Problem (4.8), if it exists, leads to the same investment decisions as the nodal pricing benchmark. Therefore, any equilibrium in zonal pricing with a two-sided capacity market will be efficient if it exists, and it only remains to determine the conditions on $\mathcal{P}$ under which existence can be guaranteed. In the following definition, we propose a condition, that we call *nodal consistency*, which is both natural to require for a well-defined set of feasible net positions and sufficient for the existence.

Definition 4.1. *The feasible set of zonal net positions $\mathcal{P}$ is said to be* ***nodal consistent*** *if there exists a solution to the nodal capacity expansion problem, with r^* the vector of values of the nodal net injections in that solution, such that*

$$p^*_{:t} \in \mathcal{P} \quad \forall t \in T$$

with

$$p^*_{zt} = \sum_{n \in N(z)} r^*_{nt} \quad \forall z \in Z, t \in T$$

In words, $\mathcal{P}$ is nodal consistent if it contains, at each time period, a vector of zonal net positions that aggregates a vector of nodal net injections that solves the nodal capacity expansion problem.

Proposition 4.1. *If $\mathcal{P}$ is nodal consistent, then any equilibrium in zonal pricing with a two-sided capacity market is efficient.*

The proof of Proposition 4.1, as well as the proofs of all propositions of this chapter, are relegated to Appendix 4.A of this chapter.

One-sided capacity markets

In the previous paragraphs we have established that two-sided capacity markets can recover the efficiency of nodal pricing. One is thus naturally motivated to pose the question whether it is possible to obtain an equivalence between nodal and zonal investment with a one-sided capacity market.

Recovering the optimal spatial configuration of investment requires, in a nodal pricing setting, to allow the market to produce electricity prices that may penalize or subsidize investment of specific technologies in given locations. The zonal price distorts this by imposing an average zonal price in all nodes of a given zone. Thus, recovering the behavior of optimal nodal prices requires the zonal market to be able to lift or depress the remuneration of specific technologies in specific locations through locational capacity charges that are sometimes positive, and sometimes negative. If we impose that these charges be only positive or only negative, then a non-trivial question emerges of whether one can still recover optimal investment. The answer is not a priori negative, because what may ultimately drive investments is *differentials* between price signals, but this needs to be examined more carefully, which is the exact task that we undertake in this section, first in the case of a negative capacity price and then, in the positive case.

Negative capacity price. This is the situation of a capacity market where investment in a certain technology and in a certain node is subsidized, but no investment is penalized. This corresponds more closely to current practices regarding capacity remuneration mechanisms in Europe, for which the goal is indeed to remunerate investment, not penalize it (see [SM21] for a discussion of the different mechanisms curretly implemented in Europe). In this case, the optimization problem of the auctioneer of the capacity market is modified in order to add a non-positivity constraint on the capacity price:

$$\max_{\pi_{in} \leq 0} \ \pi_{in}(\bar{x}_{in} - x_{in}) \tag{4.9}$$

Duality implies that, in the corresponding optimization problem, the capacity invested can only be forced to be greater than or equal to the nodal investment:

$$\begin{aligned} \min_{x_{in}, y_{int}} & \sum_{int} MC_i y_{int} + \sum_{in} IC_{in} x_{in} \\ \text{s.t. } & X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\ & p_{zt} - \sum_{i,n \in N(z)} y_{int} + D_{zt} = 0, z \in Z, t \in T \\ & p_{:t} \in \mathcal{P}, t \in T \\ & x_{in} - \bar{x}_{in} \geq 0, i \in I, n \in N \end{aligned} \tag{4.10}$$

It turns out that, in this case, efficiency can also be guaranteed under certain conditions, as we show with Proposition 4.2. In order to determine a sufficient condition for which efficiency holds in the one-sided case, we use the concept of the nodal capacity factor.

Definition 4.2. *The **nodal capacity factor** of a technology i in a node n, that we denote by γ_{in}, is the proportion of time for which the technology produces a positive quantity in a solution of the nodal capacity expansion problem, i.e.*

$$\gamma_{in} = \frac{\sum_{t \in T}(\bar{y}_{int} > 0)}{|T|} \tag{4.11}$$

where $\bar{y}_{int}$ is the production in an optimal solution of the nodal capacity expansion problem.

Proposition 4.2. *If $\mathcal{P}$ is nodal consistent and if $\forall i, j \in I, n \in N, i \neq j$,*

$$\frac{IC_{in}}{|T|} + MC_i < MC_j \tag{4.12}$$

implies that

$$\frac{IC_{in}}{|T|} - \frac{IC_{jn}}{|T|} < \gamma_{jn}(MC_j - MC_i) \tag{4.13}$$

with $\gamma_{jn} > 0$ in at least one solution to the nodal capacity expansion problem, then any equilibrium in zonal pricing and a one-sided capacity market with a nonpositive capacity price is efficient.

In order to understand better this condition, assume a very simple two-node one-zone network with a single transmission line with zero capacity. The demand in the first node (node 1) is equal to D during the first half of the capacity expansion horizon and 0 in the second half. The opposite holds for node 2, with 0 demand in the first half and D in the second. There are two candidate technologies, A and B, with respective costs IC_A, MC_A and IC_B, MC_B. Assume that

$$\frac{IC_A}{|T|} + \frac{1}{2}MC_A < \frac{IC_B}{|T|} + \frac{1}{2}MC_B \tag{4.14}$$

This implies that the optimal nodal solution is to invest two units of technology A in the two nodes, with capacities D. Is it possible that the zonal solution with a lower bound $x_A \geq D$ and $x_B \geq 0$ in the two nodes can deviate from the nodal optimal investment? This would be the case if it is beneficial for the system to invest an additional capacity D of technology B in the system so as to cover the zonal demand, which is constant at a level D during the whole horizon. This holds when

$$\frac{IC_B}{|T|} + MC_B < MC_A \tag{4.15}$$

Conditions (4.14) and (4.15) are not mutually exclusive, and it is possible to find values for the costs such that both conditions are respected (which would mean that, although the optimal investment solution is to build technology A only, the zonal solution would build technology B and thus be inefficient). Take, for instance, $IC_A = 0, MC_A = 10, IC_B/|T| = 3, MC_B = 3$. But this situation is very specific: it takes place due to the important congestion and the specific choice of costs. The condition that we impose in Proposition 4.2 prevents such cases from occurring.

Intuitively, one understands that the problem here lies in the fact that the marginal cost of the existing capacity of technology A is too high. Therefore, no matter how the investment in technology A is subsidized (through a negative capacity price), it will always be better to invest also in technology B in order to cover the zonal demand, which is inefficient.

Positive capacity price. In the same vein, under this framework one can easily model the situation in which a non-negative capacity price is imposed. This corresponds to a market design with locationally and technologically differentiated connection charges. In this case, the equivalent optimization problem to the Nash equilibrium becomes as follows:

$$\begin{aligned}
\min_{x_{in}, y_{int}} \; & \sum_{int} MC_i y_{int} + \sum_{in} IC_{in} x_{in} \\
\text{s.t.} \; & X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\
& p_{zt} - \sum_{i, n \in N(z)} y_{int} + D_{zt} = 0, z \in Z, t \in T \\
& p_{:t} \in \mathcal{P} \\
& \bar{x}_{in} - x_{in} \geq 0, i \in I, n \in N
\end{aligned} \tag{4.16}$$

The capacity of the nodal benchmark now implies an upper bound on the investment. In this case, however, the efficiency is in general not recovered. Here, the situation is quite different than with a negative price: one cannot find a simple condition for which efficiency would be obtained. Indeed, let us consider again our simple two-node one-zone example. Assume that there is only a single technology, such that the condition for efficiency of the negative price is trivially respected. In this case, the nodal solution is to invest D capacity of the only technology available in both nodes. The upper bounds, however, are not constraining in the zonal solution which will be optimal with a single unit of capacity D invested.

Technology-independent capacity markets

The previous results have all been obtained under the assumption that the capacity markets are differentiated between all technologies. Although this is a necessary condition for a capacity-based market to be efficient [Eic21], it is usually not the case in existing market-based price signals [EKH20]. This assumption can easily be lifted within our modeling framework. In this case, the dual variables of the capacity market clearing constraints depend only on the node and so do the constraints. The equivalent optimization problem remains the same, except for the capacity market clearing constraints which now read as:

$$\sum_{i \in I} x_{in} - \sum_{i \in I} \bar{x}_{in} \geq 0, n \in N \tag{4.17}$$

The loss of efficiency associated to lack of technological differentiation can easily be evaluated by solving this new optimization problem. We confirm, indeed, experimentally that this design fails to recover nodal efficiency, as discussed in detail in the case study of section 4.4.

4.3.4 Locational energy markets

The second class of instruments that we consider in this chapter are locational energy markets. These additional markets imply that, in addition to the electricity price, a locationally differentiated energy price is added, which can be either positive or negative. The models of this section can be seen as closely related to the exiting market design in Norway and Sweden, where a locationally differentiated multiplier is added on top of the zonal electricity price, as we discuss in further detail in section 4.2.3.

Two-sided markets

Let us start with the case of a two-sided energy market by describing successively the profit-maximizing problems of each type of agent.

Producers. The objective of the producers is modified by the addition of the locational energy price ν_{nt} to their marginal cost.

$$\begin{aligned} \max_{x_{in}, y_{int}} & \sum_t \Big((\rho_{Z(n)t} - \nu_{nt} - MC_i) y_{int} \Big) - IC_{in} x_{in} \\ \text{s.t. } & X_{in} + x_{in} - y_{int} \geq 0 \\ & x_{in} \geq 0, y_{int} \geq 0 \end{aligned} \tag{4.18}$$

In this setting, ν_{nt} plays the role of a nodal adder to the zonal electricity price, in an analogous way as the multiplier on the zonal price in the Norwegian and Swedish markets.

TSO. The TSO collects the zonal congestion rent as well as the revenues (or costs) from its participation to the energy market, so that the total production at each node n and time t is feasible with the nodal constraints of the grid.

$$\begin{aligned} \max_{p_{zt}, \tilde{y}_{nt}} & - \sum_{zt} p_{zt} \rho_{zt} - \sum_{nt} r_{nt} \nu_{nt} \\ \text{s.t. } & p_{:t} \in \mathcal{P} \\ & r_{:t} \in \mathcal{R} \end{aligned} \tag{4.19}$$

Note that the part of the TSO problem that relates to the zonal electricity market (and involves variable p) and to the nodal energy market (with variable r) are independent. Consequently, they could also be formulated in two separate optimization problems. Here, we chose to represent them in the single optimization problem (4.19) for conciseness.

Auctioneer of the electricity market. The problem of the auctioneer of the zonal market does not change (see Problem (4.6)).

Auctioneer of the energy market. In the locational energy market, the auctioneer determines the energy prices so that the total injection determined at node n and time t by the TSO (r_{nt}) matches the demand D_{nt} and the total power produced at node n and time t ($\sum_i y_{int}$):

$$\max_{\nu_{nt}} \ \nu_{nt}\Big(r_{nt} - \sum_i y_{int} + D_{nt}\Big) \tag{4.20}$$

Note that, ultimately, these energy prices are determined so as to restore network feasibility, represented by $\mathcal{R}$ in the TSO problem.

Equivalent optimization problem. Using the same reasoning as in the previous section, a Nash equilibrium can be obtained by solving the following equivalent optimization problem:

$$\begin{aligned}
\min_{x_{in}, y_{int}} \ & \sum_{int} MC_i y_{int} + \sum_{in} IC_{in} x_{in} \\
\text{s.t. } & X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\
& - p_{zt} + \sum_{i \in I, n \in N(z)} y_{int} - \sum_{n \in N(z)} D_{nt} = 0, z \in Z, t \in T \\
& p_{:t} \in \mathcal{P}, t \in T \\
& - r_{nt} + \sum_i y_{int} - D_{nt} = 0, n \in N, t \in T \\
& r_{:t} \in \mathcal{R}, t \in T
\end{aligned} \tag{4.21}$$

Using the definition of nodal consistency that we introduced in section 4.3.3 (Definition 4.1), it is quite straightforward to observe that (4.21) is equivalent to the nodal problem and therefore leads to the same efficiency as the nodal capacity expansion problem.

Proposition 4.3. *If $\mathcal{P}$ is nodal consistent, then zonal pricing with an energy-based tariff is efficient.*

This result is identical to that obtained in the case of capacity-based tariffs (Proposition 4.1). In both cases we are implicitly given the freedom to modify energy remuneration both upward and downward, such that the scarcity rents exactly match the investment costs of the optimal spatial technology configuration; in the capacity-based case this is implicit, because it is a direct payment on the invested capacity, whereas in the energy-based case it is clear that it is a direct remuneration of energy.

One-sided markets

As in the case of capacity-based markets, one may wonder whether the equivalence holds if we restrict ourselves to one-sided markets, i.e. when the energy price is restricted to be non-negative or non-positive. If we consider a non-negative price, which is probably the most realistic case in practice, as it could be assimilated to an energy-based tariff, using duality we can show that the equivalent optimization problem is modified by transforming the market clearing equality constraint to an inequality:

$$-r_{nt} + \sum_i y_{int} - D_{nt} \geq 0 \tag{4.22}$$

One can observe that this does not impact the equivalence with the nodal benchmark, as this inequality is always tight in any solution of Problem (4.21). Indeed, the following equivalences hold successively:

$$\begin{aligned} &\sum_{zt} p_{zt} = 0 \\ \Leftrightarrow &\sum_{zt} \Big(\sum_{i,n \in N(z)} y_{int} - \sum_{n \in N(z)} D_{nt} \Big) = 0 \\ \Leftrightarrow &\sum_{nt} \Big(\sum_i y_{int} - D_{nt} \Big) = 0 \end{aligned}$$

As $\sum_{nt} r_{nt} = 0$, we deduce that

$$\sum_{nt} \Big(- r_{nt} + \sum_i y_{int} - D_{nt} \Big) = 0$$

which, using the fact that each $-r_{nt} + \sum_i y_{int} - D_{nt}$ must be non-negative, leads to

$$-r_{nt} + \sum_i y_{int} - D_{nt} = 0, \quad \forall n \in N, t \in T$$

The reasoning holds also in the case of a non-positive price.

4.3.5 Market-based re-dispatch

The last class of models that we consider in this chapter are zonal pricing models with market-based re-dispatch. Unlike in cost-based re-dispatch that works with mandatory participation and does not influence the payoff of the agents, in market-based re-dispatch, the re-dispatch step is organized as a market with voluntary participation in which participants can bid freely and make profit. As the re-dispatch market is nodal, it is a natural candidate for restoring locational signals in zonal markets. Our goal is to understand the conditions under which zonal pricing followed by market-based re-dispatch can recover the efficiency of nodal pricing. We formulate the long-run equilibrium in the same unifying modeling framework as the one that we developed for zonal pricing with cost-based re-dispatch in the previous sections. We therefore use the same assumptions as those introduced in section 4.3.1. In addition, we assume as in [HS20] that the re-dispatch market is organized with marginal pricing as opposed to pay-as-bid pricing, the latter being the rule in most existing re-dispatch markets. The reason is that, under the assumption of perfect competition, the outcome of the two types of auctions is identical [HL15]. We also follow the assumption of [HS20] by preventing pure financial arbitrage. Using the terminology of [HS20], our model is based on *asset-backed arbitrage* which implies that participants can only bid quantities that they would be able to physically produce in both the zonal and the re-dispatch markets.

The models that we present in this section are closely related to the situation in the Netherlands, where re-dispatch is partly market-based. We refer the reader to section 4.2.3 for more details.

We now proceed to the sequential description of the profit-maximizing problems of each agent under our unifying framework.

Producers. The producers have the opportunity to participate in two different markets: the zonal market and the re-dispatch market. We denote by y_{int} the quantity cleared in the zonal market and $\tilde{y}_{int}$ the re-dispatch quantity cleared in the re-dispatch market, which is positive (negative) in case of upward (downward) re-dispatch.

$$\begin{aligned} \max_{y_{int}, \tilde{y}_{int}, x_{in}} \quad & \sum_t \Big(\rho_{Z(n)t} y_{int} + \tilde{\rho}_{nt} \tilde{y}_{int} - MC_i(y_{int} + \tilde{y}_{int}) \Big) - IC_{in} x_{in} \\ \text{s.t. } (\mu_{int}) : & X_{in} + x_{in} - y_{int} \geq 0 \\ (\tilde{\mu}_{int}) : & X_{in} + x_{in} - y_{int} - \tilde{y}_{int} \geq 0 \\ (\delta_{int}) : & y_{int} + \tilde{y}_{int} \geq 0 \\ & x_{in}, y_{int} \geq 0 \end{aligned} \tag{4.23}$$

where ρ_{zt} is the zonal price and $\tilde{\rho}_{nt}$ is the nodal re-dispatch price.

Problem (4.23) represents a situation where the producers can take advantage of arbitrage between the two markets. Indeed, observe that if $\rho_{Z(n)t} > \tilde{\rho}_{nt}$, then the producers have the possibility to extract a profit in the market without actually producing, by setting $y_{int} = -\tilde{y}_{int} = X_{in} + x_{in}$. This corresponds to a situation where the producers would be cleared a certain quantity in the zonal market that would lead to the violation of certain nodal transmission constraints. The producers would have to be re-dispatched down by the same quantity in the re-dispatch market in order to recover a feasible dispatch and would keep a revenue equal to the generating capacity multiplied by the difference between the two prices. This arbitrage is what is often called *inc-dec gaming* in the literature and in practice.

Mathematically, the marginal revenue that the producers can extract through arbitrage is quantified by the dual variable of the first constraint of Problem (4.23), that we denote by μ_{int}. In the sequel, we refer to the total revenue obtained from inc-dec gaming on the whole horizon, i.e. $\sum_{t \in T} \mu_{int}$, as the *arbitrage rent.*

We split the description of the TSO problems in the zonal and re-dispatch markets in the interest of clarity. As the two problems are completely independent, this can be done without loss of generality.

TSO in the zonal market. This problem is the same as in the previous sections (see Problem 4.5).

TSO in RDM. In the RDM, the TSO has to buy re-dispatch resources ($\tilde{r}_{nt}$, negative or positive) in order to recover a nodal dispatch (r_{nt}) that is feasible for the constraints of the DC approximation of the power flow equations.

$$\begin{aligned} \max_{r,\tilde{r}} & - \sum_n \tilde{r}_{nt}\tilde{\rho}_{nt} \\ \text{s.t. } & r_n - \sum_i y_{int} + D_{nt} - \tilde{r}_{nt} = 0 \\ & r_{:t} \in \mathcal{R} \end{aligned} \tag{4.24}$$

Auctioneer in the zonal market. This problem is the same as in the previous sections (see Problem 4.6).

Auctioneer in the RDM.

$$\max_{\tilde{\rho}_{nt}} \ \tilde{\rho}_{nt}(\tilde{r}_{nt} - \sum_i \tilde{y}_{int}) \tag{4.25}$$

It is important to note that the right solution concept for this game is that of a generalized Nash equilibrium (GNE). Indeed, the feasible set of the TSO in the RDM, as described in Problem (4.24), is based on the nodal net injections in the physical dispatch, that depend on the production (y_{int}) in the zonal market. This has consequences in terms of the properties of the equilibrium. On the one hand, the equilibrium is not equivalent to a single optimization problem. On the other hand, the question of existence and uniqueness of equilibria is not as straightforward as in the case of Nash equilibria that we encountered in the other designs considered in this chapter. Although the equilibrium is not equivalent to a single optimization problem, it can still be formulated as a single problem in the form of an MLCP.

Equivalent MLCP. The equivalent MLCP can be obtained by aggregating the KKT optimality conditions (which are necessary and sufficient for linear programs) of the profit-maximizing problem of every agent. Using Greek letters for denoting dual variables for each constraint, we obtain the following MLCP:

$$0 \leq x_{in} \perp IC_{in} - \sum_{t \in T} \mu_{int} - \sum_{t \in T} \tilde{\mu}_{int} \geq 0 \tag{4.26a}$$

$$0 \leq y_{int} \perp MC_i + \mu_{int} + \tilde{\mu}_{int} - \rho_{Z(n)t} - \delta_{int} \geq 0 \tag{4.26b}$$

$$\tilde{y}_{int} \text{ free } \perp MC_i + \tilde{\mu}_{int} - \tilde{\rho}_{nt} - \delta_{int} = 0 \tag{4.26c}$$

$$0 \leq \mu_{int} \perp X_{in} + x_{in} - y_{int} \geq 0 \tag{4.26d}$$

$$0 \leq \tilde{\mu}_{int} \perp X_{in} + x_{in} - y_{int} - \tilde{y}_{int} \geq 0 \tag{4.26e}$$

$$0 \leq \delta_{int} \perp y_{int} + \tilde{y}_{int} \geq 0 \tag{4.26f}$$

$$p_{zt} \text{ free } \perp \rho_{zt} + \sum_m V_{mz}\gamma_{mt} = 0 \tag{4.26g}$$

$$0 \leq \gamma_{mt} \perp W_m - \sum_z V_{mz} p_{zt} \geq 0 \tag{4.26h}$$

$$\tilde{r}_{nt} \text{ free } \perp \tilde{\rho}_{nt} - \nu_{nt} = 0 \tag{4.26i}$$

$$r_{nt} \text{ free } \perp \nu_{nt} + \sum_m \tilde{V}_{mn}\tilde{\gamma}_{mt} = 0 \tag{4.26j}$$

$$\nu_{nt} \text{ free } \perp -r_{nt} + \sum_i y_{int} - D_{nt} + \tilde{r}_{nt} = 0 \tag{4.26k}$$

$$0 \leq \tilde{\gamma}_{mt} \perp \tilde{W}_m - \sum_n \tilde{V}_{mn} r_{nt} \geq 0 \tag{4.26l}$$

$$\rho_{zt} \text{ free } \perp -p_{zt} + \sum_{i,n \in N(z)} y_{int} - D_{zt} = 0 \tag{4.26m}$$

$$\tilde{\rho}_{nt} \text{ free } \perp -\tilde{r}_{nt} + \sum_{in} \tilde{y}_{int} = 0 \tag{4.26n}$$

where we define $V \in \mathbb{R}^{M \times |Z|}$, $W \in \mathbb{R}^M$ and $M \in \mathbb{N}$ such that

$$p \in \mathcal{P} \Leftrightarrow W_m - \sum_z V_{mz} p_z \geq 0, m \in \{1, ..., M\}$$

and $\tilde{V} \in \mathbb{R}^{\tilde{M} \times |N|}$, $\tilde{W} \in \mathbb{R}^{\tilde{M}}$ and $\tilde{M} \in \mathbb{N}$ such that

$$r \in \mathcal{R} \Leftrightarrow \tilde{W}_m - \sum_n \tilde{V}_{mn} r_n \geq 0, m \in \{1, ..., \tilde{M}\}$$

which are well-defined as both $\mathcal{P}$ and $\mathcal{R}$ are polytopes.

Validation. When the investment decisions are fixed, the MLCP (4.26) models the same equilibrium as the one described in [HS18] and one can check that it reproduces the results of [HS18] when applied on the same instance.

Existence of solutions. It turns out that the existence of solutions is guaranteed for this problem under a few additional light assumptions, as we show in the following proposition.

Proposition 4.4. *If the marginal costs, the investment costs and the demand in all nodes are positive and if the set $\mathcal{P}$ is such that $W_m \geq 0 \;\; \forall m \in M$, then model* (4.26) *has a solution.*

Market-based re-dispatch with locational capacity market

As the long-run equilibrium of zonal pricing followed by market-based re-dispatch is not guaranteed to be efficient, which we also confirm in the experiments of section 4.4, one may wonder whether its efficiency can theoretically be recovered by adding locational capacity markets. This is what we analyze in the present section.

Two-sided markets We start by considering the case of a two-sided capacity market with locational and technological differentiation, as in section 4.3.3 for cost-based re-dispatch. An auctioneer for the capacity market is added, which leads to the following additional complementarity condition:

$$\pi_{in} \text{ free } \perp \bar{x}_{in} - x_{in} = 0 \tag{4.27}$$

The investment condition becomes

$$0 \leq x_{in} \perp IC_{in} - \sum_{t \in T} \mu_{int} - \sum_{t \in T} \tilde{\mu}_{int} + \pi_{in} \geq 0 \tag{4.28}$$

Observe that the question of the efficiency of this design in the long run reduces to the question of the efficiency of pure zonal pricing followed by market-based re-dispatch in the short run. Indeed, equation (4.27) fixes the investment to $\bar{x}_{in}$, the optimal investment of the nodal solution, while equation (4.28) becomes trivial with the addition of the free π_{in} variable. In the next proposition, we establish the short-run efficiency of this design.

Proposition 4.5. *Under the set of assumptions described in section 4.3.1, zonal pricing followed by market-based re-dispatch is efficient in the short run.*

Proposition 4.5 should be seen as being closely related to Proposition 4 of [HL15], which also states that zonal pricing with market-based re-dispatch is efficient in the short run under a slightly different framework and assumptions. In particular, [HL15] assume that the TSO sets the inter-zonal flows to their level in the efficient dispatch. This assumption is related to our assumption of perfect TSO coordination in the re-dispatch stage: both assumptions imply that the efficient dispatch can be recovered during the re-dispatch stage.

Corollary 4.6. *Zonal pricing followed by market-based re-dispatch augmented with a two-sided capacity market with locational and technological differentiation is efficient.*

One-sided capacity market One may wonder whether the same result can be obtained with a one-sided capacity market. As we discuss when we introduce the profit-maximizing problem of the producers (Problem (4.23)), market-based re-dispatch results in general in an excess of profit for producers, as they can extract additional revenues by taking advantage of arbitrage between the zonal market and the re-dispatch market. Therefore, the best candidate for recovering nodal efficiency is a one-sided market with a positive capacity price $\pi_{in} \geq 0$. The capacity market clearing condition now becomes:

$$0 \leq \pi_{in} \perp \bar{x}_{in} - x_{in} \geq 0 \tag{4.29}$$

Proposition 4.7. *There exists a locationally and technologically differentiated capacity price that recovers the efficiency of nodal pricing followed by market-based re-dispatch.*

The proof of Proposition 4.7, which can be found in Appendix 4.A, is a constructive proof that shows that it is theoretically possible to define a price π_{in} that restores the efficiency of zonal pricing with market-based re-dispatch. The idea is to simply set this charge to the arbitrage profit that the producers are expected to be able to make on the market.

Interestingly, this is a symmetrical situation compared to zonal pricing followed by cost-based re-dispatch. With cost-based re-dispatch, there is missing money for investment, and investment must be subsidized if we want to recover the efficiency of nodal pricing. In zonal pricing followed by market-based re-dispatch, on the contrary, there are excess profits to the producers that must be taxed in order to restore efficiency.

One can go even further by noticing that the arbitrage rent $\sum_t \mu_{int}$ does not depend on the technology i and will be the same for each technology at a specific node. Indeed, the arbitrage rent is defined by the following equation:

$$0 \leq y_{int} \perp MC_i + \mu_{int} + \tilde{\mu}_{int} - \rho_{Z(n)t} - \delta_{int} \geq 0 \tag{4.30}$$

From the analysis of the complementarity conditions associated to the re-dispatch, i.e. MLCP (4.43), we get that at equilibrium, the following must hold:

$$\tilde{\mu}_{int} - \delta_{int} = \rho_{nt} - MC_i \tag{4.31}$$

The complementarity condition on the zonal dispatch can thus be rewritten as:

$$0 \leq y_{int} \perp \tilde{\rho}_{nt} + \mu_{int} - \rho_{Z(n)t} \geq 0 \tag{4.32}$$

This equation implies that

$$\mu_{int} = \begin{cases} \rho_{Z(n)t} - \tilde{\rho}_{nt} \text{ if } \rho_{Z(n)t} > \tilde{\rho}_{nt} \\ 0 \text{ otherwise} \end{cases} \tag{4.33}$$

which shows that arbitrage rent does not depend on technology i. Therefore, the result of Proposition 4.7 can be strengthened in the following way:

Proposition 4.8. *There exists a locationally differentiated capacity price that recovers the efficiency of nodal pricing followed by market-based re-dispatch.*

As it stands, the proposition only states that there exists a capacity price that recovers the efficiency of nodal pricing, i.e. there exists an equilibrium that is efficient. This means that the efficiency of the design would be guaranteed to be recovered if an external entity would be able to compute and impose the price to market participants. One may wonder whether the result still holds if the charge is computed in a decentralized way, i.e. in a market.

Mathematically, the question of whether the mechanism remains efficient in this case translates into the question of whether all solutions to the MLCP augmented with the locational capacity market are efficient. In the case of a locational capacity market, the investment conditions of producers become

$$0 \leq x_{in} \perp IC_i - \sum_{t \in T} \mu_{int} - \sum_{t \in T} \tilde{\mu}_{int} + \pi_n \geq 0 \tag{4.34}$$

where π_n is the capacity price. The capacity market clearing condition becomes

$$0 \leq \pi_n \perp \sum_{i \in I} \bar{x}_{in} - \sum_{i \in I} x_{in} \geq 0 \tag{4.35}$$

The result turns out to be affirmative:

Proposition 4.9. *Zonal pricing followed by market-based re-dispatch augmented with a locationally differentiated capacity market recovers the efficiency of nodal pricing.*

4.3.6 Feasible set of zonal net positions

In the previous sections, we kept the discussion abstract of considerations regarding the specific definition of the feasible set of zonal net positions $\mathcal{P}$. We have seen, however, that our results depend on the choice of $\mathcal{P}$. In capacity and energy markets, our results of efficiency hold when $\mathcal{P}$ satisfies nodal consistency. The shape of $\mathcal{P}$ can also influence the magnitude of the prices of the instruments needed to recover efficiency. In this section, we describe some of the possible definitions for $\mathcal{P}$ and discuss their implications in the context of the present study.

Price aggregation

The feasible set of zonal net positions with price aggregation, that we denote by $\mathcal{P}^{\mathrm{PA}}$, was introduced in the previous chapter, section 3.3.2. It can be readily checked that $\mathcal{P}^{\mathrm{PA}}$ is nodal consistent.

Flow-based market coupling

FBMC is the default methodology for market coupling in Europe. As we discuss in chapter 3, the specificity of FBMC is that it depends on the installed capacity x_{in}, which implies some inefficiency in the context of capacity expansion. In the present chapter, it also implies that the definition of nodal consistency must be adapted to this situation.

Definition 4.3. *The feasible set of zonal net positions $\mathcal{P}$ is said to be* ***nodal consistent*** *if there exists a solution to the nodal capacity expansion problem with r^* and x^* respectively the vector of values of the nodal net injections and the investment in that solution, such that*

$$p^*_{:t} \in \mathcal{P}(x^*) \quad \forall t \in T$$

with

$$p^*_{zt} = \sum_{n \in N(z)} r^*_{nt} \quad \forall z \in Z, t \in T$$

It can be shown that our results of efficiency remain true with this new definition and it can be easily checked that $\mathcal{P}^{\text{FBMC-EP}}$,which is defined in equation (1.4), is nodal consistent.

Min-RAM

By definition, a zonal market ignores the flows associated to intra-zonal trade, which implies that re-dispatch is needed in general. As discussed in [Mee20], TSOs tend to limit cross-border trade in order to decrease re-dispatch. As a consequence, the European Commission introduced a minimum requirement of 70% for the capacity that should be made available for cross-border trade by TSOs. This requirement, sometimes referred to as a min-RAM requirement, can easily be modeled within our framework. Indeed, if we start from the polytope $\mathcal{P}^{\text{PA}}$, we observe that the flow variables f_k represent purely flows due to inter-zonal trade. In zonal pricing with the 70% min-RAM, these flows on each line should be allowed to reach 70% of the line capacity. Let us denote by η the min-RAM requirement. The feasible set of zonal net positions under the min-RAM rule can be written as:

$$\begin{aligned}\mathcal{P}^{\text{MR}}_{\eta} = \Big\{ p \in \mathbb{R}^{|Z|} \Big| \; & \exists (f,r) \in \mathbb{R}^{|K|} \times \mathbb{R}^{|N|} : p_z = \sum_{n \in N(z)} r_n \ \ \forall z \in Z, \\ & f_k = \sum_n PTDF_{kn} \cdot r_n \ \ \forall k \in K, \sum_n r_n = 0, \\ & -\eta \cdot TC_k \leq f_k \leq \eta \cdot TC_k \ \ \forall k \in K \Big\}\end{aligned} \tag{4.36}$$

Note that $\mathcal{P}^{\text{MR}}_{\eta}$ is not guaranteed to be nodal consistent, which implies that our results of efficiency in capacity and energy markets do not hold in this case.

4.4 Simulation results

In this section, we present simulation results for a reduced instance of the CWE network area. We use the same dataset as the one used in the previous chapter and that is presented in section 3.5.1.

Most models that we have presented in section 4.3 can be formulated as linear programs (LP) and can thus be readily solved by state-of-the-art LP solvers. The only one that cannot be formulated as an LP is the long-run equilibrium of zonal pricing with market based re-dispatch which, instead, is equivalent to the MLCP (4.26). As mentioned in section 4.3.5, this equilibrium corresponds in fact to a GNE. In this sense, this problem is similar to the decentralized capacity expansion problem with FBMC, described in the previous chapter in section 3.4.3. For this reason, the same splitting-based algorithm can be used for solving MLCP (4.26). The algorithm is described in Appendix B.

For market-based re-dispatch, we modify the MLCP (4.26) in order to make it more realistic by limiting the amount of arbitrage rent that the producers can extract. Indeed, as we discuss in section 4.3.5, even in the absence of market power, producers are incentivized to deviate from bidding at marginal cost in the zonal market when it is followed by market-based re-dispatch. In practice, this deviation could catch the attention of the regulator if it is too important and it is therefore unrealistic to assume that it takes place to its full possible extent. Although our models assume that market participants are price takers and only play on quantities, their bidding behavior can be inferred from the results of MLCP (4.26). Indeed, when the investment is fixed, the solution of the zonal market can be obtained by solving a modified version of the zonal economic dispatch problem in which the marginal costs of the producers are replaced by the re-dispatch price at their node (see Problem (4.47) and the proof of Proposition 4.7 in Appendix 4.A). For the simulation results of the case study that we present in this section, the MLCP (4.26) is thus modified by adding a nodal capacity tax that corresponds to excess profits obtained from arbitrage. Profits are considered excessive when they correspond to an equivalent bidding behavior of 30% more (less) than the marginal cost of the peak-load (base-load) technology in the zonal market.

Unless specified otherwise, we use the price aggregation methodology for zonal pricing, i.e. set $\mathcal{P}^{\mathrm{PA}}$ defined in equation (3.4).

4.4.1 Relative performance of policies

Table 4.3 presents the performance in terms of total operational and investment costs of the different policies. Two policies are able to reproduce the efficiency of the nodal pricing benchmark: the energy signal policy with full temporal granularity and market-based re-dispatch with locational connection charges. In the simulations, we assumed that the capacity-based signals apply only on the new investment, not on existing units, which explains why the two-sided capacity market does not recover the efficiency of nodal pricing.

Regarding these capacity markets, our simulation results confirm the theoretical result of Proposition 4.2, which states that the efficiency of the two-sided capacity market instrument can be obtained with its one-sided version with negative capacity price: the two policies obtain indeed the same efficiency

Policy	Op. cost	Inv. cost [M€/yr]	Total cost
Benchmark			
Nodal	15,810	10,433	26,243
Zonal with PA	16,835	10,909	27,744
Capacity market			
Two-sided (TS)	16,041	10,839	26,880
Negative one-sided	16,041	10,839	26,880
Positive one-sided (POS)	16,899	10,795	27,694
TS no technology differentiation	16,705	10,619	27,324
TS with 70% rule	15,929	11,088	27,016
Energy market			
Two-sided	15,809	10,433	26,242
TS low granularity (÷4)	15,911	10,493	26,404
Market-based re-dispatch			
Market-based re-dispatch (MBR)	15,860	11,646	27,506
MBR with POS capacity market	15,810	10,433	26,243

Table 4.3: Performance comparison of the different policies.

in our simulations. Our results also highlight the importance of differentiating technologies in the capacity market as well as the importance of the design of the zonal transmission constraints for the efficiency of the instruments. The loss of efficiency of the capacity market policy when it is not differentiated by technologies, i.e. the difference between "TS" and "TS no technology differentiation", is evaluated at 2.4% of the total cost. The loss of efficiency associated to the use of the feasible set of zonal net positions based on the 70% rule, that is not nodal-consistent, is estimated at 1.3% (difference between "TS" and "TS with 70% rule").

The energy market must have full temporal granularity in order to recover the efficiency of the nodal benchmark. In practice, this implies that in the case of an hourly day-ahead auction, the energy price that is added on top of the zonal day-ahead electricity price must also be updated on an hourly basis. This renders the policy complicated to implement in practice and one may wonder what happens if the energy market is implemented with a lower temporal granularity. The energy market policy with low granularity that we have simulated has 4 times less temporal granularity than the full energy market. Its loss of efficiency is estimated in our simulations at 161 M€/year, which amounts to 0.6% of the total cost.

We observe that, in the long run, zonal pricing followed by market-based re-dispatch is significantly more costly than the nodal pricing benchmark and only slightly more efficient that the zonal pricing benchmark, even though excessive arbitrage profits are prevented through a modification of the MLCP 4.26, as we discuss in the beginning of section 4.4. We obtain a loss of efficiency compared to nodal pricing of 4.8% of the total cost. This rise in the total cost is due almost exclusively to the investment cost, which is to be expected in regards of the analysis of section 4.3. Indeed, the arbitrage between the zonal electricity market and nodal re-dispatch market allows producers to extract a rent that translates to significantly more investment in the long run, but that does not improve the operational cost. The operational costs of the nodal policy and the MBR policy are indeed comparable.

Although the MBR policy does not seem to be a good candidate to obtain an efficient design, it should be highlighted that this policy is probably the best candidate when appropriate locational instruments are used to steer the right investment on top of market prices. Indeed, in our simplified framework, the efficiency of the nodal benchmark is restored with a locational capacity market, which is probably the easiest instrument to implement among all that we have discussed in this study.

4.4.2 Value of the instruments and capacity mix

We now turn to the comparison of the different policies based on the value of the instruments and the capacity mix.

Figure 4.1 shows the different nature of the instruments. The capacity market is differentiated among technologies, the two-sided energy market has

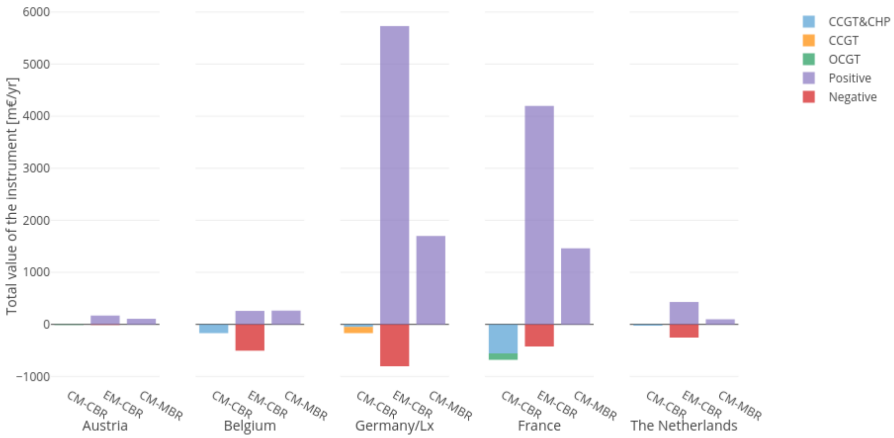
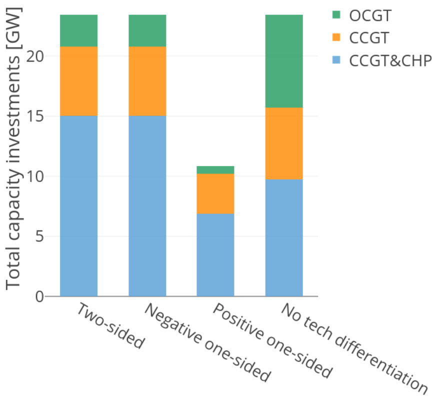


Figure 4.1: Total value of the different types of instrument as a function of the bidding zone. CM-CBR stands for capacity market with cost-based re-dispatch. EM-CBR is the energy market policy with cost-based re-dispatch. CM-MBR is the combination of a capacity market and market-based re-dispatch. Different colors are used to differentiate technologies in the CM-CBR policy. The two different colors for EM-CBR differentiate the positive and negative prices in the two-sided version of the market. The value of an instrument is defined as the total amount of money received or paid by the producers in the corresponding policy, i.e. the price of the instrument multiplied by the total volume involved.

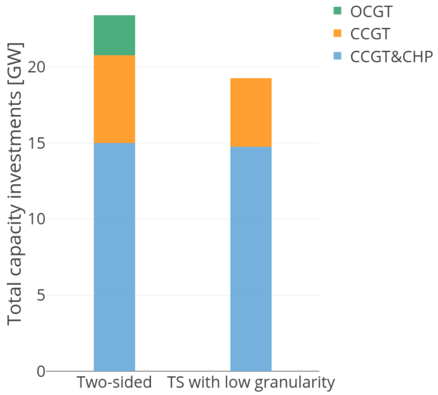
both positive and negative prices and the market-based re-dispatch with one-sided capacity market has only positive prices. We observe that among the two policies that are able to recover the efficiency of nodal pricing, the one with market-based re-dispatch requires a significantly smaller total value for the instrument. This is in part due to the fact that the loss of efficiency in market-based re-dispatch is moderated by the limit on the bidding behavior of market participants. In comparison, the capacity market policy with cost-based re-dispatch, which in this situation does not lead to a full recovery of the nodal efficiency, leads to even lower total value for the instruments. It is interesting to observe that, in general in energy markets, the total value of positive prices is more important than for negative prices. This means that the electricity price must in general be augmented by a positive energy price. However, the opposite is true for Belgium where the zonal electricity price must more frequently be decreased by negative energy prices in order to recover the efficiency. This can be related to an analysis of the Belgian regulator [CRE19] which observes that the Belgian day-ahead price that results from market coupling is distorted. The prices increase in Belgium due to high North-South loop-flows that originate from congestion in Germany [CRE19].

In Figure 4.2, we compare the investment in new generation capacity between the different designs, separated in the three main classes of instruments. Subfigure 4.2a corresponds to zonal pricing with locational capacity markets. As anticipated by the theory (see Proposition 4.2), the two-sided capacity market and one-sided capacity market with negative price lead to the same investment. This is not the case, however, for the one-sided version with positive price. This highlights the fact that zonal pricing with cost-based re-dispatch leads, in general, to a lack of investment that must be corrected with additional revenues to the investors, not additional costs. There is also an inefficiency associated to zonal pricing with locational capacity markets when the capacity prices are not differentiated among technologies, as already observed in [GKL+19]. In this case, as one can observe in Figure 4.2, the policy will tend to favor peak technologies that exhibit low investment costs. Subfigure 4.2b relates to zonal pricing with locational energy markets. Efficiency can be recovered in this case under the condition that the energy price has full temporal granularity. When the temporal granularity is not complete, the situation is somehow reverse to the one observed in capacity markets: peak-load technologies, that exhibit high marginal costs, tend to be penalized. The situation with market-based re-dispatch is analyzed in subfigure 4.2c. Interestingly, MBR leads to a large increase in peak-load technology investments, way above the optimal level, which confirms the intuition of [HS20]. In order to profit from arbitrage rents, large amounts of capacity with limited investment costs are built. The investment, however, will not benefit the system as most of this capacity will not be used for actually producing electricity. The investments in the two other capacities with larger investments costs, in contrast, are close to their optimal level.

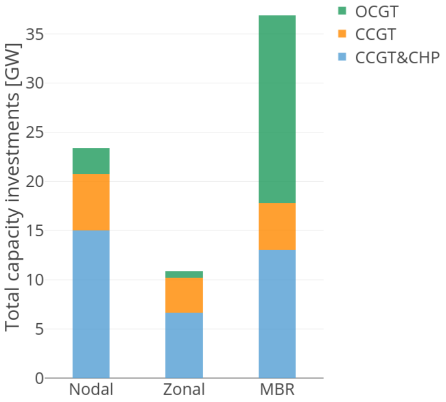
Finally, we are interested in analyzing further the market-based re-dispatch



(a) Capacity-based instruments.



(b) Energy-based instruments.



(c) Market-based re-dispatch and benchmarks.

Figure 4.2: Total capacity investments for the different policies.

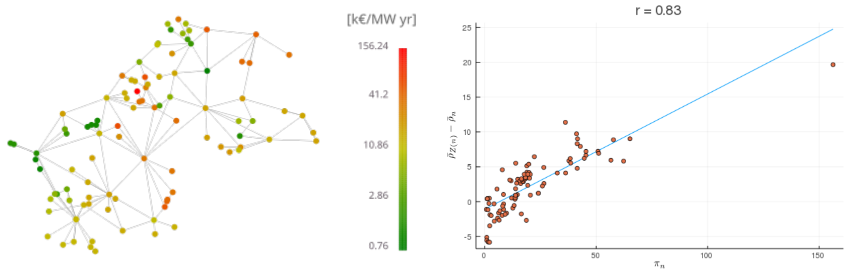


Figure 4.3: Left: value of the optimal connection charge in zonal pricing followed by MBR. Right: difference between the zonal and the nodal price as a function of the value of the optimal connection charge instrument in MBR.

policy. In Figure 4.3, we present in the left panel the value of the optimal capacity price under this policy for each node of the network. In the right panel, we plot the difference between zonal and nodal prices as a function of the optimal capacity price. We observe large differences between the different values of the optimal capacity price. In particular, although the large majority of the charges are below 50k€/MW, one node at the border between Belgium and France receives a value of more than 150k€/MW. This large value can be explained by the fact that it is only connected to nodes in France while it belongs to the Belgian bidding zone. For this reason, it exhibits a large difference between its zonal and nodal price which makes it particularly prone to arbitrage. As displayed on the right panel, there is a high correlation between the difference of nodal and zonal prices in a location and the value of the optimal capacity price needed to restore the efficiency of MBR, which suggests that this difference is the main driver for the inefficiency.

4.5 Conclusion

The question of the best market-based way of allocating transmission capacity remains at the center of intense discussions among European stakeholders. In the academic literature, it is largely treated as a dichotomy in the form of the nodal vs zonal pricing debate. In this chapter, we approach the question from a different perspective and, instead, take zonal pricing as a given for the European electricity market. From this starting point, we investigate the potential of additional market-based instruments for restoring the efficiency of zonal pricing. We consider three main classes of instruments: additional capacity-based and energy-based markets as well as re-dispatch markets. The chapter aims at comparing the efficiency of the three different classes of instruments both theoretically and empirically on the basis of a unifying modeling framework.

We conclude that, theoretically, the efficiency of the nodal design can be recovered in zonal pricing with additional markets, that can be of each of the three classes that we considered. This, however, holds only under strong conditions that are unlikely to be met in practice. For locational capacity markets, efficiency can be recovered in the long run only if the price is differentiated among each type of producing unit. For energy markets, the drawback is that the price should have full temporal and locational granularity which would be difficult to implement. The situation is different for zonal pricing with the market-based re-dispatch policy. Although it is subject to inefficiencies due to arbitrage between the zonal and re-dispatch markets, it can be corrected by means of an additional capacity market that does not need to be differentiated by technology. The main drawback of this policy lies in its complexity as it is comprised of two additional instruments. Additionally, our analysis highlights that these theoretical results are subject to conditions on the cooperation of TSOs and coherence of the zonal capacity calculation methodology, that are currently not entirely fulfilled in practice. For these reasons, it seems unlikely that long-run efficiency could be restored in electricity markets with zonal pricing by means of a practical additional market-based instrument.

4.A Proofs

Proposition 4.1. *If $\mathcal{P}$ is nodal consistent, then any equilibrium in zonal pricing with a two-sided capacity market is efficient.*

Proof of Proposition 4.1. As $\mathcal{P}$ is nodal consistent, existence of a solution to Problem (4.8) follows from the existence of a solution to the nodal capacity expansion problem. The efficiency follows from the short-run efficiency of zonal pricing followed by cost-based re-dispatch. □

Proposition 4.2. *If $\mathcal{P}$ is nodal consistent and if $\forall i, j \in I, n \in N, i \neq j$,*

$$\frac{IC_{in}}{|T|} + MC_i < MC_j \tag{4.12}$$

implies that

$$\frac{IC_{in}}{|T|} - \frac{IC_{jn}}{|T|} < \gamma_{jn}(MC_j - MC_i) \tag{4.13}$$

with $\gamma_{jn} > 0$ in at least one solution to the nodal capacity expansion problem, then any equilibrium in zonal pricing and a one-sided capacity market with a nonpositive capacity price is efficient.

Proof of Proposition 4.2. Let us denote with the symbol $(\hat{\cdot})$ a solution to Problem (4.10) and $(\tilde{\cdot})$ a solution to Problem (4.8). By associating a dual variable to each constraint, the KKT conditions of Problem (4.10) can be written as follows:

$$0 \leq x_{in} \perp IC_{in} - \sum_{t \in T} \mu_{int} - \pi_{in} \geq 0 \tag{4.37a}$$

$$0 \leq y_{int} \perp MC_i + \mu_{int} - \rho_{Z(n)t} \geq 0 \tag{4.37b}$$

$$0 \leq \mu_{int} \perp X_{in} + x_{in} - y_{int} \geq 0 \tag{4.37c}$$

$$p_{zt} \text{ free } \perp \rho_{zt} + \sum_{m} V_{mz}\gamma_{mt} = 0 \tag{4.37d}$$

$$0 \leq \gamma_{mt} \perp W_m - \sum_{z} V_{mz} p_{zt} \geq 0 \tag{4.37e}$$

$$\rho_{zt} \text{ free } \perp -p_{zt} + \sum_{i,n \in N(z)} y_{int} - D_{zt} = 0 \tag{4.37f}$$

$$0 \leq \pi_{in} \perp x_{in} - \bar{x}_{in} \geq 0 \tag{4.37g}$$

where we define $V \in \mathbb{R}^{M \times |Z|}$, $W \in \mathbb{R}^M$ and $M \in \mathbb{N}$ such that

$$p \in \mathcal{P} \Leftrightarrow W_m - \sum_{z} V_{mz} p_z \geq 0, m \in \{1, ..., M\}$$

These conditions are necessary and sufficient for solution $(\hat{\cdot})$.

Let us first show that if there exists $i \in I, n \in N$ such that $\hat{x}_{in} > \bar{x}_{in}$, then there exists $i, j \in I$ and $n \in N$ such that $IC_{in} + MC_i < MC_j$. If $\hat{x}_{in} > \bar{x}_{in}$ then, by condition (4.37g) we have that $\hat{\pi}_{in} = 0$ which implies, by condition (4.37a), that $IC_{in} - \sum_{t \in T} \hat{\mu}_{int} = 0$. Now, we denote by $S \subset T$ the set of time periods t for which $\hat{\rho}_{Z(n)t} > MC_i$. Note that in these time periods, we have by (4.37b) that $\hat{\mu}_{int} > 0$ which implies that:

$$\hat{y}_{int} = X_{in} + \hat{x}_{in} > X_{in} + \tilde{x}_{in} \geq \tilde{y}_{int} \tag{4.38}$$

Then, by conditions (4.37b) and (4.37c), we have that

$$IC_{in} - \sum_{t \in S} \big(\hat{\rho}_{Z(n)t} - MC_i\big) = 0 \tag{4.39}$$

We will show that $\hat{\rho}_{Z(n)t}$ is bounded from above by the marginal cost of at least one technology, for all $t \in S$. To do that, we discuss two possible cases, depending on whether $\hat{p}_{Z(n)t}$ is $\leq$ or $>$ than $\tilde{p}_{Z(n)t}$.

Case 1: $\hat{p}_{Z(n)t} \leq \tilde{p}_{Z(n)t}$ In this case, we have that there exist $j \in I, m \in Z(n)$ such that the following implications hold:

$$y_{jmt} < X_{jm} + \hat{x}_{jm} \Rightarrow \hat{\mu}_{jmt} = 0 \Rightarrow MC_j > \rho_{Z(n)t}$$

Case 2: $\hat{p}_{Z(n)t} > \tilde{p}_{Z(n)t}$ In this case, the technology j that does not reach its maximum capacity might not be in the same zone as node n. The additional thing to notice is that the optimality of the profit maximizing problem of the TSO (i.e. Problem (4.5)) implies that

$$-\sum_{zt} \hat{p}_{zt}\hat{\rho}_{zt} \geq -\sum_{zt} \tilde{p}_{zt}\hat{\rho}_{zt}$$

which, in turn, implies that there exists $z \neq Z(n)$ such that $\tilde{p}_{zt} > \hat{p}_{zt}$ and $\hat{\rho}_{zt} \geq \hat{\rho}_{Z(n)t}$. We conclude by observing that in that zone, by the same argument as in case 1, there exists j such that $MC_j > \rho_{zt}$.

We have shown that, for each $t \in S$, $\hat{\rho}_{Z(n)t}$ can be bounded from above by the marginal cost of one technology. By taking the maximum of the marginal cost of these technologies and denoting it by MC_j, we can transform equation (4.39) into the following inequality:

$$IC_{in} - \sum_{t \in S} \big(MC_j - MC_i\big) < 0$$

In the worst case, $S = T$ and we get

$$IC_{in} - |T|\big(MC_j - MC_i\big) < 0 \tag{4.40}$$

which concludes the first part of the proof.

Now, we proceed by contradiction. Let us assume that there exists $i \in I, n \in N$ such that $\hat{x}_{in} > \bar{x}_{in}$. We have just shown that this implies equation (4.40), which in turn implies, by the main assumption of the proposition, that

$$\frac{IC_{in}}{|T|} + \gamma_{jn} MC_i < \frac{IC_{jn}}{|T|} + \gamma_{jn} MC_j$$

But then, this means that we can find $\epsilon > 0$ such that, in the nodal solution, decreasing x_{jn} by ϵ and increasing x_{in} by ϵ decreases the cost by at least $(IC_{jn} + \gamma_{jn}|T|MC_j - IC_{in} - \gamma_{jn}|T|MC_i)\epsilon$, which is positive. This is a contradiction to the optimality of the nodal investment $\bar{x}_{in}$. □

Proposition 4.3. *If $\mathcal{P}$ is nodal consistent, then zonal pricing with an energy-based tariff is efficient.*

Proof of Proposition 4.3. It suffices to observe that the feasible set of (4.21) is included in the feasible set of (4.1) and that any solution of (4.1) is feasible for (4.21) by definition of nodal consistency. □

Proposition 4.4. *If the marginal costs, the investment costs and the demand in all nodes are positive and if the set $\mathcal{P}$ is such that $W_m \geq 0 \;\; \forall m \in M$, then model* (4.26) *has a solution.*

Proof of Proposition 4.4. We first note that if $MC, IC, D > 0$, then the redispatch prices $\tilde{\rho}_{nt}$ are always nonnegative and the balance constraints can be transformed into inequalities without affecting the solution. Indeed, assume by contradiction that $\exists n, t$ s.t. $\tilde{\rho}_{nt} < 0$. Then, by equation (4.26e), we get that $\delta_{in} > 0 \;\; \forall i \in I$ which implies by (4.26f) that $y_{int} + \tilde{y}_{int} = 0 \;\; \forall i$. Then, combining (4.26k), (4.26n) and the fact that $\sum_n r_{nt} = 0 \;\; \forall t$, we get

$$\sum_{in} \left(y_{int} + \tilde{y}_{int} \right) = D_{nt}$$

which is a contradiction.

Equation (4.26k) can thus be transformed into

$$0 \leq \nu_{nt} \perp -r_{nt} + \sum_i y_{int} - D_{nt} + \tilde{r}_{nt} \geq 0$$

Let us denote by M and q respectively the matrix and the vector of independent terms associated to our MLCP. By Theorem 3.8.6 of [CPS09], if M is copositive and if for all solutions v^* of the homogeneous MLCP, it holds that $q^\top v^* \geq 0$, then there exists a solution to $MLCP(q, M)$. We will use this theorem to prove existence.

Let us first show that M is copositive. To do this, we note that M is the sum of two matrices:

$$M = \tilde{M} + \tilde{N}$$

where $\tilde{M}$ is skew-symmetric (it is the matrix associated to the equivalent MLCP of the centralized problem, i.e. the welfare-maximizing problem) and where $\tilde{N}$ is of the following form (in block formulation):

$$\tag{4.41}$$

$$\tilde{N} = \begin{array}{cc} & \nu_{nt} \\ y_{int} & \begin{pmatrix} 0 & \dots & I & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \end{array} \tag{4.42}$$

Here, I is the rectangular identity matrix, i.e. a matrix with 1 in the entries associated to line y_{int} and column ν_{nt}, and 0 otherwise. This implies that

$$v^\top M v = v^\top \tilde{M} v + v^\top \tilde{N} v = v^\top \tilde{N} v = \sum_{int} y_{int} \nu_{nt}$$

where v is the full vector of variables of the MLCP. This expression is indeed nonnegative if each y_{int} and ν_{nt} are nonnegative.

Now, let $v^* = (x^*_{in}, y^*_{int}, \dots, \rho^*_{zt})$ be a solution to the homogeneous version of the MLCP. We have

$$\begin{aligned} q^\top v^* = & \sum_{in} IC_{in} x^*_{in} + \sum_{int} MC_i y^*_{int} + \sum_{int} X_{in} \mu^*_{int} + \sum_{int} X_{in} \tilde{\mu}^*_{int} + \\ & \sum_{mt} W_m \gamma_{mt} + \sum_{mt} \tilde{W}_m \tilde{\gamma}_{mt} + \\ & \sum_{int} MC_i \tilde{y}^*_{int} - \sum_{nt} D_{nt} (\rho^*_{Z(n)t} + \nu^*_{nt}) \end{aligned}$$

All the terms of the first and second line are non-negative, as they correspond to the product of non-negative quantities. From equation (4.26m) and the fact that $\sum_z p_{zt} = 0$, we deduce that $y^*_{int} = 0$. This implies from equation (4.26f) that $\tilde{y}^*_{int} \geq 0$ and thus that the first term of the third line is also non-negative. From equations (4.26f), (4.26i) and (4.26k), we get that $0 \leq \nu^*_{nt} = \tilde{\rho}^*_{nt} = -\delta^*_{nt} \leq 0$, which implies that these three quantities must be equal to 0. We then deduce from (4.26b) that $\rho^*_{zt} \leq 0$, which, together with $\nu^*_{nt} = 0$ yields the non-negativity of the third term of the second line and concludes the proof. □

Proposition 4.5. *Under the set of assumptions described in section 4.3.1, zonal pricing followed by market-based re-dispatch is efficient in the short run.*

Proof of Proposition 4.5. The short-run problem is the MLCP (4.26) where equation (4.26a) is removed, x_{in} is fixed to $\bar{x}_{in}$, and the time period index t is fixed. Let us consider the MLCP obtained by isolating the equations related to the re-dispatch problem only, i.e. the MLCP consisting of equations

(4.26c), (4.26e), (4.26f), (4.26i), (4.26k), (4.26l) and (4.26n). If we denote by $\bar{y}_{int} = y_{int} + \tilde{y}_{int}$ the physical dispatch of technology i in node n and period t, one can simplify this MLCP by eliminating variables $\delta_{int}, \nu_{nt}, \tilde{r}_{nt}$. We get:

$$\begin{aligned}
&0 \leq \bar{y}_{int} \perp MC_i + \tilde{\mu}_{int} - \tilde{\rho}_{nt} \geq 0\\
&0 \leq \tilde{\mu}_{int} \perp X_{in} + \bar{x}_{in} - \bar{y}_{int} \geq 0\\
&r_{nt} \text{ free } \perp \tilde{\rho}_{nt} + \sum_m \tilde{V}_{mn}\tilde{\gamma}_{mt} = 0\\
&\tilde{\rho}_{nt} \text{ free } \perp r_{nt} - \sum_i \bar{y}_{int} + D_{nt} = 0\\
&0 \leq \tilde{\gamma}_{mt} \perp \tilde{W}_m - \sum_n \tilde{V}_{mn} r_{nt} \geq 0
\end{aligned} \tag{4.43}$$

MLCP (4.43) is exactly the set of KKT conditions of the nodal economic dispatch problem:

$$\begin{aligned}
\min\ & \sum_{in} MC_i \bar{y}_{int}\\
\text{s.t. } & X_{in} - \bar{y}_{int} \geq 0, i \in I, n \in N\\
& r_{nt} - \sum_{in} \bar{y}_{int} + D_{nt} = 0, n \in N\\
& r_{:t} \in \mathcal{R}
\end{aligned} \tag{4.44}$$

which shows that any solution of zonal pricing followed by market-based re-dispatch has the same operating cost as nodal pricing in the short run. □

Proposition 4.7. *There exists a locationally and technologically differentiated capacity price that recovers the efficiency of nodal pricing followed by market-based re-dispatch.*

Proof of Proposition 4.7. With the capacity price added to the market, the KKT condition associated to the investment becomes:

$$0 \leq x_{in} \perp IC_i - \sum_{t \in T} \mu_{int} - \sum_{t \in T} \tilde{\mu}_{int} + \pi_{in} \geq 0 \tag{4.45}$$

We need to show that there exists a solution to the MLCP that consists of equations (4.45), (4.26b) - (4.26n), (4.27) with $\pi^*_{in} \geq 0$ for all $i \in I, n \in N$. This solution can be obtained by solving sequentially two distinct optimization problems, one nodal with investment variables and one zonal with the investment fixed. The nodal problem is the following:

$$\begin{aligned}
\min\ & \sum_{in} IC_{in}x_{in} + \sum_{int} MC_i\bar{y}_{int} \\
\text{s.t. } & X_{in} + x_{in} - \bar{y}_{int} \geq 0, i \in I, n \in N, t \in T\ \ [\tilde{\mu}_{int}] \\
& r_{nt} - \sum_{int} \bar{y}_{int} + D_{nt} = 0, n \in N, t \in T\ \ [\tilde{\rho}_{nt}] \\
& \tilde{W}_m - \sum_n \tilde{V}_{mn} r_{nt} \geq 0, m \in \{1, ..., \tilde{M}\}, t \in T\ \ [\tilde{\gamma}_{mt}] \\
& \bar{y}_{int} \geq 0, i \in I, n \in N, t \in T\ \ [\delta_{int}]
\end{aligned} \tag{4.46}$$

We use the notation $(\cdot)^*$ to denote the optimal value of a primal or dual variable in this problem. The zonal problem is the following:

$$\begin{aligned}
\min\ & \sum_{int} \tilde{\rho}^*_{nt} y_{int} \\
\text{s.t. } & X_{in} + x^*_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T\ \ [\mu_{int}] \\
& p_{zt} - \sum_{i, n \in N(z)} y_{int} + D_{zt} = 0, z \in Z, t \in T\ \ [\rho_{zt}] \\
& W_m - \sum_z V_{mz} p_{zt} \geq 0, m \in \{1, ..., M\}, t \in T\ \ [\gamma_{mt}]
\end{aligned} \tag{4.47}$$

We now let

$$\begin{aligned}
\tilde{y}^*_{int} &= \bar{y}^*_{int} - y^*_{int} \\
\nu^*_{nt} &= -\tilde{\rho}^*_{nt} \\
\tilde{r}^*_{nt} &= r^*_{nt} - \sum y^*_{int} + D_{nt} \\
\pi^*_{in} &= \sum_{t \in T} \mu^*_{int}
\end{aligned}$$

Clearly, π^*_{in} is positive and it can be checked that $\left(x^*_{in}, y^*_{int}, \tilde{y}^*_{int}, \mu^*_{int}, \tilde{\mu}^*_{int}, \delta^*_{int}, p^*_{zt}, \gamma^*_{mt}, \tilde{r}^*_{nt}, r^*_{nt}, \nu^*_{nt}, \tilde{\gamma}^*_{mt}, \rho^*_{zt}, \tilde{\rho}^*_{nt}\right)$ is a solution of the MLCP (4.45), (4.26b) - (4.26n), (4.27). □

Proposition 4.8. *There exists a locationally differentiated capacity price that recovers the efficiency of nodal pricing followed by market-based re-dispatch.*

Proof of Proposition 4.8. This can be proven in the same way as Proposition 4.7 with the additional observation that the arbitrage rent does not depend on the technology. □

Proposition 4.9. *Zonal pricing followed by market-based re-dispatch augmented with a locationally differentiated capacity market recovers the efficiency of nodal pricing.*

Proof of Proposition 4.9. We need to show that all solutions of MLCP (4.34), (4.26b) - (4.26n), (4.35) are as efficient as the nodal solution. Let $(\hat{\cdot})$ be an arbitrary solution of the MLCP. Assume by contradiction that

$$\sum_{in} IC_i \hat{x}_{in} + \sum_{int} MC_i(\hat{y}_{int} + \hat{\tilde{y}}_{int}) > \sum_{in} IC_i \bar{x}_{in} + \sum_{int} MC_i \bar{y}_{int} \tag{4.48}$$

We start by showing that, in this case, $\sum_{i \in I} \hat{x}_{in} = \sum_{i \in I} \bar{x}_{in}$. Let us denote by ν_n the arbitrage rent in this solution, i.e. $\nu_n = \sum_{i \in I} \hat{\mu}_{int}$. Assume by contradiction that there exists a nonempty set of indices $Q \subset N$ such that

$$\sum_{i \in I} \hat{x}_{in} < \sum_{i \in I} \bar{x}_{in} \;\; \forall n \in Q \tag{4.49}$$

and denote by $\bar{Q}$ the one with maximum cardinality among such sets. Now, observe that the corresponding variables of the solution $(\hat{\cdot})$ must also be a solution of the following optimization problem:

$$\begin{aligned} \min \; & \sum_{in} (IC_{in} - \nu_n + \hat{\pi}_n) x_{in} + \sum_{int} MC_i(y_{int} + \tilde{y}_{int}) \\ \text{s.t. } & X_{in} + x_{in} - y_{int} - \tilde{y}_{int} \geq 0, i \in I, n \in N, t \in T \;\; [\tilde{\mu}_{int}] \\ & r_{nt} - \sum_{int} (y_{int} + \tilde{y}_{int}) + D_{nt} = 0, n \in N, t \in T \;\; [\tilde{\rho}_{nt}] \\ & \tilde{W}_m - \sum_n \tilde{V}_{mn} r_{nt}, m \in \tilde{M}, t \in T \;\; [\tilde{\gamma}_{mt}] \\ & y_{int} + \tilde{y}_{int} \geq 0, i \in I, n \in N, t \in T \;\; [\delta_{int}] \end{aligned} \tag{4.50}$$

By definition, the solution $(\bar{\cdot})$ is the solution of the non-perturbed nodal capacity expansion problem, i.e. it is a solution of:

$$\begin{aligned} \min \; & \sum_{in} IC_{in} x_{in} + \sum_{int} MC_i y_{int} \\ \text{s.t. } & X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \;\; [\tilde{\mu}_{int}] \\ & r_{nt} - \sum_{int} y_{int} + D_{nt} = 0, n \in N, t \in T \;\; [\tilde{\rho}_{nt}] \\ & \tilde{W}_m - \sum_n \tilde{V}_{mn} r_{nt}, m \in \tilde{M}, t \in T \;\; [\tilde{\gamma}_{mt}] \\ & y_{int} \geq 0, i \in I, n \in N, t \in T \;\; [\delta_{int}] \end{aligned} \tag{4.51}$$

We deduce, as the feasible sets of both problems are the same, that the following should hold:

$$\begin{aligned} & \sum_{in} (IC_i - \nu_n + \hat{\pi}_n) \bar{x}_{in} + \sum_{int} MC_i \bar{y}_{int} \geq \\ & \sum_{in} (IC_i - \nu_n + \hat{\pi}_n) \hat{x}_{in} + \sum_{int} MC_i(\hat{y}_{int} + \hat{\tilde{y}}_{int}) \end{aligned} \tag{4.52}$$

By reorganising the terms, we get

$$\begin{aligned}&\sum_{in} IC_i\bar{x}_{in} + \sum_{int} MC_i\bar{y}_{int} - \sum_{in} IC_i\hat{x}_{in} - \sum_{int} MC_i(\hat{y}_{int} + \hat{\hat{y}}_{int}) \geq \\ &\sum_{n\in\bar{Q}}(-\nu_n + \hat{\pi}_n)\big(\sum_{i\in I}(\bar{x}_{in} - \hat{x}_{in})\big) + \sum_{n\in N\setminus\bar{Q}}(-\nu_n + \hat{\pi}_n)\big(\sum_{i\in I}(\bar{x}_{in} - \hat{x}_{in})\big)\end{aligned} \tag{4.53}$$

Note that, as $\bar{Q}$ is the set of maximum cardinality among all for which (4.49) holds, we have that

$$\sum_{n\in N\setminus\bar{Q}}\sum_{i\in I}(\bar{x}_{in} - \hat{x}_{in}) = 0$$

Moreover, for all $n \in \bar{Q}$, we have $\hat{\pi}_n = 0$ by the complementarity condition (4.35). Therefore, equation (4.53) simplifies to

$$\begin{aligned}&\sum_{in} IC_i\bar{x}_{in} + \sum_{int} MC_i\bar{y}_{int} - \sum_{in} IC_i\hat{x}_{in} - \sum_{int} MC_i(\hat{y}_{int} + \hat{\hat{y}}_{int}) \geq \\ &\sum_{n\in\bar{Q}}(-\nu_n)\big(\sum_{i\in I}(\hat{x}_{in} - \bar{x}_{in})\big)\end{aligned} \tag{4.54}$$

The left-hand side in (4.54) is strictly negative by our first contradiction assumption whereas the right-hand side in (4.54) is non-negative, which leads to a contradiction of our second contradiction assumption. This implies that, in any solution $(\hat{\cdot})$ of the MLCP (4.45), (4.26b) - (4.26n), (4.35) such that (4.48) holds, $\sum_{i\in I}\hat{x}_{in} = \sum_{i\in I}\bar{x}_{in}$ for all $n \in N$.

Now, by the optimalitity of $(\hat{\cdot})$ for the perturbed problem (i.e. problem (4.50)), we have

$$\begin{aligned}&\sum_{in} IC_i\hat{x}_{in} + \sum_{n\in N}(-\nu_n + \hat{\pi}_n)\sum_{i\in I}\hat{x}_{in} + \sum_{int} MC_i(\hat{y}_{int} + \hat{\hat{y}}_{int}) \leq \\ &\sum_{in} IC_i\bar{x}_{in} + \sum_{n\in N}(-\nu_n + \bar{\pi}_n)\sum_{i\in I}\hat{x}_{in} + \sum_{int} MC_i\bar{y}_{int}\end{aligned}$$

which implies

$$\sum_{in} IC_i\hat{x}_{in} + \sum_{int} MC_i(\hat{y}_{int} + \hat{\hat{y}}_{int}) \leq \sum_{in} IC_i\bar{x}_{in} + \sum_{int} MC_i\bar{y}_{int}$$

which is a contradiction. $\square$

5 Conclusions and future perspectives

5.1 Summary of the contributions

In the context of the climate and energy crises, it is important to ensure that the design of electricity markets is well suited for accompanying the necessary energy transition. In the European market, the fitness of the methodology for allocating transmission capacity, that is based on zonal electricity pricing, is recurrently questioned. In this dissertation, we contribute to the assessment of the efficiency of zonal pricing by proposing models and algorithms for analyzing this pricing paradigm in the context of the European market. Our contributions are structured in three main chapters.

In chapter 2, we focus on the impacts of transmission switching on the short-term efficiency of zonal electricity markets. We propose a two-stage model of the short-term market with zonal pricing that accounts for transmission switching at both the day-ahead and real-time stages. We show how the day-ahead problem with switching can be formulated as an adaptive robust optimization problem with mixed-integer recourse and present a new algorithm for solving the adversarial max-min problem that obeys the structure of an interdiction game.

In chapter 3, we turn to the analysis of zonal pricing in the long run. We model the capacity expansion problem with flow-based market coupling, which is the market coupling methodology that is currently used in Europe, and show that the equivalence between centralized and decentralized formulations ceases to hold in this case, unlike in markets that implement nodal pricing.

Finally, in chapter 4, we abstract from the direct comparison of nodal vs zonal pricing. Instead, we take zonal pricing as a given for the European market and investigate whether the long-run efficiency of zonal pricing can be restored by means of additional locational instruments. We introduce a common modeling framework for comparing zonal pricing with different classes of locational instruments: capacity-based signals, energy-based signals and redispatch markets.

The models developed in both chapters 3 and 4 for analyzing the long-run efficiency of zonal pricing lead us to identify a common pattern in the loss of efficiency: externalities imposed by the producers of electricity on the TSO. We show that the models that exhibit these kinds of inefficiency in the present thesis can be modeled as mixed complementarity problems and solved with an algorithm based on matrix splitting, which has an economic interpretation, that we discuss in Appendix B.

For all three of the main chapters of the thesis, we apply the models and algorithms presented on a realistic instance of the Central Western European system and discuss the impacts of zonal pricing on the efficiency of the electricity market.

5.2 Summary of the findings

Transmission switching

- **The efficiency of proactive and reactive switching is similar:** When considering the impacts of transmission switching, one must distinguish between the proactive case (when employed in day-ahead and real-time) and the reactive case (when employed only in real-time). We find that the two cases have similar efficiency. This suggests that considering the optimization of the topology when computing the zonal transmission constraints that are used in the day ahead delivers limited benefits. This is important, when we consider how much complexity switching adds to day-ahead operations. The interest of transmission switching in zonal pricing is therefore mainly found in re-dispatch and balancing.

- **Transmission switching is more beneficial in zonal compared to nodal pricing:** We find that reactive switching leads to a larger cost reduction in zonal pricing than in nodal pricing. This can be understood in light of the fact that zonal pricing uses simplified transmission constraints in the day-ahead market. This leads to suboptimal unit commitment with respect to the full representation of the network which can in part be mitigated by the additional flexibility that switching provides.

- **Transmission switching is not sufficient for restoring the efficiency of zonal pricing in the short term:** Although zonal pricing benefits more from transmission switching, transmission switching is not sufficient for fully compensating the loss of efficiency associated to a zonal network representation in the day ahead. Inefficiencies remain and are especially important in cases where the system is subject to a line contingency close to real time.

Zonal pricing and investment

- **The equivalence between centralized and decentralized formulations ceases to hold in zonal pricing with flow-based market**

coupling: In market design, a key question is whether one can find a set of prices that leads to the recovery of the centralized solution in a decentralized way. It is well known that it is indeed possible in nodal pricing. We show that this equivalence between centralized and decentralized formulations holds in some well-defined variations of zonal pricing, but that it ceases to hold in zonal pricing with FBMC, the current methodology used in Europe.

- **Important efficiency losses are associated to zonal pricing in the long run:** As zonal prices are computed based on a simplified representation of transmission constraints, they lead to investment decisions that are suboptimal with respect to the real network constraints. We quantify these losses of efficiency at 3% of the total investment and operating costs in the best case.

Additional locational instruments

- **In theory, the efficiency of zonal pricing can be restored with additional locational instruments:** We find that, in theory, the efficiency of nodal pricing can be recovered in zonal pricing through additional market-based locational instruments that are added on top of the electricity market. This theoretical result holds both in the case where re-dispatch is cost-based or market-based and where the instruments are capacity or energy-based.

- **In practice, the recovery of efficiency is subject to strong conditions that are unlikely to be satisfied in practice:** For the recovery of the nodal efficiency to hold, capacity-based instruments must have full technological and locational differentiation and the energy-based instruments must have full temporal granularity, which makes them complicated to implement. Moreover, the restoration of efficiency requires full inter-TSO cooperation and is subject to conditions on the definitions of zonal transmission constraints, that are currently not observed in practice.

- **Market-based re-dispatch can be detrimental to long-run efficiency if it is not corrected by a capacity-based instrument:** We find that when there is no locational capacity-based instrument, zonal pricing followed by market-based re-dispatch leads to highly inefficient investment. This loss of efficiency is due to the arbitrage opportunity that the re-dispatch market offers, that translates into additional profits for the producers in the short term and, as a consequence, additional investment in unnecessary peak-load capacity in the long term.

5.3 Discussion: importance of institutional context

It is now largely recognized that zonal pricing results in various types of inefficiencies regarding congestion management, which is confirmed by the results obtained in this thesis. The reader may wonder why, knowing these inefficiencies, Europeans do not switch to nodal.

In this section, we suggest that one reason why Europeans insist on zonal pricing is to be found in the institutional context that accompanied the development of the single market in the EU. We attempt to provide an intuitive explanation; a full discussion is beyond the scope of this dissertation. We argue that the European zonal system developed as a natural interpretation of legal requirements imposed by the "Completion of the Internal Market". It also reflects the respective domains of responsibilities of European and national authorities, which makes it institutionally feasible. This does not make it the best possible system, but deficiencies so far have been managed at an affordable cost.

The restructuring of the power system in the US was initiated by the introduction of some competition in the Public Utility Regulatory Policies Act (PURPA) of 1978; its evolution took place under the requirement of the 1935 Federal Power Act that electricity prices must be "just and reasonable", which FERC recognized to be satisfied by the nodal system. There was no such initiating event or requirement in the EU. The origin of the restructuring must be found elsewhere.

The European Community (and later the European Union) were founded on the postulate that economic integration between Member States would prevent new wars such as those that had twice destroyed the continent in the first part of the 20th century. Better economic integration was the most that could be hoped for; political integration was out of reach. The underpinning idea was that this integration could be achieved by the removal of barriers to trade between Member States. This principle was to apply to all sectors with possible exceptions for activities that provided "services of general economic interest". The slow progress towards that goal was suddenly accelerated in 1986 by the Single European Act that set 1992 as the deadline for the completion of the "Single Market". Extensions of the deadline were foreseen for network industries because of their complexity. The implicit reasoning was that, as in other sectors, competition would then develop on the market and achieve its integrating role: this is all that was intended and expected. The Treaties did not give any other power to the European Commission, but the Single European Act facilitated the passing of new legislation for moving towards the Single Market.

Member States are by nature geographical zones. Barriers to trade differ depending on the sectors: technical norms were the standard barriers against trading goods and services. Exclusive (monopoly) rights in generation and transmission were obvious barriers that prevented the trading of electricity.

Removing those rights would then make generation competitive provided transmission could take place through the grid owned by the incumbent generators. It was clear that access to the grid, even if subject to an open access constraint, could be difficult, all the more when it had to be negotiated, as in Germany.

The European Commission first proposed a two-tier approach that had worked well in telecommunications: a first Directive, enacted by the European Commission on the basis of EU competition law, would remove exclusive rights and equivalent effects. A second Directive would specify the more detailed aspects required by electricity. This second directive would be enacted by the Council and the Parliament, which (simplifying things) means that it would result from a consensus between Member States to reach the assigned goal. The sole idea of the European Commission using its own power to apply EU law in electricity was almost seen as a "casus belli" by Member States. The approach was abandoned and the whole task passed to a Council and Parliament Directive. Electricity was and remained essentially a national affair.

Directive 96/92 was the first outcome of this process. Not surprisingly, it restricted itself to the strict minimum: it removed exclusive rights but left maximum freedom on how to do so. [Han98] called it "a framework in the loosest sense of the word: its objectives are laid down in very general terminology and moreover, Member States are given a substantial degree of choice in how they go about introducing more competition into their electricity markets. Indeed the margin is so substantial that it would seem possible for the determined anti-market countries to avoid introducing any meaningful degree of competition at all".

The European Commission also reacted with dismay in 1998 to the situation: it immediately took the initiative of a second stronger directive and initiated the Florence Regulatory Forum consisting of the members of the Commission, network operators and regulators to come up with more meaningful proposals. The second Directive (2003/54EC) was accompanied by a Regulation (1228/2003) dedicated to transmission. We argue that this Regulation was instrumental in shaping the European zonal system and ensuring its persistence up to now.

"Congestion" is a key element of the Regulation: its definition is technically flawed but institutionally quite appropriate. Article 2 defines congestion as a problem encountered on interconnections due to international trade actions. Trading possibilities are defined in article 5(3) by transfer capacities on interconnections. Barriers to trade thus occur because of congestion on interconnections of insufficient transfer capacity. The language set the stage: Member States are zones and barriers to trade between zones occur because of congestion on interconnections characterized by transmission capacities. Possible congestion within zones does not matter in the process: it is the responsibility of the National Regulatory Authorities (article 23-2(a) of the Directive) without any relation to possible barrier to trade mentioned in the Regulation. What happens in the zone and on the interconnections are two different things: one is the concern of the Member State, the other is a barrier to trade between

Member States, which is the responsibility of the European Commission. Remarkably, this zonal view emerged in 2003, that is almost immediately after the 2000-2001 California crisis. A possible explanation for the fact that a more rigid zonal system than the one of California (zones could be split in California) was proposed in 2003 is that this was as far as one could go within the context of the European Treaties. Moving beyond that point would have required a much deeper interaction between Member States that could not be enforced at the time and remains difficult today. In other words, the zonal system became the reference framework because it was the only one that reflected the responsibilities enshrined in the institutions at the time and still today. Part of the work of the Florence Regulatory Forum for the next twenty years would consist in trying to make it work.

It is useful to note that nodal pricing, operational in PJM since 1998, was perfectly understood in some important European continental companies. Boiteux in EDF, who invented time-differentiated electricity prices (peak load pricing) had also written a paper on spatially differentiated electricity prices based on the interpretation of marginal cost obtained from optimal dispatch [BS52]. This was in 1952, well before the ground-breaking work of [SCTB88]. These economic concepts were also operational in some of the (published) EDF computational models [DM79]. For some reason, neither the industry nor the Member States pushed these ideas and there was no legal way the European Commission could have imposed them. Had one country implemented them, it could have created a burgeoning nodal system as observed in the progressive extension of PJM in the US or the enlargement of market splitting in Norway to the other Nordic countries[1].

The last 20 years were thus devoted to the discovery of the unintended difficulties of the zonal system. We only mention a sample of them. (i) Transmission capacity is a convenient concept for writing legal texts or policy recommendations but it cannot be defined in an unambiguous way; it also lacks basic properties like addition and subtraction. (ii) In contrast to what was initially thought, energy and transmission are not two separated activities that can be auctioned separately (the explicit auctions at the time) but they need to be

[1]The Nordic system is a special case among zonal systems. Its reputation is that it works well, which generated some questions among researchers, especially after the California debacle (see [AB06]). Besides the causes mentioned in the above paper, we can mention that, in contrast with the continental market that developed on the basis of the closed-area system with interconnections enabling occasional cross-border transactions, the Nordic grid was built for wholesale transport between its Northern part with its massive hydro resources and Denmark with, at the time, massive coal and combined heat and power. This was meant to take advantage of economic exchanges created by changing hydro conditions; these have now been replaced by the massive provisions of flexibility from the North to the large wind Danish generation installations. The very linear structure of the grid, mainly from North to South, and the geographic dispersion of resources, which drove the variable zone structure (market splitting) may have also helped (Sweden, which was a single zone, also had to introduce market splitting to abide to complaints of market power due to congestion). These are only intuitions; researchers in the Nordic countries have sometimes advocated moving to a nodal system, but the argument was not further pursued.

treated jointly (the move to implicit auctions[2]). (iii) Even though the Florence Forum never went as far as the nodal system, it introduced flowgates to replace transmission capacities to cross some borders. (iv) Efforts were made to avoid having flowgates inside zones as this would split them, which would have logically required two domestic prices and introduced a direct link between "European" and domestic affairs; this would have destroyed the institutional logic of the system. (v) Countertrading (foreseen in Regulation 1228/2003) did not cost much in the beginning but became expensive later on. Notwithstanding these technical difficulties, the fiction of the zonal system remains convenient for certain stakeholders, which can continue arguing in terms of transmission capacities between Member States. This is in particular the case for the so-called 70% rule of the "Clean Energy Package".

Many stakeholders have now realized these problems. The rumor at the time of this writing is that many are convinced that the zonal system is bound to cause real problems and that one should go to the nodal system. However, this requires unraveling the large legislation that developed on the basis of the zonal system, which could be a real issue. Texas took several years to move from zonal to nodal, one may imagine what it would take to do so in a system covering more than 40 zones. This would also require a general consensus among Member States as the construction of a nodal system is probably not something that the European Commission could impose in the immediate future.

5.4 Future perspectives

We mention here a number of future research perspectives that are triggered by the present work.

Consideration of uncertainty

In what concerns the models, the consideration of uncertainty would make for an interesting extension. Uncertainty is indeed a major component of electric power systems and it is of growing importance as large amount of renewable energy sources continue to be integrated to the grid. In the context of this dissertation, it would be particularly relevant to consider the following two types of uncertainty: (i) regulatory uncertainty and (ii) uncertainty related to the re-dispatch price in market-based re-dispatch.

As we discuss in the introduction, the regulations associated to transmission capacity allocation in the European electricity markets are continuously evolving. These changes of regulation influence the electricity prices and, therefore, create additional investment risks. As the regulation is expected to keep evolving in order to accompany the energy transition, with a possible transition to

[2]Explicit auctioning is again the new paradigm between the UK and the EU after Brexit, although the British government seems to have decided to review these new arrangements, see: `https://www.reuters.com/business/energy/britain-seeks-views-plugging-back-into-european-power-market-2021-09-30/`

nodal pricing in the medium term, it would be interesting to understand the impacts of regulatory risk associated to transmission capacity allocation on investment. The deterministic models of the long-run equilibrium presented in this dissertation would probably be a promising starting point.

In chapter 4, we propose a model of the long-run equilibrium of zonal pricing followed by market-based re-dispatch. We assume in this model that the re-dispatch price is known to producers when they bid in the day-ahead market. In practice, however, some uncertainty is associated to the re-dispatch price. It would be interesting to extend the models presented in chapter 4 to account for this specific type of uncertainty, in order to understand how it modifies the arbitrage behavior of producers.

Multiplicity of solutions

In this work, we identify two designs of transmission capacity allocation for which the long-term equilibrium is not equivalent to a single optimization problem: (i) the current methodology of flow-based market coupling and (ii) zonal pricing with market-based re-dispatch. We show that these two situations can be modeled mathematically as generalized Nash equilibrium problems. These types of problems lead in general to a multiplicity of solutions. It would be interesting to understand further the implications of this multiplicity of solutions on the efficiency of the market.

Convergent algorithm

We propose in Appendix B a splitting-based algorithm for identifying one solution to these two long-term equilibrium problems. Although we observe convergence for certain starting points, convergence of this algorithm is not guaranteed for the class of problems that we study in the thesis. The question of whether there exists a way to modify the iterations in order to guarantee convergence remains open.

A Dimensionality reduction of the CWE dataset

In this appendix, we describe in further detail the dimensionality reduction that we perform on the Central Western Europe dataset in order to render it tractable for investment problems. We perform a reduction on both the load duration curve and the network.

A.1 Load duration curve

We start with hourly data of demand and renewable production on the entire year and fix the number of time periods of the reduced load duration curve to 20. We then compute the best approximation in the sense of the Euclidean norm of the hourly net load duration curve by a piecewise constant function of 20 pieces, as shown on Figure A.1. We use the dynamic programming algorithm presented in [KK88] to solve for the best approximation.

A.2 Network reduction

For the network reduction, we also fix *a priori* the targeted number of nodes to 100. The nodes are clustered with hierarchical clustering using the Euclidean Commute Time (ECT) distance [YVW$^+$05] on the graph corresponding to the network, with the edge weights set to the difference in nodal prices between each pair of nodes. The cross-zonal lines are removed from the network for the clustering in order to obtain only clusters of nodes from the same bidding zone. The ECT distance is chosen for the clustering in order to favor clusters that are strongly connected. The results of the clustering can be visualized in Figure A.2.

Once the buses of the network have been clustered, it remains to compute the PTDF matrix as well as the thermal capacities of the reduced network. For the reduced PTDF matrix, we use the injection-independent method described in [FDS18]. We then compute the thermal capacities of the lines of the reduced network in a way that minimizes the Euclidean norm of the difference between

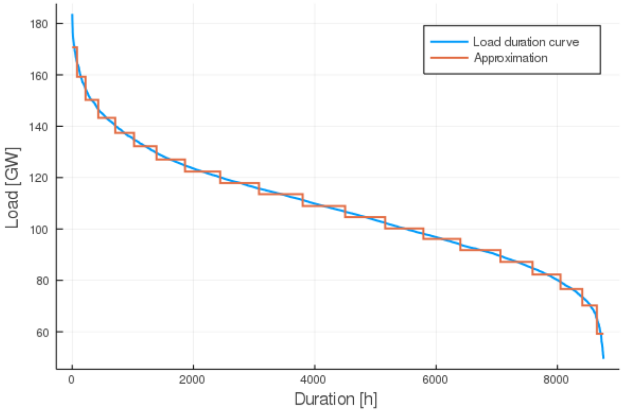
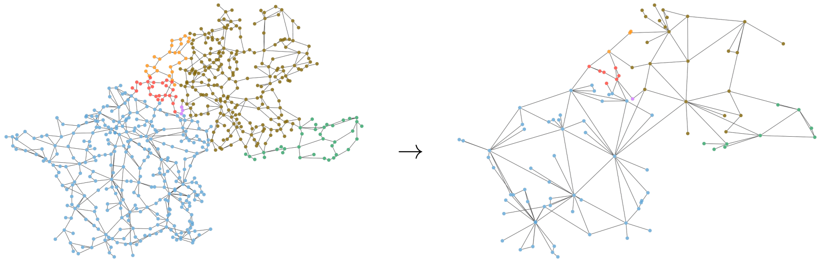


Figure A.1: Best piecewise constant approximation of the aggregate load duration curve of the CWE area.



(a) Original network

(b) Reduced network

Figure A.2: Visualization of the reduction of the CWE network the average nodal price of the nodes in each cluster and the new nodal price of the cluster. More precisely, let us denote by N the initial set of nodes in the network and by M the set of nodes in the reduced network, with $|N| = 632$ and $|M| = 100$. Let $M(n)$ be the aggregated cluster to which node $n \in N$ belongs and let $N(m)$ be the set of nodes of the initial network that belongs to cluster m. We first compute the zonal prices on the 20 periods obtained by the dimensionality reduction of the load duration curve, that we denote by $\bar{\rho}_{nt}$. Based on these prices, we can compute the average of the nodal prices for each cluster, i.e.

$$\bar{\rho}_{mt} = \frac{\sum_{n \in N(m)} \bar{\rho}_{nt}}{|N(m)|}, \ \ \forall m \in M$$

We can now formulate the problem of the minimization of Euclidean distance from the new prices ρ_{mt} to the average prices of the initial network $\bar{\rho}_{mt}$:

$$\min_{\substack{TC \\ y,s,r,f \\ \rho,\mu,\psi,\phi,\lambda}} \sum_{m \in M, t \in T} (\rho_{mt} - \bar{\rho}_{mt})^2 \tag{A.1a}$$

$$\sum_{i \in I, m \in M} MC_i \cdot y_{imt} + \sum_{m \in M} VOLL \cdot s_{mt} =$$

$$\sum_{n} D_{mt}\rho_{mt} - \sum_{im} X_{im}\mu_{imt} - \sum_{k} TC_k(\lambda^+_{kt} + \lambda^-_{kt}), t \in T \tag{A.1b}$$

$$X_{im} - y_{imt} \geq 0, i \in I, m \in M, t \in T \tag{A.1c}$$

$$D_{mt} - s_{mt} \geq 0, m \in M, t \in T \tag{A.1d}$$

$$- r_{mt} + \sum_{i \in I} y_{imt} + s_{mt} - D_{mt} = 0, m \in M, t \in T \tag{A.1e}$$

$$f_{kt} - \sum_{m \in M} PTDF_{km} \cdot r_{mt} = 0, k \in K, t \in T \tag{A.1f}$$

$$\sum_{m \in M} r_{mt} = 0, t \in T \tag{A.1g}$$

$$- TC_k \leq f_{kt} \leq TC_k, k \in K, t \in T \tag{A.1h}$$

$$\rho_{mt} + \sum_{k} PTDF_{km}\psi_{kt} - \phi = 0, m \in M, t \in T \tag{A.1i}$$

$$MC_i - \rho_{mt} + \mu_{imt} \geq 0, i \in I, m \in M, t \in T \tag{A.1j}$$

$$VOLL - \rho_{mt} + \delta_{mt} \geq 0, m \in M, t \in T \tag{A.1k}$$

$$- \psi_{kt} - \lambda^-_{kt} + \lambda^+_{kt} = 0, k \in K, t \in T \tag{A.1l}$$

$$y \geq 0, s \geq 0, \mu, \lambda^+, \lambda^- \geq 0 \tag{A.1m}$$

In this problem, constraint (A.1b) represents strong duality of the market clearing problem, constraints (A.1c)-(A.1h) are the primal constraints and constraints (A.1i)-(A.1l) are the dual constraints. Note that the goal is to find the thermal capacities TC_k of the lines, that are thus variables of the problem. The bilinear terms in the strong duality constraint (A.1b) make the problem non-convex. These non-convexities and the large size of the problem make it complicated to solve directly. Therefore, we do not pass it directly to a solver, but instead we solve it using the alternating direction method. That is, we start by fixing the capacities TC_k to an initial value and we solve (A.1) by alternatively fixing TC_k and λ_k^+, λ_k^- until no more progress can be made, which yields a suboptimal solution.

B A splitting-based algorithm for solving generalized Nash equilibrium problems

In this appendix, we present a splitting-based algorithm for solving the GNE that is associated to a general competitive market with externalities. This method is a basic algorithm for solving LCPs. We note that it has a natural economic interpretation in relation with the theory of Pigouvian taxation and missing markets in this context. Our goal is to describe this economic interpretation, to draw the link with the splitting methods for LCPs, and to comment on the adequacy of this method for studying failures in electricity market design, such as the ones that we study in this thesis.

We consider a general competitive economy on which the set of goods traded is described by $K = \{1, ..., |K|\}$. We assume that there are a set of producers $I = \{1, ..., |I|\}$ and a set of consumers $J = \{1, ..., |J|\}$ that trade in the markets of one or several of the goods. We start by describing the competitive equilibrium when there is no externality. Then, we introduce externalities and, finally, we present the splitting-based algorithm for computing one equilibrium in the latter case.

B.1 No externality

There are three types of agents in our economy: producers, consumers and a Walrasian auctioneer that clears the markets. Let us describe the profit-maximizing problems of each type of agent:

Producers. We assume that the decision problem of the producers is described as a general linear optimization problem that includes revenues from the sale of the goods produced. Producer i decides on x^p_{ik}, its production of good k, and y^p_{il}, the value of its additional decision l, by solving the following optimization problem:

$$\begin{aligned} \mathcal{P}_i = \max_{x^p_{ik}, y^p_{il}} \quad & \sum_{k\in K} c^p_{ik} x^p_{ik} + \sum_{l\in L_i} d^p_{il} y^p_{il} + \sum_{k\in K} \rho_k x^p_{ik} \\ \text{s.t.} \quad & \sum_{k\in K} A^p_{ikm} x^p_{ik} + \sum_{l\in L_i} B^p_{ilm} y^p_{il} + f^p_{im} \geq 0, \;\; \forall m \in M_i \end{aligned} \tag{B.1}$$

Here, L_i is the set of indices of its additional decision variables y^p_{il}, M_i is the set of indices of its linear constraints and c^p_{ik}, d^p_{il} are the coefficients of the goods and additional decision variables in its objective function, A^p_{ikm}, B^p_{ilm} the corresponding coefficients in the constraints, and f^p_{im} the independent terms in the constraints.

Consumers. The problem of the consumers is similar, but now the objective includes expenses from the purchase of the goods consumed. It can be written as follows:

$$\begin{aligned} \mathcal{C}_j = \max_{x^c_{jk}, y^c_{jl}} \quad & \sum_{k\in K} c^c_{jk} x^c_{jk} + \sum_{l\in L_j} d^c_{jl} y^c_{jl} - \sum_{k\in K} \rho_k x^c_{jk} \\ \text{s.t.} \quad & \sum_{k\in K} A^c_{jkm} x^c_{jk} + \sum_{l\in L_j} B^c_{jlm} y^c_{jl} + f^c_{jm} \geq 0, \;\; \forall m \in M_j \end{aligned} \tag{B.2}$$

Auctioneer. There is a Walrasian auctioneer that decides the prices ρ_k by clearing the market. Its representative profit-maximizing problem can be written as:

$$\mathcal{W}_k = \max_{\rho_k} \quad \rho_k \Big(\sum_i x^p_{ik} - \sum_j x^c_{jk} \Big) \tag{B.3}$$

By slightly abusing notation by omitting the concatenation on all indices, we can say that $(\bar{x}^p_{ik}, \bar{y}^p_{il}, \bar{x}^c_{jk}, \bar{y}^c_{j,l}, \bar{\rho}_k)$ is a competitive equilibrium if

$$\begin{cases} (\bar{x}^p_{ik}, \bar{y}^p_{il}) \text{ solves } \mathcal{P}_i \text{ given } \bar{\rho}_k \\ (\bar{x}^c_{jk}, \bar{y}^c_{jl}) \text{ solves } \mathcal{C}_j \text{ given } \bar{\rho}_k \\ (\bar{\rho}_k) \text{ solves } \mathcal{W}_k \text{ given } (\bar{x}^p_{ik}, \bar{x}^c_{jk}) \end{cases} \tag{B.4}$$

The equilibrium can be obtained by solving the following welfare-maximizing optimization problem:

$$\begin{aligned} \max_{x,y,\rho} \quad & \sum_{i\in I} \Big(\sum_{k\in K} c^p_{ik} x^p_{ik} + \sum_{l\in L_i} d^p_{il} y^p_{il} \Big) + \sum_{j\in J} \Big(\sum_{k\in K} c^c_{jk} x^c_{jk} + \sum_{l\in L_j} d^c_{jl} y^c_{jl} \Big) \\ \text{s.t.} \quad & \sum_{k\in K} A^p_{ikm} x^p_{ik} + \sum_{l\in L_i} B^p_{ilm} y^p_{il} + f^p_{im} \geq 0, \;\; \forall m \in M_i, i \in I \\ & \sum_{k\in K} A^c_{jkm} x^c_{jk} + \sum_{l\in L_j} B^c_{jlm} y^c_{jl} + f^c_{jm} \geq 0, \;\; \forall m \in M_j, j \in J \end{aligned} \tag{B.5}$$

The equivalence between the welfare-maximizing problem (B.5) and the Nash equilibrium described by problems (B.1)-(B.3) can be easily observed by comparing their necessary and sufficient KKT conditions.

B.2 Externality

We are interested in equilibria that arise in the presence of externalities and in particular when decisions of some agent influence the feasible set of the profit-maximizing problem of another agent. We restrict ourselves to the case of producer-to-producer externalities through the additional variables y for simplicity of the exposition, but the analysis would be similar for a more general kind of externality. The problem of the producers can be written as follows:

$$\begin{aligned}\mathcal{P}_i = \max_{x_{ik}^p, y_{il}^p} \quad & \sum_{k\in K} c_{ik}^p x_{ik}^p + \sum_{l\in L_i} d_{il}^p y_{il}^p + \sum_{k\in K} \rho_k x_{ik}^p \\ \text{s.t.} \quad & \sum_{k\in K} A_{ikm}^p x_{ik}^p + \sum_{l\in L_i} B_{ilm}^p y_{il}^p + \sum_{\substack{e\in I\setminus\{i\} \\ l\in L_e}} E_{el}^i y_{el}^p + f_{im}^p \geq 0, \ \ \forall m \in M_i\end{aligned} \tag{B.6}$$

where $\sum_{\substack{e\in I\setminus\{i\} \\ l\in L_e}} E_{el}^i y_{el}^p$ is the additional term that represents the externality. In this case, an equilibrium, if it exists, is not necessarily efficient. Efficiency can however be restored by means of a Pigouvian tax[1], which can also be interpreted as the price of the externality if a market for this externality would be available. Let us denote by π_{iel} the price of the externality that agent $e \in I\setminus\{i\}$ causes to agent $i \in I$ through its decision associated to variable y_{el}^p, with $l \in L_e$. Taking into account the markets for externalities and the change in revenues that it incurs to the producers, the profit-maximizing problem of the producers can be written as follows:

$$\begin{aligned}\mathcal{P}_i = \max_{x_{ik}^p, y_{il}^p} \quad & \sum_{k\in K} c_{ik}^p x_{ik}^p + \sum_{l\in L_i} d_{il}^p y_{il}^p + \sum_{k\in K} \rho_k x_{ik}^p + \sum_{l\in L_i} \big(\sum_{e\in I\setminus\{i\}} \pi_{eil} \big) y_{li}^p - \\ & \sum_{\substack{e\in I\setminus\{i\} \\ l\in L_e}} \pi_{iel} y_{iel}^p \\ \text{s.t.} \quad & \sum_{k\in K} A_{ikm}^p x_{ik}^p + \sum_{l\in L_i} B_{ilm}^p y_{il}^p + \sum_{\substack{e\in I\setminus\{i\} \\ l\in L_e}} E_{el}^i y_{iel}^p + f_{im}^p \geq 0, \ \ \forall m \in M_i\end{aligned} \tag{B.7}$$

The markets of externalities are associated to the following market clearing conditions:

$$\pi_{iel} \text{ free } \ \perp y_{iel}^p - y_{el}^p = 0 \ \ \forall i \in I, e \in I\setminus\{i\}, l \in L_e \tag{B.8}$$

[1]The optimal tax level that restores efficiency in a market with externalities is named after [Pig32]. More background on Pigouvian taxation can be obtained from [MCWG95, Chapter 11.B].

The restoration of the missing markets for externalities brings us back to the conditions of the first welfare theorem, and the equilibirum in this case can be obtained by soling the welfare-maximizing problem:

$$\begin{aligned}
\max_{x,y}\ & \sum_{i\in I}\Big(\sum_{k\in K} c^p_{ik}x^p_{ik} + \sum_{l\in L_i} d^p_{il}y^p_{il}\Big) + \sum_{j\in J}\Big(\sum_{k\in K} c^c_{jk}x^c_{jk} + \sum_{l\in L_j} d^c_{jl}y^c_{jl}\Big)\\
\text{s.t. } & \sum_{k\in K} A^p_{ikm}x^p_{ik} + \sum_{l\in L_i} B^p_{ilm}y^p_{il} + \sum_{\substack{e\in I\setminus\{i\}\\ l\in L_e}} E^i_{el}y^p_{iel} + f^p_{im} \geq 0,\ \ \forall m\in M_i, i\in I\\
& \sum_{k\in K} A^c_{jkm}x^c_{jk} + \sum_{l\in L_j} B^c_{jlm}y^c_{jl} + f^c_{jm} \geq 0,\ \ \forall m\in M_j, j\in J\\
& y^p_{iel} - y^p_{el} = 0\ \ \forall i\in I, e\in I\setminus\{i\}, l\in L_e
\end{aligned} \tag{B.9}$$

In this optimization problem, the optimal value of the dual variable associated to the last constraint gives the price of the externality or, equivalently, the value of the Pigouvian tax needed to restore efficiency.

Note that given the optimal value of the dual variables $\bar{\pi}_{iel}$, problem (B.9) can also be formulated by applying the principle of Lagrangian duality as follows:

$$\begin{aligned}
\max_{x,y}\ & \sum_{i\in I}\Big(\sum_{k\in K} c^p_{ik}x^p_{ik} + \sum_{l\in L_i} d^p_{il}y^p_{il}\Big) + \sum_{j\in J}\Big(\sum_{k\in K} c^c_{jk}x^c_{jk} + \sum_{l\in L_j} d^c_{jl}y^c_{jl}\Big) +\\
& \sum_{\substack{i\in I\\ e\in I\setminus\{i\}\\ l\in L_e}} \bar{\pi}_{iel}(y^p_{iel} - y^p_{el})\\
\text{s.t. } & \sum_{k\in K} A^p_{ikm}x^p_{ik} + \sum_{l\in L_i} B^p_{ilm}y^p_{il} + \sum_{\substack{e\in I\setminus\{i\}\\ l\in L_e}} E^i_{el}y^p_{iel} + f^p_{im} \geq 0,\ \ \forall m\in M_i, i\in I\\
& \sum_{k\in K} A^c_{jkm}x^c_{jk} + \sum_{l\in L_j} B^c_{jlm}y^c_{jl} + f^c_{jm} \geq 0,\ \ \forall m\in M_j, j\in J
\end{aligned} \tag{B.10}$$

When there is no market for externalities, however, the producer that causes an externality is not exposed to its cost and the clearing constraint $y^p_{iel} - y^p_{el} = 0$ is totally internalized by producer i. Mathematically, this translates into dropping variable y^p_{el} of the objective in problem (B.10) and the question is now to find the values of primal variables (x, y) *and* dual variables π such that they solve the following problem:

$$\begin{aligned}
\max_{x,y,\pi} \quad & \sum_{i\in I}\Big(\sum_{k\in K} c^p_{ik}x^p_{ik} + \sum_{l\in L_i} d^p_{il}y^p_{il}\Big) + \sum_{j\in J}\Big(\sum_{k\in K} c^c_{jk}x^c_{jk} + \sum_{l\in L_j} d^c_{jl}y^c_{jl}\Big) - \\
& \sum_{\substack{i\in I\\ e\in I\setminus\{i\}\\ l\in L_e}} \pi_{iel}y^p_{el} \\
\text{s.t.} \quad & \sum_{k\in K} A^p_{ikm}x^p_{ik} + \sum_{l\in L_i} B^p_{ilm}y^p_{il} + \sum_{\substack{e\in I\setminus\{i\}\\ l\in L_e}} E^i_{el}y^p_{iel} + f^p_{im} \geq 0, \;\; \forall m\in M_i, i\in I \\
& \sum_{k\in K} A^c_{jkm}x^c_{jk} + \sum_{l\in L_j} B^c_{jlm}y^c_{jl} + f^c_{jm} \geq 0, \;\; \forall m\in M_j, j\in J \\
[\pi_{iel}] \;\; & y^p_{iel} - y^p_{el} = 0 \;\; \forall i\in I, e\in I\setminus\{i\}, l\in L_e
\end{aligned} \tag{B.11}$$

where variable π_{iel} appears both in the objective and as the dual variable of the last constraint.

B.3 Algorithm

Iterative scheme There is a very natural iterative scheme that can be employed in order to solve for an equilibrium in the presence of externality, as described by problem (B.11):

Step 0. *Initialization.* Let π^0_{iel} be an arbitrary initial value and $\epsilon > 0$ a given tolerance. Set $\nu = 0$.

Step 1. *General iteration.* Given π^ν_{iel}, solve (B.11) with π_{iel} fixed to π^ν_{iel} and let $\pi^{\nu+1}_{iel}$ be an optimal dual variable associated to the last constraint.

Step 2. *Test for termination.* If $\|\pi^{\nu+1}_{iel} - \pi^\nu_{iel}\| < \epsilon$, terminate. Otherwise, return to Step 1 with ν replaced by $\nu + 1$.

Link with the splitting method As it turns out, this problem has an interpretation as a splitting algorithm for solving a generic LCP problem. Indeed, let us first note that the equilibrium with externality can be formulated as an $LCP(q, M)$ by taking the joint KKT conditions of the linear profit-maximizing problems of each agent, where M is the matrix of the LCP and q is the independent term.

This LCP is almost equivalent to the LCP that corresponds to the welfare-maximizing problem, problem (B.9) in our case. The two LCPs only slightly differ by their matrices: in the welfare-maximizing problem, the value corresponding to entry (y^p_{el}, π_{iel}) is equal to one, whereas it is zero in $LCP(q, M)$.

Let us denote by B the matrix of the LCP associated to the welfare-maximizing problem. We have that

$$M = B + C \tag{B.12}$$

where C is of the following form (in block formulation):

$$C = \begin{array}{c} \\ y^p_{el} \\ \\ \\ \\ \end{array} \begin{array}{c} \pi_{iel} \\ \begin{pmatrix} 0 & \dots & I & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \end{array}$$

Here, I is the rectangular identity matrix, i.e. a matrix with 1 in the entries associated to line y^p_{el} and column π_{iel}, and 0 otherwise.

Our iterative scheme is a direct application of the splitting algorithm for solving $LCP(q, M)$ with splitting (B.12), as presented for instance in [CPS09] (Algorithm 5.2.1).

Applications The algorithm presented in this section is particularly well suited when there is only a limited number of externalities, as this implies that the deviation from welfare-maximization is limited and one can hope for faster convergence. As it turns out, this applies to a number of problems related to electricity market design with transmission constraints. In part II of this dissertation, we have presented two different problems that obey this structure: in chapter 3, the long-term equilibrium with flow-based market coupling corresponds to an equilibrium with externalities, where the investment in capacity by private firms influences the feasible set of the TSO. In chapter 4, the long-term equilibrium of zonal pricing with market-based re-dispatch also obeys this structure: the production decision of the producers in the zonal market influences the feasible set of the TSO in the re-dispatch market.

Convergence Finally, let us note that there is no convergence guarantee for the splitting algorithm applied to the problems that we presented in this dissertation. Experimentally, we have observed convergence for both problems when starting from the pure welfare-maximization problem, i.e. when $\pi^0_{iel} = 0$, but that the iterates can diverge for a general starting point. Moreover, when convergence is observed, it is not monotonic[2].

[2] In our case, the convergence would be monotonic if the sequence of distances between two iterates $\|\pi^{\nu+1}_{iel} - \pi^{\nu}_{iel}\|$ would be decreasing.

Bibliography

[AB06] Eirik S. Amundsen and Lars Bergman. Why has the Nordic electricity market worked so well? *Utilities Policy*, 14(3):148–157, September 2006. URL: https://ideas.repec.org/a/eee/juipol/v14y2006i3p148-157.html.

[ACE22] ACER. Final assessment of the eu wholesale electricity market design. Technical report, April 2022. URL: https://documents.acer.europa.eu/Official_documents/Acts_of_the_Agency/Publication/ACER's%20Final%20Assessment%20of%20the%20EU%20Wholesale%20Electricity%20Market%20Design.pdf.

[AGK+20] Mirjam Ambrosius, Veronika Grimm, Thomas Kleinert, Frauke Liers, Martin Schmidt, and Gregor Zöttl. Endogenous price zones and investment incentives in electricity markets: An application of multilevel optimization with graph partitioning. *Energy Economics*, 92:104879, 2020. URL: https://www.sciencedirect.com/science/article/pii/S014098832030219X, doi:https://doi.org/10.1016/j.eneco.2020.104879.

[AHT19] Victor Ahlqvist, Pär Holmberg, and Thomas Tangerås. Central- versus Self-Dispatch in Electricity Markets. Working Papers EPRG 1902, Energy Policy Research Group, Cambridge Judge Business School, University of Cambridge, January 2019. URL: https://ideas.repec.org/p/enp/wpaper/eprg1902.html.

[ALPS21] Ignacio Aravena, Quentin Lété, Anthony Papavasiliou, and Yves Smeers. Transmission capacity allocation in zonal electricity markets. *Operations Research*, 69(4):1240–1255, 2021. doi:10.1287/opre.2020.2082.

[AP17] I. Aravena and A. Papavasiliou. Renewable energy integration in zonal markets. *IEEE Transactions on Power Systems*, 32(2):1334–1349, March 2017. doi:10.1109/TPWRS.2016.2585222.

[Aus20] Austrian Energy Agency. Increasing Transparency for Flow-Based Market Coupling in European Electricity Trading. https://oesterreichsenergie.at/fileadmin/user_upload/IncreasingTransparencyForFBMC_20201216.pdf, 2020.

[AVFM20] G. Antonopoulos, S. Vitiello, G. Fulli, and M. Masera. Nodal pricing in the European Internal Electricity Market. Technical report, 2020. doi:10.2760/41018.

[AZ06] Parviz Adib and Jay Zarnikau. Chapter 11 - texas: The most robust competitive market in north america1. In Fereidoon P. Sioshansi and Wolfgang Pfaffenberger, editors, *Electricity Market Reform*, Elsevier Global Energy Policy and Economics Series, pages 383–417. Elsevier, Oxford, 2006. URL: https://www.sciencedirect.com/science/article/pii/B9780080450308500138, doi:https://doi.org/10.1016/B978-008045030-8/50013-8.

[BBB12] C. Barrows, S. Blumsack, and R. Bent. Computationally efficient optimal transmission switching: Solution space reduction. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–8, July 2012. doi:10.1109/PESGM.2012.6345550.

[BC19] Ronan Bolton and Thies Clausen. Ambiguous interventions: Capacity remuneration mechanisms and the politics of energy transition in britain and germany. *Energy Research & Social Science*, 56:101218, 2019. URL: https://www.sciencedirect.com/science/article/pii/S2214629618311654, doi:https://doi.org/10.1016/j.erss.2019.101218.

[BEKS17] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017. URL: https://doi.org/10.1137/141000671.

[BJ01] Mette Bjørndal and Kurt Jørnsten. Zonal Pricing in a Deregulated Electricity Market. *The Energy Journal*, 0(1):51–73, 2001. URL: https://ideas.repec.org/a/aen/journl/2001v22-01-a03.html.

[BJ05] Mette Bjørndal and Kurt Jørnsten. The deregulated electricity market viewed as a bilevel programming problem. *Journal of Global Optimization*, 33(3):465–475, November 2005. doi:10.1007/s10898-004-1939-9.

[BJ08] Mette Bjørndal and Kurt Jørnsten. *Investment Paradoxes in Electricity Networks*, pages 593–608. Springer New York, New York, NY, 2008. doi:10.1007/978-0-387-77247-9_23.

[Boi60] Marcel Boiteux. Peak-Load Pricing. *The Journal of Business*, 33(2):157–179, 1960.

[Boi64] Marcel Boiteux. *The Choice of Plant and Equipment for the Production of Electric Energy.* 1964. in James Nelson, ed. *Marginal Cost Pricing in Practice*, Englewood Cliffs, N.J., Prentice-Hall.

[BS52] Marcel Boiteux and Paul Stasi. Sur la détermination des prix de revient de dévelopment dans un système interconnecté de production-distribution, 1952. Reprinted in G. Morlat et F. Bessière (eds.), *Vingt-cinq ans d'économie électrique*, Paris Dunod, 361-400.

[BS01] Jacqueline Boucher and Yves Smeers. Alternative models of restructured electricity systems, Part 1: No market power. *Operations Research*, 49(6):821–838, 2001. `doi:10.1287/opre.49.6.821.10017`.

[CCLW16] Alberto Caprara, Margarida Carvalho, Andrea Lodi, and Gerhard J. Woeginger. Bilevel knapsack with interdiction constraints. *INFORMS Journal on Computing*, 28(2):319–333, 2016. `arXiv:https://doi.org/10.1287/ijoc.2015.0676`, `doi:10.1287/ijoc.2015.0676`.

[Com03] Commission for Energy Regulation. Existing structure of tariffs in ireland: Transmission, distribution, supply. Technical report, December 2003.

[CPS09] Richard Cottle, Jong-Shi Pang, and Richard Stone. *The Linear Complementarity Problem.* SIAM, 2009. `doi:10.1137/1.9780898719000`.

[CRE17] CREG. Functioning and design of the central west european day-ahead flow based market coupling for electricity: Impact of tsos discretionary actions. Technical report, December 2017. Available online: `https://www.creg.be/sites/default/files/assets/Publications/Studies/F1687EN.pdf`. Accessed: April 26 2021.

[CRE19] CREG. Study on the best forecast of remedial actions to mitigate market distortion. Technical report, October 2019. URL: `https://www.creg.be/sites/default/files/assets/Publications/Studies/F1987EN.pdf`.

[CS05] Peter Cramton and Steven Stoft. A capacity market that makes sense. *The Electricity Journal*, 18(7):43–54, 2005. URL: `https://www.sciencedirect.com/science/article/pii/S1040619005000850`, `doi:https://doi.org/10.1016/j.tej.2005.07.003`.

[CSF19] Michele Conforti, Marco Di Summa, and Yuri Faenza. Balas formulation for the union of polytopes is optimal. *Mathematical Programming*, 180(1-2):311–326, January 2019. doi:10.1007/s10107-018-01358-9.

[DH02] Laurens J. De Vries and Rudi. A. Hakvoort. An Economic Assessment of Congestion Management Methods for Electricity Transmission Networks. *Competition and Regulation in Network Industries*, 3(4):425–466, 2002. doi:10.1177/178359170200300403.

[DHL17] Iain Dunning, Joey Huchette, and Miles Lubin. Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320, 2017. doi:10.1137/15M1020575.

[DM79] J.C. Dodu and A. Merlin. Recent improvements of the mexico model for probabilistic planning studies. *International Journal of Electrical Power & Energy Systems*, 1(1):46–56, 1979. doi:https://doi.org/10.1016/0142-0615(79)90031-0.

[Dua19] Vincent Duane. Nodal vs zonal pricing for the electricity market, February 2019. FSR Energy interview. [Audio file]. Retrieved from https://soundcloud.com/fsregulation-energy-and-climate/zonal-versus-nodal-electricity-pricing-the-pjm-experience-vincent-p-duane on 9 September 2021.

[DW11] Justin Dijk and Bert Willems. The effect of counter-trading on competition in electricity markets. *Energy Policy*, 39(3):1764–1773, 2011. URL: https://www.sciencedirect.com/science/article/pii/S0301421511000188, doi:https://doi.org/10.1016/j.enpol.2011.01.008.

[EGK+21] Jonas Egerer, Veronika Grimm, Thomas Kleinert, Martin Schmidt, and Gregor Zöttl. The impact of neighboring markets on renewable locations, transmission expansion, and generation investment. *European Journal of Operational Research*, 292(2):696–713, 2021. URL: https://www.sciencedirect.com/science/article/pii/S0377221720309486, doi:https://doi.org/10.1016/j.ejor.2020.10.055.

[Eic21] Anselm Eicke. Optimal generation investment under suboptimal dispatch. Working paper, Internationale Energiewirtschaftstagung an der TU Wien, 2021. URL: https://iewt2021.eeg.tuwien.ac.at/download/contribution/fullpaper/176/176_fullpaper_20210831_082627.pdf.

[EKH20] Anselm Eicke, Tarun Khanna, and Lion Hirth. Locational investment signals: How to steer the siting of new generation capacity in power systems? *Energy Journal*, 41(1):281–304, 2020. doi:10.5547/01956574.41.6.aeic.

[ENT18] ENTSO-E. Overview of transmission tariffs in europe: Synthesis 2018. Technical report, May 2018. Available online: https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/mc-documents/TTO_Synthesis_2018.pdf. Accessed: April 13 2021.

[ENT20] ENTSO-E. Mid-term Adequacy Forecast Executive Summary 2020 Edition, 2020. Available online: https://eepublicdownloads.entsoe.eu/clean-documents/sdc-documents/MAF/2020/MAF_2020_Executive_Summary.pdf.

[ENT21] ENTSO-E. Options for the design of european electricity markets in 2030, 2021. Abailable online: https://eepublicdownloads.entsoe.eu/clean-documents/Publications/Market%20Committee%20publications/210331_Market_design%202030.pdf.

[ENT22] ENTSO-E. Single Day-ahead Coupling (SDAC). https://www.entsoe.eu/network_codes/cacm/implementation/sdac/, 2022. Accessed: 2022-07-05.

[ES05] Andreas Ehrenmann and Yves Smeers. Inefficiencies in European congestion management proposals. *Utilities Policy*, 13(2):135 – 152, 2005. doi:10.1016/j.jup.2004.12.007.

[ES11] Andreas Ehrenmann and Yves Smeers. Generation capacity expansion in a risky environment: A stochastic equilibrium analysis. *Operations Research*, 59(6):1332–1346, 2011. doi:10.1287/opre.1110.0992.

[ES22] Anselm Eicke and Tim Schittekatte. Fighting the wrong battle ? A critical assessment of arguments against nodal electricity prices in the European debate. Working paper, MIT Energy Initiative, 2022. URL: https://energy.mit.edu/wp-content/uploads/2022/02/MITEI-WP-2022-01.pdf.

[Eur] European Parliament, Council of the European Union. Regulation (EU) 2019/943 of the European Parliament and if the Council of 5 june 2019 on the internal market for electricity.

[Eur09] European Parliament, Council of the European Union. Regulation (EC) No 714/2009 of the European Parliament and of the Council of 13 July 2009 on conditions for access to the network for cross-border exchanges in electricity and repealing Regulation (EC) No 1228/2003, July 2009.

[Eur15] European Commission. Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management, July 2015.

[Eur18a] European Commission. Communication from the Commission. A Clean Planet for all, November 2018. COM(2018) 773 final `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0773`.

[Eur18b] European TSOs (CORE). Explanatory note on the day-ahead and intraday common capacity calculation methodologies for the core ccr, June 2018.

[Eur19] European Parliament, Council of the European Union. Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity, June 2019.

[Eur20] European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Stepping up Europe's 2030 climate ambition, September 2020. COM(2020) 562 final `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0562`.

[FDS18] P. Fortenbacher, T. Demiray, and C. Schaffner. Transmission network reduction method using nonlinear optimization. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7, 2018. `doi:10.23919/PSCC.2018.8442974`.

[FHK+21] Christoph Fraunholz, Dirk Hladik, Dogan Keles, Dominik Möst, and Wolf Fichtner. On the long-term efficiency of market splitting in Germany. *Energy Policy*, 149(February 2020), 2021. `doi:10.1016/j.enpol.2020.111833`.

[FLMS19] Matteo Fischetti, Ivana Ljubić, Michele Monaci, and Markus Sinnl. Interdiction games and monotonicity, with application to knapsack problems. *INFORMS Journal on Computing*, 31(2):390–410, 2019. `arXiv:https://doi.org/10.1287/ijoc.2018.0831`, `doi:10.1287/ijoc.2018.0831`.

[FOF08] E. B. Fisher, R. P. O'Neill, and M. C. Ferris. Optimal transmission switching. *IEEE Transactions on Power Systems*, 23(3):1346–1355, Aug 2008. `doi:10.1109/TPWRS.2008.922256`.

[FP03] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY, 2003 edition, February 2003.

[FP08] Dominique Finon and Virginie Pignon. Electricity and long-term capacity adequacy: The quest for regulatory mechanism compatible with electricity market. *Utilities Policy*, 16(3):143–158, 2008. Capacity Mechanisms in Imperfect Electricity Markets. URL: https://www.sciencedirect.com/science/article/pii/S0957178708000039, doi:https://doi.org/10.1016/j.jup.2008.01.002.

[FRC12] J. D. Fuller, R. Ramasra, and A. Cha. Fast heuristics for transmission-line switching. *IEEE Transactions on Power Systems*, 27(3):1377–1386, Aug 2012. doi:10.1109/TPWRS.2012.2186155.

[GKL+19] Veronika Grimm, Thomas Kleinert, Frauke Liers, Martin Schmidt, and Gregor Zöttl. Optimal price zones of electricity markets: a mixed-integer multilevel model and global solution approaches. *Optimization Methods and Software*, 34(2):406–436, 2019. arXiv:https://doi.org/10.1080/10556788.2017.1401069, doi:10.1080/10556788.2017.1401069.

[GMS+16] Veronika Grimm, Alexander Martin, Martin Schmidt, Martin Weibelzahl, and Gregor Zöttl. Transmission and generation investment in electricity markets: The effects of market splitting and network fee regimes. *European Journal of Operational Research*, 254(2):493–509, oct 2016. doi:10.1016/j.ejor.2016.03.044.

[GMS+18] Veronika Grimm, Alexander Martin, Christian Sölch, Martin Weibelzahl, and Gregor Zöttl. Market-Based Redispatch May Result in Inefficient Dispatch. *SSRN Electronic Journal*, (2008):1–27, 2018. doi:10.2139/ssrn.3120403.

[GMWZ16] Veronika Grimm, Alexander Martin, Martin Weibelzahl, and Gregor Zöttl. On the long run effects of market splitting: Why more price zones might decrease welfare. *Energy Policy*, 94:453–467, 2016. doi:10.1016/j.enpol.2015.11.010.

[GOS13] G. Gürkan, O. Ozdemir, and Y. Smeers. Generation Capacity Investments in Electricity Markets : Perfect Competition. Discussion Paper 2013-045, Tilburg University, Center for Economic Research, 2013. URL: https://ideas.repec.org/p/tiu/tiucen/97828b2c-3630-4b66-8ac9-f24c469cefbd.html.

[GRSZ19] Veronika Grimm, Bastian Rückel, Christian Sölch, and Gregor Zöttl. Regionally differentiated network fees to affect incentives for generation investment. *Energy*, 177:487–502, 2019. doi:10.1016/j.energy.2019.04.035.

[GRSZ21] Veronika Grimm, Bastian Rückel, Christian Sölch, and Gregor Zöttl. The impact of market design on transmission and generation investment in electricity markets. *Energy Economics*, 93:104934, 2021. URL: https://www.sciencedirect.com/science/article/pii/S0140988320302747, doi:https://doi.org/10.1016/j.eneco.2020.104934.

[Han98] Leigh Hancher. Delimitation of energy law jurisdiction: the eu and its member states: From organisational to regulatory conflicts. *Journal of Energy & Natural Resources Law*, 16(1):42–67, 1998. arXiv:https://doi.org/10.1080/02646811.1998.11433126, doi:10.1080/02646811.1998.11433126.

[Har91] Patrick T. Harker. Generalized Nash games and quasi-variational inequalities. *European Journal of Operational Research*, 54(1):81–94, sep 1991. doi:10.1016/0377-2217(91)90325-P.

[Har18] Nicolas Hary. Delimitation of bidding zones for electricity markets in europe and the consideration of internal congestions. Technical report, 2018. Deloitte Newsletter Power & Utilities. Available online: https://www2.deloitte.com/content/dam/Deloitte/fr/Documents/financial-advisory/economicadvisory/deloitte_delimitation-zones-marches-electriques-Europe-et-consideration-des-congestions-internes.pdf. Accessed 26 April 2021.

[HB13] N. Hutcheon and J. W. Bialek. Updated and validated power flow model of the main continental european transmission network. In *2013 IEEE Grenoble Conference*, pages 1–5, 2013.

[HFO+10] K. W. Hedman, M. C. Ferris, R. P. O'Neill, E. B. Fisher, and S. S. Oren. Co-optimization of generation unit commitment and transmission switching with n-1 reliability. *IEEE Transactions on Power Systems*, 25(2):1052–1063, May 2010.

[HL15] Pär Holmberg and Ewa Lazarczyk. Comparison of congestion management techniques: Nodal, zonal and discriminatory pricing. *The Energy Journal*, Volume 36(2):145–166, Apr 2015. URL: http://www.iaee.org/en/publications/ejarticle.aspx?id=2611, doi:10.5547/01956574.36.2.7.

[HO19] Benjamin F. Hobbs and Shmuel S. Oren. Three waves of u.s. reforms: Following the path of wholesale electricity market restructuring. *IEEE Power and Energy Magazine*, 17(1):73–81, 2019. doi:10.1109/MPE.2018.2873952.

[HOFO10] K. Hedman, R. O'Neill, E. Fisher, and S. Oren. Optimal transmission switching with contingency analysis. In *IEEE PES General Meeting*, pages 1–1, July 2010. doi:10.1109/PES.2010.5589437.

[Hog92] William Hogan. Contract networks for electricity power transmission: technical reference. *Journal of Regulatory Economics*, 242(September 1990):211–242, 1992.

[Hog98] William Hogan. Competitive electricity market design: A wholesale primer, 1998. Available online: `https://scholar.harvard.edu/whogan/files/empr1298.pdf`. Accessed 26 April 2021.

[Hog99] William W Hogan. Restructuring the Electricity Market : Institutions for Network Systems. Technical Report April, 1999. URL: `https://scholar.harvard.edu/whogan/files/hjp0499.pdf`.

[Hog13] William W. Hogan. Electricity scarcity pricing through operating reserves. *Economics of Energy & Environmental Policy*, Volume 2(Number 2), 2013. URL: `https://EconPapers.repec.org/RePEc:aen:eeepjl:2_2_a04`.

[HP16] J. Han and A. Papavasiliou. The impacts of transmission topology control on the european electricity network. *IEEE Transactions on Power Systems*, 31(1):496–507, 2016.

[HS18] Lion Hirth and Ingmar Schlecht. Market-Based Redispatch in Zonal Electricity Markets. *SSRN Electronic Journal*, (055), 2018. `doi:10.2139/ssrn.3286798`.

[HS20] Lion Hirth and Ingmar Schlecht. Market-Based Redispatch in Zonal Electricity Markets: The Preconditions for and Consequence of Inc-Dec Gaming. Working paper, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg, 2020. URL: `http://hdl.handle.net/10419/222925`.

[HSMT19] Lion Hirth, Ingmar Schlecht, Christoph Maurer, and Bernd Tersteegen. Cost or marketbased? future redispatch procurement in germany. Technical report, October 2019. *Neon neue energieökonomik and Consentec*. Available online: `https://www.bmwi.de/Redaktion/EN/Publikationen/Studien/future-redispatch-procurement-in-germany.pdf?__blob=publicationFile&v=2`, accessed 19 August 2021.

[Int20] International Atomic Energy Agency. Country nuclear power profiles - france, 2020. Available online: `https://cnpp.iaea.org/countryprofiles/France/France.htm`. Accessed 03/03/21.

[Jos06] Paul L. Joskow. Competitive Electricity Markets and Investment in New Generating Capacity. *AEI-Brookings Joint Center Working Paper No. 06-14*, May 2006. `doi:10.2139/ssrn.902005`.

[Jos07] Paul Joskow. The new energy paradigm, 2007. Oxford University Press, Ch. Competitive Electricity Markets and Investment in New Generation Capacities, pp. 76–122.

[Kev11] Wood Kevin. Bilevel network interdiction models: Formulations and solutions, 2011. URL: https://calhoun.nps.edu/handle/10945/38416.

[KJD+16] Burak Kocuk, Hyemin Jeon, Santanu S. Dey, Jeff Linderoth, James Luedtke, and Xu Andy Sun. A cycle-based formulation and valid inequalities for dc power transmission problems with switching. *Operations Research*, 64(4):922–938, 2016. arXiv:https://doi.org/10.1287/opre.2015.1471, doi:10.1287/opre.2015.1471.

[KK88] Hiroshi Konno and Takahito Kuno. Best piecewise constant approximation of a function of single variable. *Operations Research Letters*, 7(4):205–210, 1988. URL: https://www.sciencedirect.com/science/article/pii/0167637788900302, doi:https://doi.org/10.1016/0167-6377(88)90030-2.

[LSP22] Quentin Lété, Yves Smeers, and Anthony Papavasiliou. An analysis of zonal electricity pricing from a long-term perspective. *Energy Economics*, 107:105853, 2022. URL: https://www.sciencedirect.com/science/article/pii/S0140988322000366, doi:https://doi.org/10.1016/j.eneco.2022.105853.

[Män15] Petri Mäntysaari. *Transmission Marketplaces*, pages 347–417. Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-16513-4_5.

[MCWG95] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Number 9780195102680 in OUP Catalogue. Oxford University Press, 1995. URL: https://ideas.repec.org/b/oxp/obooks/9780195102680.html.

[Mee20] Leonardo Meeus. *The Evolution of Electricity Markets in Europe*. Edward Elgar Publishing, Cheltenham, UK, 2020.

[MLTW13] A. Marien, P. Luickx, A. Tirez, and D. Woitrin. Importance of design parameters on flowbased market coupling implementation. In *2013 10th International Conference on the European Energy Market (EEM)*, pages 1–8, May 2013. doi:10.1109/EEM.2013.6607298.

[MT19] Irena Milstein and Asher Tishler. On the effects of capacity payments in competitive electricity markets: Capacity adequacy, price cap, and reliability. *Energy Policy*, 129:370–385, 2019. URL: https://www.sciencedirect.com/science/article/pii/S0301421519301119, doi:https://doi.org/10.1016/j.enpol.2019.02.028.

[Nat22] National Grid ESO. Net Zero Market Reform - Phase 3 Assessment and Conclusions. https://www.nationalgrideso.com/document/258871/download, May 2022.

[NB11] Karsten Neuhoff and Rodney Boyd. International Experiences of Nodal Pricing Implementation - Frequently Asked Questions, 2011.

[New16] David Newbery. Missing money and missing markets: Reliability, capacity auctions and interconnectors. *Energy Policy*, 94:401–410, 2016. URL: https://www.sciencedirect.com/science/article/pii/S0301421515301555, doi:https://doi.org/10.1016/j.enpol.2015.10.028.

[NHN11] Karsten Neuhoff, Benjamin F. Hobbs, and David Newbery. Congestion Management in European Power Networks: Criteria to Assess the Available Options. Discussion Papers of DIW Berlin 1161, DIW Berlin, German Institute for Economic Research, 2011. URL: https://ideas.repec.org/p/diw/diwwpp/dp1161.html.

[OBH+05] R. P. O'Neill, R. Baldick, U. Helman, M. H. Rothkopf, and W. Stewart. Dispatchable transmission in rto markets. *IEEE Transactions on Power Systems*, 20(1):171–179, Feb 2005. doi:10.1109/TPWRS.2004.841168.

[OHK+10] Richard P. O'Neill, Kory W. Hedman, Eric A. Krall, Anthony Papavasiliou, and Shmuel S. Oren. Economic analysis of the n-1 reliable unit commitment and transmission switching problem using duality concepts. *Energy Systems*, 1(2):165–195, May 2010. doi:10.1007/s12667-009-0005-6.

[Ope20] Open Power System Data. Data package conventional power plants, 2020. Version 2020-10-01. doi:https://doi.org/10.25832/conventional_power_plants/2020-10-01.

[Ope21] Open Energy Information. Transparent cost database, 2021. Available online: https://openei.org/apps/TCDB. Accessed 03/03/21.

[OS13] Giorgia Oggioni and Yves Smeers. Market failures of Market Coupling and counter-trading in Europe: An illustrative model based discussion. *Energy Economics*, 35:74–87, jan 2013. doi:10.1016/j.eneco.2011.11.018.

[Ozd13] O. Ozdemir. *Simulation modeling and optimization of competitive electricity markets and stochastic fluid systems.* PhD thesis, Tilburg University, 2013.

[Pig32] Arthur Cecil Pigou. The Economics of Welfare, 1932. London: Macmillan.

[PMVDB05] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans. Usefulness of dc power flow for active power flow analysis. In *IEEE Power Engineering Society General Meeting, 2005*, pages 454–459 Vol. 1, 2005. doi:10.1109/PES.2005.1489581.

[POO11] A. Papavasiliou, S. S. Oren, and R. P. O'Neill. Reserve requirements for wind power integration: A scenario-based stochastic programming framework. *IEEE Transactions on Power Systems*, 26(4):2197–2206, Nov 2011. doi:10.1109/TPWRS.2011.2121095.

[RPZ+09] P. A. Ruiz, C. R. Philbrick, E. Zak, K. W. Cheung, and P. W. Sauer. Uncertainty management in the unit commitment problem. *IEEE Transactions on Power Systems*, 24(2):642–651, May 2009. doi:10.1109/TPWRS.2008.2012180.

[RTE19] RTE. Is the nodal design adapted to the european energy transition ?, 2019.

[SCC17] Kathryn M. Schumacher, Amy E. M. Cohn, and Richard Li-Yang Chen. Algorithm for the n-2 security-constrained unit commitment problem with transmission switching. *INFORMS Journal on Computing*, 29(4):645–659, 2017. arXiv:https://doi.org/10.1287/ijoc.2017.0751, doi:10.1287/ijoc.2017.0751.

[SCTB88] Fred C. Schweppe, Michael C. Caramanis, Richard D. Tabors, and Roger E. Bohn. *Spot Pricing of Electricity.* Springer US, 1988. doi:10.1007/978-1-4613-1683-1.

[SL17] X. Andy Sun and Álvaro Lorca. *Robust Optimization in Electric Power Systems Operations*, pages 227–258. Springer International Publishing, Cham, 2017. doi:10.1007/978-3-319-55581-2_7.

[SM21] Tim Schittekatte and Leonardo Meeus. Capacity Remuneration Mechanisms in the EU: today, tomorrow, and a look further ahead. Working paper, Robert Schuman Centre for Advanced Studies, 2021. URL:

https://cadmus.eui.eu/bitstream/handle/1814/72460/RSC%202021_71.pdf?sequence=1&isAllowed=y.

[SMA14] A. Street, A. Moreira, and J. M. Arroyo. Energy and reserve scheduling under a joint generation and transmission security criterion: An adjustable robust optimization approach. *IEEE Transactions on Power Systems*, 29(1):3–14, Jan 2014. doi:10.1109/TPWRS.2013.2278700.

[SMBP19] Kaarthik Sundar, Sidhant Misra, Russell Bent, and Feng Pan. Spatial and Topological Interdiction for Transmission Systems. Technical Report arXiv:1904.08330v1 [math.OC], April 2019. URL: https://arxiv.org/pdf/1904.08330.pdf.

[Sta19] Stattnet. Transmission grid tariffs for 2019: Model description and rates. Technical report, 2019.

[SZ20] Lukas ; Schmidt and Jonas Zinke. One price fits all? Wind power expansion under uniform and nodal pricing in Germany. EWI Working Paper, No. 20/06, 2020. URL: http://hdl.handle.net/10419/227510.

[Tan05] Yoshihiro Tanaka. On the existence of duality gaps for mixed integer programming. *International Journal of Systems Science*, 36(6):375–379, 2005. doi:10.1080/00207720500111756.

[Ten19] TenneT. Product specifications - Reserve Power Other Purposes, April 2019. https://netztransparenz.tennet.eu/fileadmin/user_upload/SO_NL/Productspecifications_reserve_power_other_purposes_-_English_version.pdf.

[THE19] THEMA Consulting Group. Review of the swedish tranmission grid tariff model. Technical report, April 2019.

[TW21] Ryan Triolo and Frank Wolak. Quantifying the benefits of nodal market design in the texas electricity market. Working paper, Program on Energy and Sustainable Development, 2021. URL: https://fsi.stanford.edu/publication/quantifying-bene-ts-nodal-market-design-texas-electricity-market.

[VBD16] Kenneth Van den Bergh, Jonas Boury, and Erik Delarue. The flow-based market coupling in central western europe: Concepts and definitions. *The Electricity Journal*, 29(1):24 – 29, 2016. doi:https://doi.org/10.1016/j.tej.2015.12.004.

[VD16] Kenneth Van den Bergh and Erik Delarue. An improved method to calculate injection shift keys. *Electric Power Systems Research*, 134:197–204, 2016. URL: https://www.sciencedirect.com/

science/article/pii/S0378779616300062, doi:https://doi.org/10.1016/j.epsr.2016.01.020.

[vdWH11] Adriaan Hendrik van der Weijde and Benjamin F. Hobbs. Locational-based coupling of electricity markets: Benefits from coordinating unit commitment and balancing markets. *Journal of Regulatory Economics*, 39(3):223–251, 2011. doi:10.1007/s11149-011-9145-4.

[WC13] J. Wu and K. W. Cheung. On selection of transmission line candidates for optimal transmission switching in large power networks. In *2013 IEEE Power Energy Society General Meeting*, pages 1–5, July 2013. doi:10.1109/PESMG.2013.6672912.

[Wei17] Martin Weibelzahl. Nodal, zonal, or uniform electricity pricing: how to deal with network congestion. *Frontiers in Energy*, 11(2):210–232, 2017. doi:10.1007/s11708-017-0460-z.

[YVW+05] Luh Yen, Denis Vanvyve, Fabien Wouters, François Fouss, Michel Verleysen, and Marco Saerens. Clustering using a random walk based distance measure. In *ESANN 2005, European Symposium on Artificial Neural Networks*, pages 317–324, Bruges (Belgium), April 2005.

[ZZ12] Long Zhao and Bo Zeng. An Exact Algorithm for Two-stage Robust Optimization with Mixed Integer Recourse Problems. Technical report, 2012. URL: https://pdfs.semanticscholar.org/6b15/8459656b321a0791cc9df4a09af161c8f5e2.pdf.

[ZZ13a] Bo Zeng and Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *OR Letters*, 41(5):457 – 461, 2013. URL: http://www.sciencedirect.com/science/article/pii/S0167637713000618, doi:https://doi.org/10.1016/j.orl.2013.05.003.

[ZZ13b] L. Zhao and B. Zeng. Vulnerability analysis of power grids with line switching. *IEEE Transactions on Power Systems*, 28(3):2727–2736, Aug 2013. doi:10.1109/TPWRS.2013.2256374.