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

Quentin Lété*, Yves Smeers, Anthony Papavasiliou

Center for Operations Research and Econometrics, UCLouvain

Abstract

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 paper, we revisit the question of capacity allocation in zonal markets from a long-term perspective. We model the capacity expansion problem in zonal markets in which inter-zonal transmission capacity allocation is organized through flow-based market coupling, which is an approximation of power flow equations in aggregate networks that is employed in European market design. We demonstrate that the classical result of equivalence between centralized and decentralized formulations in transmission-constrained markets ceases to hold in this case. We propose a model of the decentralized capacity expansion problem with flow-based market coupling as a generalized Nash equilibrium that we formulate as a linear complementarity problem. We then perform simulations of the capacity expansion problem with nodal pricing and three variations of zonal pricing on a realistic instance of the Central Western European network and comment on the impacts of flow-based market coupling on investment.

Keywords: Zonal electricity markets, Transmission capacity allocation, Capacity expansion

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 paper, 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

*Corresponding author.
Email address: quentin.lete@uclouvain.be (Quentin Lété)

an 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 accruing by plants 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 Joskow (2007) 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 Boiteux (1960) and Boiteux (1964), 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 Boiteux (1960) 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 paper is precisely to understand the impacts of the difference in market designs on investment.

We concentrate in this paper on market designs that implement nodal and zonal systems. Nodal and zonal systems differ primarily in the way transmission constraints are considered in the market. A nodal system considers every node in the transmission grid and clears the market on the basis of a direct current (DC) approximation of the power flow equations in which every transmission line is accounted for. In the zonal market, intra-zonal congestion is omitted and the market is cleared on the basis of simplified transmission constraints. This implies some cost that, depending on how it is allocated, may or may not modify the cash flows accruing to plants. European legislation recently introduced a new version of the zonal system, known as the 70% rule. This allocation rule has an impact on congestion, and the need to remove it in real time, and thus also an effect on cash flows.

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. The main reason for this is that 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 for the design that is currently being implemented in the European electricity market, called flow-based market coupling (FBMC). 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 the paper.

Existing literature on missing money has, to a large extent, omitted considerations related to congestion. Cramton and Stoft (2005) do 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 Cramton and Stoft (2005); Finon and Pignon (2008), and scarcity pricing based on operating reserve demand curves Hogan (2013). There are, nowadays, still active discussions about these remedies and their ability to solve the missing money problem with sometimes contradictory results: Milstein and Tishler (2019) find that capacity markets can mitigate the missing money problem, while Newbery (2016) 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 paper.

The paper is organized as follows: we start by reviewing existing literature on the modeling of zonal transmission constraints and on the long-term impacts of a zonal design on investment in section 2. In section 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 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 paper on a realistic instance of the Central Western Europe (CWE) system. Section 7 provides a brief conclusion.

2. Literature review and contributions

2.1. Literature review

In this section, we discuss existing literature relevant to our work. We first discuss the modeling of zonal transmission constraints. We then discuss results regarding the long-term impacts of zonal markets.

In a zonal market, the exchange between nodes of the same zone are assumed to be unlimited. Transmission constraints are only applied on the net position (exports - imports) of every zone. In contrast, a nodal market will impose constraints on the nodal net injections. However, unlike in a nodal design, in the zonal design there is no unique methodology for defining the transmission constraints on the set of zonal net positions, and several methods have been proposed and implemented. Bjørndal and Jornsten (2001) is one of the first papers that describe a zonal market clearing problem. It has been referred to as ideal zonal pricing in subsequent literature (Ehrenmann and Smeers, 2005; Weibelzahl, 2017). The idea of this zonal pricing model is to specify all transmission constraints with a nodal resolution and to add constraints that impose that prices within the same zone should be equal. The authors propose to use this model in a setting of dynamic bidding zones: starting from a unique zone, if the market cannot be cleared with the ideal zonal pricing model, the bidding zone is split into two or more zones. This procedure is repeated until the model is feasible. This is in contrast to the zonal pricing models described in Ehrenmann and Smeers (2005), where the delimitation of the zones is assumed to be fixed at the time of market clearing. The idea is then to define an aggregate network and to compute transfer capacities on the interconnectors. The models in Ehrenmann and Smeers (2005) are similar to the historical methodology for market coupling that is used in Europe, which is referred to as Available Transfer Capacity (ATC) market coupling. With pre-defined bidding zones, a limit on the bilateral exchange between each pair of zones (the ATCs) is computed. More recently, part of the European market has transitioned to FBMC in order to improve the efficiency of market coupling. The idea of FBMC is to compute an available margin of capacity for cross-border trade on cross-zonal lines and on a reduced set of internal lines, based on the expected state of the grid in the period that is considered. Constraints on zonal net positions are then imposed based on these available margins. A detailed description of FBMC is provided in Van den Bergh et al. (2016). One drawback of the current methodology used for FBMC is that it is complicated to model and its results are hard to replicate. The reason is that it relies on a series of parameters that are decided by Transmission System Operators (TSO), and different TSOs have different ways of computing these parameters. In a previous publication (Aravena et al., 2021), we have developed models of ATCMC and FBMC that abstract from the definitions of these parameters and focus on the main principles of ATCMC and FBMC, as stated in EU regulations. In the present paper, the FBMC model of Aravena et al. (2021) serves as our starting point for studying capacity expansion based on FBMC.

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 Grimm et al. (2016a). 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 Grimm et al. (2021) 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 Grimm et al. (2016a) 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 Grimm et al. (2016a). Here, ATC market coupling is assumed with exogenous ATC values. A third paper in this stream of work is Egerer et al. (2021), 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 Grimm et al. (2016b), 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 (Grimm et al., 2019) is focused on methods for solving the large tri-level mixed-integer mathematical program in the form of which the studied problem can be formulated. 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 Ambrosius et al. (2020), 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 Fraunholz et al. (2021) 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.

2.2. Contributions

In this section, we specify our contributions 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 Aravena et al. (2021) in order to account for generation investment by private firms. As we show in section 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 3.2.

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 paper by employing a model of FBMC which is independent of exogenous parameters, as we propose in previous work (Aravena et al., 2021). 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 paper 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 (Ehrenmann and Smeers, 2011; Ozdemir, 2013).

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. Capacity expansion models in transmission-constrained electricity markets

Quoting Paul Joskow in (Joskow, 2006), "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 (Boiteux, 1960). 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. This section serves as an introduction to the more advanced section that follows on capacity expansion with FBMC, which is our main modeling contribution. Its aim is to introduce our modeling assumptions regarding capacity expansion. It will also be the basis for comparing FBMC with nodal pricing and other versions of zonal pricing, as they relate to investment.

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 define the set of all net injections at the network buses that are feasible for the DC power flow equations, denoted by $\mathcal{R}$:[1]

$$\begin{aligned}\mathcal{R} = \Big\{ & r \in \mathbb{R}^{|N|} \;\Big|\; \exists f \in \mathbb{R}^{|K|} \; : \\ & f_k = \sum_{n \in N} PTDF_{kn} \cdot r_n, k \in K \\ & \sum_{n \in N} r_n = 0, -TC_k \leq f_k \leq TC_k, k \in K \Big\}\end{aligned} \tag{1}$$

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 of line k. This set $\mathcal{R}$ describes completely the network constraints in the case of nodal pricing. Using this set, one

[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 (Aravena et al., 2021).

can define the capacity expansion model from the central planner perspective as follows:[2]

$$\min_{x,y,s,r} \sum_{i \in I, n \in N} IC_i \cdot x_{in} + \sum_{i \in I, n \in N, t \in T} MC_i \cdot y_{int} + \sum_{n \in N, t \in T} VOLL \cdot s_{nt} \tag{2a}$$

$$(\mu_{int}) : y_{int} \leq x_{in} + X_{in}, i \in I, n \in N, t \in T \tag{2b}$$

$$(\rho_{nt}) : r_{nt} = \sum_{i \in I} y_{int} + s_{nt} - D_{nt}, n \in N, t \in T \tag{2c}$$

$$r_{:t} \in \mathcal{R}, t \in T \tag{2d}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{2e}$$

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 (2b), the network constraints (2d) and nodal balance (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 Ozdemir (2013). 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. In particular, the TSO maximizes the value of its grid, i.e. the congestion rent, in line with the literature on markets with transmission operations (Hogan, 1992; Boucher and Smeers, 2001; Ozdemir, 2013). Regarding consumers, we assume that electricity is priced at VOLL by the regulator in case of demand curtailment. Figure 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 characterized the decentralized market-based model here, but we simply highlight one important complementarity condition from the set of KKT conditions of this problem:

$$0 \leq x_{in} \perp IC_i - \sum_{t \in T} \mu_{int} \geq 0 \;\; \forall i \in I, n \in N \tag{3}$$

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 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 paper in order to illustrate the different models that we present. The data of the example is presented in Figure 2. The instance is a three-node,

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

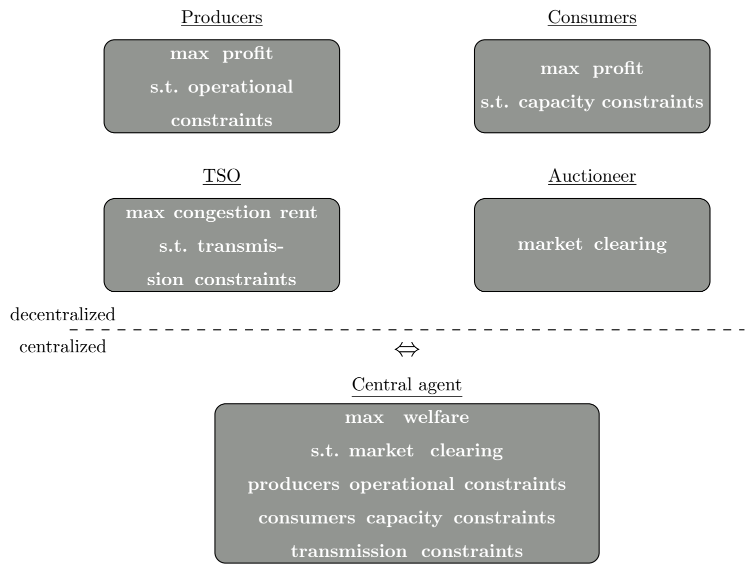
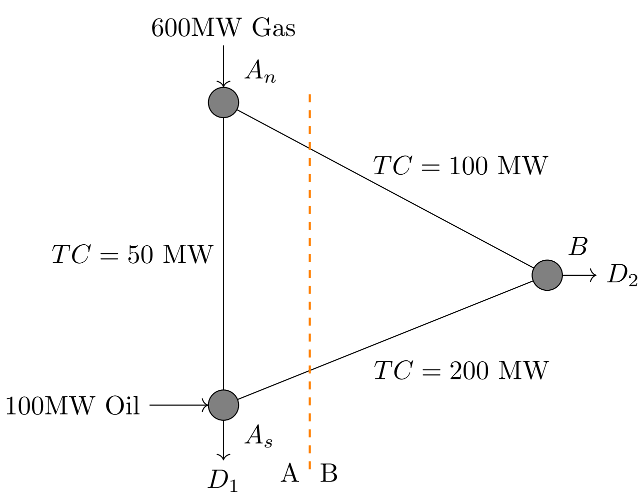


Figure 1: Equivalence between the decentralized game of the 4 groups of agents in the market and the welfare maximization problem of the central agent.



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

two-zone system with a load duration curve that is aggregated into three demand blocks. 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 €.

3.2. 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 C for the details of this derivation. We now denote by $\mathcal{P}^{\mathrm{PA}}$ 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 subsequent variations of the set of zonal net positions that we will present in section 4. Mathematically, $\mathcal{P}^{\mathrm{PA}}$ can be defined as follows:

$$\begin{aligned}\mathcal{P}^{\mathrm{PA}} = \Bigg\{ p \in \mathbb{R}^{|Z|} \,\Bigg|\, & \exists (f,r) \in \mathbb{R}^{|K|} \times \mathbb{R}^{|N|} : p_z = \sum_{n \in N(z)} r_n \;\; \forall z \in Z, \\ & f_k = \sum_n PTDF_{kn} \cdot r_n \;\; \forall k \in K, \sum_n r_n = 0, -TC_k \leq f_k \leq TC_k \;\; \forall k \in K \Bigg\}\end{aligned} \tag{4}$$

Note that although every line of the network is accounted for in equation (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 (4) is not equivalent to the ideal zonal pricing model proposed in Bjørndal and Jornsten (2001). In fact, as discussed in Weibelzahl (2017), 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

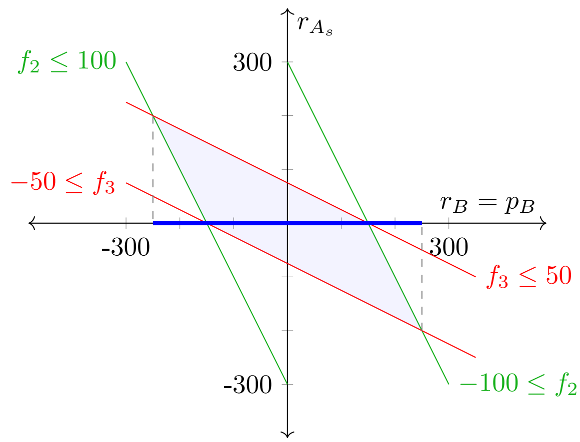


Figure 3: Illustration of the feasible set of net injections in the nodal model (light blue area) and the feasible net positions in the zonal model (thick blue line) for the 3-node 2-zone network. The feasible set of zonal net positions in the PA model is a projection of the set of feasible nodal net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

pricing might be infeasible, as mentioned in Ehrenmann and Smeers (2005). 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 2, the set $\mathcal{P}^{\text{PA}}$ can be made explicit as follows:

$$
\begin{aligned}
\mathcal{P}^{\text{PA}} = \Bigg\{ p \in \mathbb{R}^2 \,\Bigg|\, & \exists (r, f) \in \mathbb{R}^3 \times \mathbb{R}^3 : \\
& p_A = r_{A_n} + r_{A_s}, p_B = r_B \\
& f_1 = \frac{1}{3} r_{A_n} + \frac{2}{3} r_{A_s}, -200 \leq f_1 \leq 200 \\
& f_2 = \frac{2}{3} r_{A_n} + \frac{1}{3} r_{A_s}, -100 \leq f_2 \leq 100 \\
& f_3 = \frac{1}{3} r_{A_n} - \frac{1}{3} r_{A_s}, -50 \leq f_3 \leq 50 \\
& r_{A_n} + r_{A_s} + r_B = 0 \Bigg\}
\end{aligned}
\tag{5}
$$

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, 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 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 $\mathcal{P}^{\mathrm{PA}}$ set, as follows:

$$\min_{x,y,s,p} \sum_{i\in I, z\in Z} IC_i \cdot x_{iz} + \sum_{i\in I, z\in N, t\in T} MC_i \cdot y_{izt} + \sum_{n\in N, t\in T} VOLL \cdot s_{zt} \tag{6a}$$

$$(\mu_{izt}) : y_{izt} \leq x_{iz} + X_{iz}, i \in I, z \in Z, t \in T \tag{6b}$$

$$(\rho_{zt}) : p_{zt} = \sum_{i\in I} y_{izt} + s_{zt} - D_{zt}, z \in Z, t \in T \tag{6c}$$

$$p_{:t} \in \mathcal{P}^{\mathrm{PA}}, t \in T \tag{6d}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{6e}$$

It should be noted that 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[3]. 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 (6). We note that this does not mean that internal congestion is omitted. As one can observe in equation (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 CREG (2017), 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 4 indeed holds. But one can also mention other definitions of the copper plate assumption that do not lead to the equivalence: Van den Bergh and Delarue (2016) defines it as "ignoring transmission constraints within a zone", and Hary (2018) 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 1, we clarify the conditions for an equivalence between the two pricing models based on physical quantities.

Proposition 1. Let us define the zonal network as the network obtained by aggregating the nodes of a zone into a single node 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 (2) and the zonal model (6) are equivalent.

[3] We elaborate more on the links between the European institutional setting and the market design choices in section 6.

Proof. We prove this statement formally in the end of Appendix 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 vs zonal pricing in the US (Hogan, 1998).

The reasoning on the decentralization of the solution extends easily to the case of zonal pricing using model (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

$$0 \leq x_{iz} \perp IC_i - \sum_{t \in T} \mu_{izt} \geq 0 \ \ \forall i \in I, z \in Z \tag{7}$$

holds from the KKT conditions. The main difference between model (2) and model (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 an important property of zonal markets: the dispatch obtained after clearing the market does not in general respect the actual nodal network constraints and the market clearing must be followed by an out-of-market procedure for restoring the feasibility of the dispatch. This procedure is called *re-dispatch* and is undertaken in Europe by TSOs. 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 is a market-based re-dispatch (Hirth and Schlecht, 2018). 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 (Grimm et al., 2018; Hirth et al., 2019). 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 (European Parliament, Council of the European Union, 2019), 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 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. This is thus a simplification regarding current practice in the Netherlands. We also mention that modeling and solving the model on a large scale in the case of market-based re-dispatch would be more challenging for two reasons: (i) both the spot market and the re-dispatch influence the payoff (and thus the investment decisions) of the market participants and (ii) even agents with no market power have an incentive to deviate from truthful bidding, because of inc-dec gaming opportunities (Hirth et al., 2019). In theory, if generators are completely flexible and there are no unit-commitment decisions made based on 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[4]. In practice, however, a loss of welfare is associated to zonal unit commitment (Aravena et al., 2021).

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.[5]

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:

$$\min_{x,y,s,r,f} \sum_{i\in I,n\in N,t\in T} MC_i \cdot y_{int} + \sum_{n\in N,t\in T} VOLL \cdot s_{nt} \tag{8a}$$

$$\sum_{n\in N(z)} x_{in} = \bar{x}_{iz}, \ \ \forall z \in Z \tag{8b}$$

$$y_{int} \leq x_{in} + X_{in}, i \in I, n \in N, t \in T \tag{8c}$$

$$r_{nt} = \sum_{i\in I} y_{int} + s_{nt} - D_{nt}, n \in N, t \in T \tag{8d}$$

$$r_{:t} \in \mathcal{R}, t \in T \tag{8e}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{8f}$$

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

[4]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 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 paper).

[5]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 (Adib and Zarnikau, 2006).

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 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 $\mathcal{P}$. In Europe, two approaches are currently employed for market coupling: ATC-based market coupling (ATCMC) and FBMC. The idea of ATCMC is to impose constraints on the maximum power that can be exchanged between each pair of zones. ATCMC is currently used in the intra-day market throughout the continent and in the day-ahead market in part of it. In 2015, the countries of the CWE network area transitioned to using the FBMC approach for day-ahead market clearing (50Hertz et al., 2018). The idea is to define the feasible set of net injections as a poyhedral set by approximating the expected flows on inter-connectors. FBMC can be interpreted as a generalization of ATCMC in that it allows for representing more complicated relationships between zonal net positions and is therefore expected to perform better. Viewed from the perspective of short-term market operations, FBMC is thus more efficient than the pure zonal market based on $\mathcal{P}^{\text{PA}}$. From a long-term perspective, however, FBMC raises certain concerns that we shall discuss in detail in the next section.

4. Flow-based market coupling in the context of capacity expansion

As we have just mentioned, FBMC deviates from the Price Aggregation (PA) model by introducing a set of rules in order to approximate the expected flows on inter-connectors. The goal is to have a feasible set of zonal net positions $\mathcal{P}^{\text{FBMC}}$ that is tighter than $\mathcal{P}^{\text{PA}}$ thereby resulting in a reduced re-dispatch cost. The expected flows on inter-connectors are computed using a forecast of the demand and a knowledge of the capacity installed at every node of the network. In practice, however, this methodology is complicated to model because it relies on a set of exogenous parameters defined by TSOs. Each TSO follows a different definition for these parameters, and it has been shown that the market outcome is sensitive to these definitions (Marien et al., 2013). For these reasons, Aravena et al. (2021) proposes a methodology for modeling FBMC that does not rely on the definitions of these exogenous parameters but, instead, is based on the principles stated in EU regulations on the internal market for electricity (European Parliament, Council of the European Union, 2009; European Commision, 2015). This modeling framework is particularly interesting for our present study as it highlights the important property of FBMC related to capacity expansion: the dependence of $\mathcal{P}^{\text{FBMC}}$ on a forecast of the demand and on the *installed capacity* at every node. Aravena et al. (2021) studied the short-term efficiency of FBMC using the newly introduced polytope of feasible zonal net positions $\mathcal{P}^{\text{FBMC}}$. In this section, we will integrate $\mathcal{P}^{\text{FBMC}}$ into a capacity expansion model in order to focus on the efficiency of FBMC from a long-term perspective.

We start by describing Aravena's model of the network constraints in FBMC in section 4.1. We then model respectively the capacity expansion model of a central planner under FBMC and its decentralized version in sections 4.2 and 4.3, as well as their respective results on the small illustrative example introduced in the previous section.

4.1. Network constraints in FBMC

Using Aravena's model (Aravena et al., 2021), the network constraints in FBMC can be written as follows:

$$
\begin{aligned}
\mathcal{P}^{\text{FBMC}} = \Big\{ p \in \mathbb{R}^{|Z|} \,\Big|\, & \exists (f, r, \tilde{y}) \in \mathbb{R}^{|K|} \times \mathbb{R}^{|N|} \times \mathbb{R}^{|I||N|} : p_z = \sum_{n \in N(z)} r_n \;\; \forall z \in Z, \\
& r_n = \tilde{y}_{int} - D_{nt} \;\; \forall n \in N,\, 0 \leq \tilde{y}_{int} \leq X_{in} \;\; \forall i \in I, n \in N, \\
& f_k = \sum_n PTDF_{kn} \cdot r_n \;\; \forall k \in K,\, \sum_n r_n = 0, -TC_k \leq f_k \leq TC_k, \;\; \forall k \in k \Big\}
\end{aligned}
\tag{9}
$$

In this set of equations, variables $\tilde{y}_{int}$ can be understood as an auxiliary nodal dispatch. By introducing $\mathcal{P}^{\text{FBMC}}$, 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 paper, 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 $\mathcal{P}^{\text{FBMC}}$. We denote this extended set by $\mathcal{PX}^{\text{FBMC}}$ that is now defined on the space of zonal net positions p *and* zonal investment x:

$$
\begin{aligned}
\mathcal{PX}^{\text{FBMC}} = \Big\{ p \in \mathbb{R}^{|Z|}, x \in \mathbb{R}^{|I||Z|} \,\Big|\, & \exists (f, r, \tilde{y}, \tilde{x}) \in \mathbb{R}^{|L|} \times \mathbb{R}^{|N|} \times \mathbb{R}^{|I||N|} \times \mathbb{R}_+^{|I||N|} : \\
& p_z = \sum_{n \in N(z)} r_n \;\; \forall z \in Z, x_{iz} = \sum_{n \in N(z)} \tilde{x}_{in} \;\; \forall i \in I, z \in Z, \\
& r_n = \tilde{y}_{in} - D_n \;\; \forall n \in N,\, 0 \leq \tilde{y}_{in} \leq X_{in} + \tilde{x}_{in} \;\; \forall i \in I, n \in N, \\
& f_k = \sum_n PTDF_{kn} \cdot r_n \;\; \forall k \in K,\, \sum_n r_n = 0, -TC_k \leq f_k \leq TC_k, \;\; \forall k \in K \Big\}
\end{aligned}
\tag{10}
$$

Note that the set $\mathcal{PX}^{\text{FBMC}}$ 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 4.3 and in Appendix 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 constraints 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

$$
(p, x) \in \mathcal{PX}^{\text{FBMC}} \Leftrightarrow \sum_{z \in Z} V_{mz} p_z + \sum_{i \in I, z \in Z} U_{miz} x_{iz} + W_m \geq 0 \;\; \forall m \in \{1, ..., M\}
\tag{11}
$$

4.2. Centralized capacity expansion under FBMC

Using set $\mathcal{PX}^{\text{FBMC}}$ and its expression as linear constraints defined in (11), one can easily define the capacity expansion problem from the central planner's perspective:

$$\min_{x,y,s,p} \sum_{i\in I,z\in Z} IC_i \cdot x_{iz} + \sum_{i\in I,z\in N,t\in T} MC_i \cdot y_{izt} + \sum_{n\in N,t\in T} VOLL \cdot s_{zt} \tag{12a}$$

$$(\mu_{izt}) : y_{izt} \leq x_{iz} + X_{iz}, i \in I, z \in Z, t \in T \tag{12b}$$

$$(\rho_{zt}) : p_{zt} = \sum_{i\in I} y_{izt} + s_{zt} - D_{zt}, z \in Z, t \in T \tag{12c}$$

$$(\gamma_m) : \sum_{z\in Z} V_{mz} p_z + \sum_{i\in I,z\in Z} U_{miz} x_{iz} + W_m \geq 0, \ \ \forall m \in \{1, ..., M\}, t \in T \tag{12d}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{12e}$$

Problem (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 (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:

$$0 \leq x_{iz} \perp IC_i - \sum_{t\in T} \mu_{izt} - \sum_{m\in\{1,...,M\}} U_{miz}\gamma_m \geq 0 \ \ \forall i \in I, z \in Z \tag{13}$$

One observes an important difference between condition (13) and the condition under the PA zonal pricing model (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 (12) cannot be readily decentralized using zonal prices ρ_{zt} associated to constraint (12c). Energy-only markets under FBMC are thus incomplete and network constraints must be 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

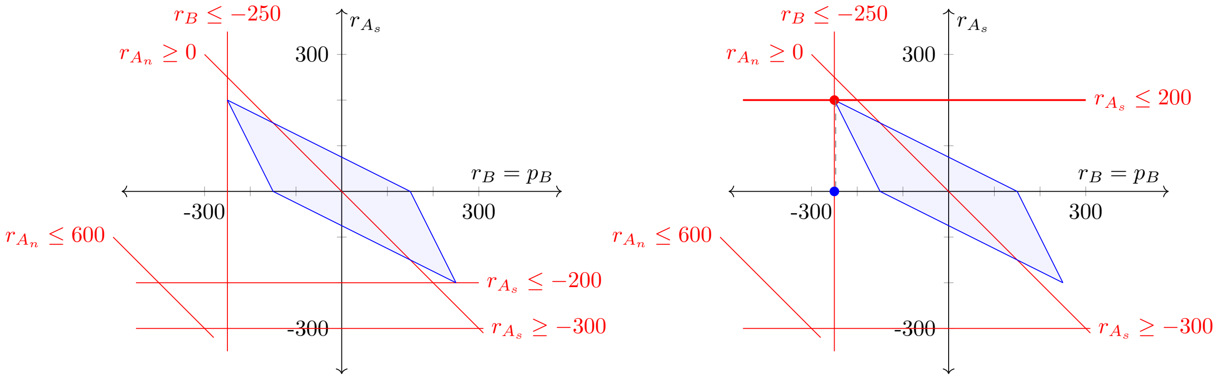


Figure 4: Representation of the flow-based constraints on the space of r_{A_s} and r_B. The PA polytope is shown in blue. The additional constraints imposed in FBMC are shown in red with no oil capacity invested (left) and 400MW of oil capacity invested in node A_s (right). One observes that the problem 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).

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 $\mathcal{PX}^{\text{FBMC}}$ on the space (r_{A_s}, r_B) in Figure 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 (left panel), 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. 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 (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 $\mathcal{PX}^{\text{FBMC}}$ 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 $\mathcal{PX}^{\text{FBMC}}$, not restrict it. For that reason, the term in γ_m is always positive and can thus be interpreted as a subsidy.

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 (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 (12). This corresponds to replacing equation (13) by (7) in its KKT conditions. One obtains a new complementarity problem. This problem is square 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 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 modifies 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 incomplete market. The coordination between generation and transmission achieved by pricing the network constraints (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 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 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 5. We insist on the fact that the difficulty discussed here regarding the transition from an 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 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.

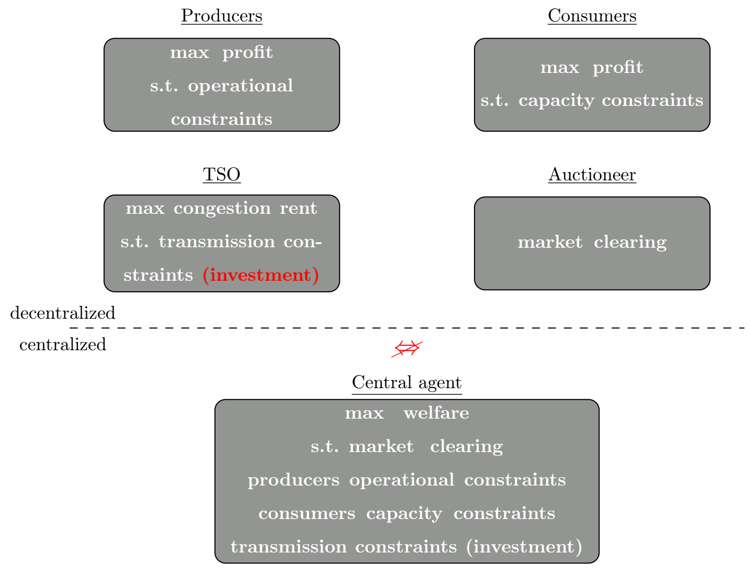


Figure 5: Schematic representation of the decentralized game and the problem of the central planner. In the case of FBMC, the feasible set of the TSO depends on generation investment, which corresponds to decision variables of the producers. The appropriate solution concept of the game is thus the GNE and the equivalence between the centralized and decentralized formulations is broken.

The GNE structure also raises questions regarding existence and unicity. 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 (Bolton and Clausen, 2019). 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 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

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 1: Summary of the results of the different policies on the illustrative example. PA-NR is the PA model presented in section 3.2, with network reserve. FBMC-C is the centralized FBMC model of section 4.2 and FBMC-D is its decentralized version (section 4.3).

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. 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 1.

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 (Bezanson et al., 2017) using JuMP v0.21.5 (Dunning et al., 2017) and solved with Gurobi

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

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 F for the details of our solution methodology.

5.1. Dataset

Our starting point is the dataset used in previous work by the authors (Aravena et al., 2021). The network data is an updated version of the European grid model of Hutcheon and Bialek (2013). The generation data is obtained from Open Power System Data (2020). Table 2 presents the total installed capacity per generator type for the entire CWE region.

The network model of Hutcheon and Bialek (2013) 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. 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 E.

The models that we use for the case study are generalized versions of the models presented in sections 3 and 4. In the models of the CWE case study, we also consider revenues from reserve provision, where reserve is assumed to be cleared simulatenously with energy. We also consider fixed operating and maintenance costs. Units that cannot cover their fixed costs are decommissioned. We assume that investment is possible in 3 different technologies, similarly to Ambrosius et al. (2020): CCGT units, OCGT units and Combined Heat and Power CCGT units. We

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

use the same cost data as Ambrosius et al. (2020), which are presented in Table 3.

Wind and solar expansion are accounted for in an exogenous way. The fixed and marginal cost of existing capacity are also sourced from Ambrosius et al. (2020) and completed from Open Energy Information (2021) when missing. We separate each existing unit into open-cycle and combined-cycle generators. For combined-cycle units, we increase the fixed cost by 77% and reduce the marginal cost range by 39%, similarly to what is assumed for natural gas units in investment. We alse 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 (Open Power System Data, 2020). 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 (International Atomic Energy Agency, 2020). F 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 (Ahlqvist et al., 2019; Antonopoulos et al., 2020). This can be related to the proposal of Hogan for contract networks (Hogan, 1992, 1999), 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 Hogan (1999). 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

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 5: Performance comparison of the different policies.

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 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, hypothetical. 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 (Neuhoff and Boyd, 2011; Duane, 2019).

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 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 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 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 paper. Comparing the nodal policy with FBMC-C, one observes that nodal pricing 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

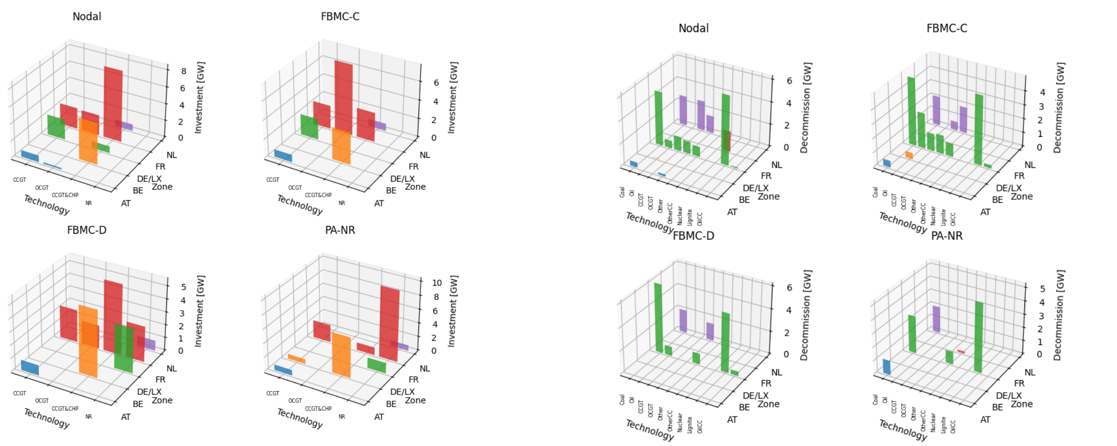


Figure 6: Total investment (left) and decommissioning (right) per zone and per technology for each policy.

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).

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 energy mix, ii) the decommissioning behavior of the different policies, and iii) welfare re-allocation.

5.3.1. Energy mix

In Figure 6, we present the total investment and decommissioning per zone and per technology for each policy. We have already observed that the nodal policy leads to more decommissioning. In this figure, 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 7 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 CREG (2019) 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.

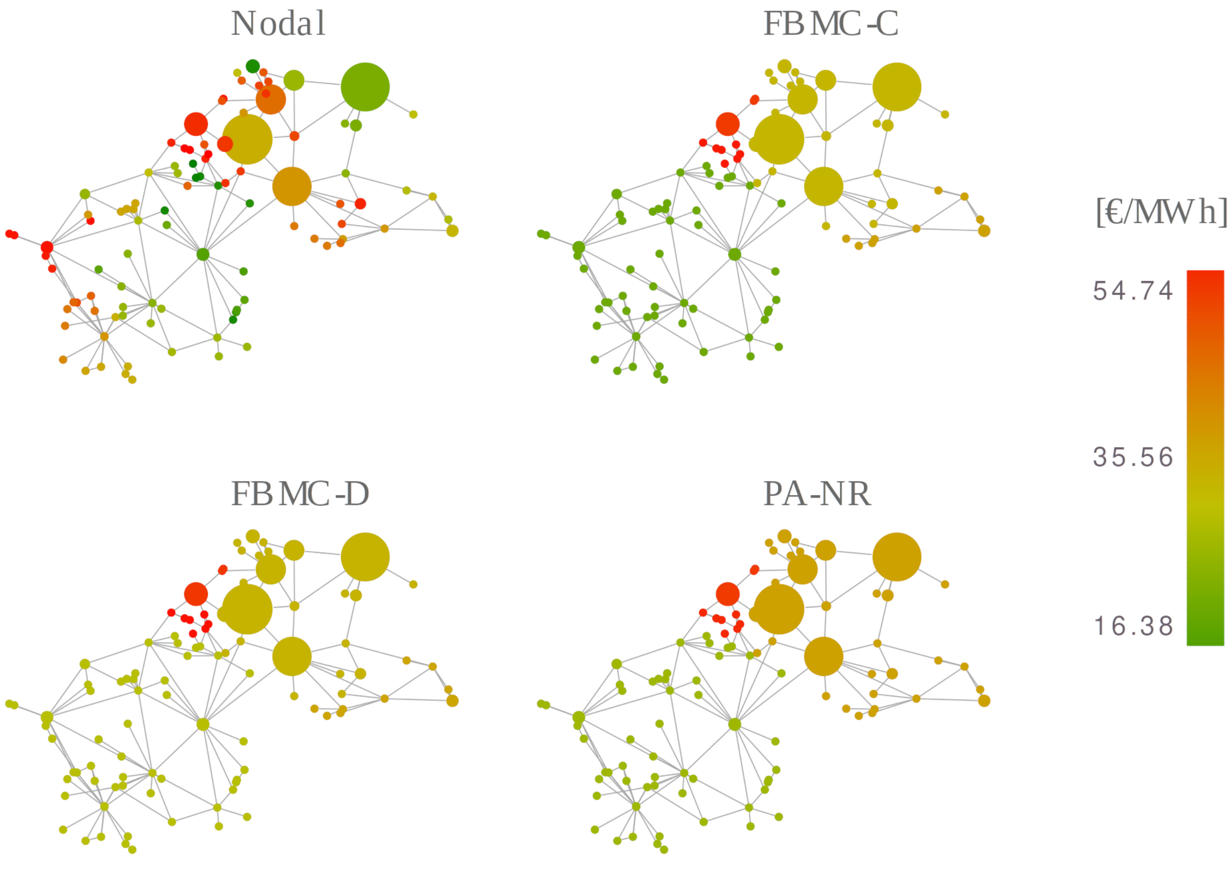


Figure 7: Locational distribution of the average price under the four different policies. The size of the nodes is proportional to the amount of existing coal and lignite at these locations.

Figure 8: 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.

5.3.2. Decommissioning behavior

Another interesting observation that can be based on Figure 6 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, 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 (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.

5.3.3. Welfare re-allocation

Finally, we wish to stress that, while each policy leads to a different total welfare as discussed in section 5.2, the allocation of welfare is also not homogeneous between the different agents. Figure 7 already highlights the important welfare re-allocation between the different locations. In Figure 8, 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 co-exist 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 (ENTSO-E, 2018) and among industrial and local consumers (Mäntysaari, 2015). 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 paper, 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, as discussed in Eicke et al. (2020).

6. Discussion

In this paper, we have described a new type of inefficiency that arises in zonal systems in the context of capacity expansion. It is now largely recognized that zonal pricing results in various types of inefficiencies regarding congestion management and the reader may wonder why, knowing these inefficiencies, Europeans do not switch to nodal. In the US, several markets started with a zonal design but quickly moved to a nodal one due to market failures. The PJM zonal market collapsed after only a year, due to dispatching difficulties. It took a bit more time in California, but the market eventually collapsed due to inc-dec gaming. As things have not unraveled at such pace in the EU, there has been more investment in the zonal model. The European Commission appears to be rather favorable towards Member States adding zones, as witnessed for example by the CACM Regulation which instructs ACER to proceed with a bidding zone review every three years. Several EU countries consist of several zones, and the configuration of bidding zone is recurrently debated. The most recent bidding zone reconfiguration took place in 2018 with the split of the German-Austrian zone, following the first bidding zone review. Poland is even going further by investigating the possibility to directly move to a nodal design. On the other hand, congestion management is not the only aspect that needs to be taken into account. It is sometimes claimed in the literature through non-quantitative arguments that zonal pricing may also influence liquidity and the exercise of market power (Ahlqvist et al., 2019; Antonopoulos et al., 2020).

In this section, we suggest that one reason why Europeans stick to zonal 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 paper and the expertise of the authors. 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. Hancher (1998) 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 (Boiteux and Stasi, 1952). This was in 1952, well before the ground-breaking work of Schweppe et al. (1988). These economic concepts were also operational in some of the (published) EDF computational models (Dodu and Merlin, 1979). 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[6].

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 substraction. (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 treated jointly (the move to implicit auctions[7]). (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 (in Central Western Europe). (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 43 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.

7. Conclusion

The capacity expansion problem is a key analytical tool in an era of energy transition. In this paper, 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

[6]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 Amundsen and Bergman (2006)). 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.

[7]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/`

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 discussed in this paper: 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 energy 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 make 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.

This paper focuses on identifying and assessing the missing money problems associated to zonal pricing that lead to inefficient investment in the long run. Future research will focus on remedies to these missing money problems and will investigate the extent to which additional market instruments (capacity markets, market-based re-dispatch, locational transmission tariffs) can restore the theoretical efficiency of investment under zonal pricing.

Acknowledgements

Quentin Lété is a Research Fellow of the Fonds de la Recherche Scientifique - FNRS. The work has received funding from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No. 850540). The authors thank Alain Marien and Clara Verhelst (CREG) as well as Ross Baldick for their helpful comments and discussions during the development of this work.

References

50Hertz , Amprion , APG , Creos , Elia , EPEXSpot , RTE , TenneT , Transnet BW . Documentation of the CWE FB MC solution. 2018.

Adib P, Zarnikau J. Chapter 11 - texas: The most robust competitive market in north america1. In: Sioshansi FP, Pfaffenberger W, editors. Electricity Market Reform. Oxford: Elsevier; Elsevier Global Energy Policy and Economics Series; 2006. p. 383–417. URL: `https://www.sciencedirect.com/science/article/pii/B9780080450308500138`. doi:`https://doi.org/10.1016/B978-008045030-8/50013-8`.

Ahlqvist V, Holmberg P, Tangerås T. Central- versus Self-Dispatch in Electricity Markets. Working Papers EPRG 1902; Energy Policy Research Group, Cambridge Judge Business School, University of Cambridge; 2019. URL: https://ideas.repec.org/p/enp/wpaper/eprg1902.html.

Ambrosius M, Grimm V, Kleinert T, Liers F, Schmidt M, Zöttl G. Endogenous price zones and investment incentives in electricity markets: An application of multilevel optimization with graph partitioning. Energy Economics 2020;92:104879. URL: https://www.sciencedirect.com/science/article/pii/S014098832030219X. doi:https://doi.org/10.1016/j.eneco.2020.104879.

Amundsen ES, Bergman L. Why has the Nordic electricity market worked so well? Utilities Policy 2006;14(3):148–57. URL: https://ideas.repec.org/a/eee/juipol/v14y2006i3p148-157.html.

Antonopoulos G, Vitiello S, Fulli G, Masera M. Nodal pricing in the European Internal Electricity Market. Technical Report; 2020. doi:10.2760/41018.

Aravena I, Lété Q, Papavasiliou A, Smeers Y. Transmission capacity allocation in zonal electricity markets. Operations Research 2021;69(4):1240–55. doi:10.1287/opre.2020.2082.

Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. SIAM review 2017;59(1):65–98. URL: https://doi.org/10.1137/141000671.

Bjørndal M, Jornsten K. Zonal Pricing in a Deregulated Electricity Market. The Energy Journal 2001;0(1):51–73. URL: http://www.jstor.org/stable/41322907.

Boiteux M. Peak-Load Pricing. The Journal of Business 1960;33(2):157–79.

Boiteux M. 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.

Boiteux M, Stasi P. 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.

Bolton R, Clausen T. Ambiguous interventions: Capacity remuneration mechanisms and the politics of energy transition in britain and germany. Energy Research & Social Science 2019;56:101218. URL: https://www.sciencedirect.com/science/article/pii/S2214629618311654. doi:https://doi.org/10.1016/j.erss.2019.101218.

Cramton P, Stoft S. A capacity market that makes sense. The Electricity Journal 2005;18(7):43–54. URL: https://www.sciencedirect.com/science/article/pii/S1040619005000850. doi:https://doi.org/10.1016/j.tej.2005.07.003.

CREG . Functioning and design of the Central West European day-ahead flow based market coupling for electricity: Impact of TSOs Discretionary Actions. Technical Report; 2017. Available online: https://www.creg.be/sites/default/files/assets/Publications/Studies/F1687EN.pdf. Accessed: April 26 2021.

CREG . Study on the best forecast of remedial actions to mitigate market distortion. Technical Report; 2019. Available online: https://www.creg.be/sites/default/files/assets/Publications/Studies/F1987EN.pdf. Accessed: April 13 2021.

Dodu J, Merlin A. Recent improvements of the mexico model for probabilistic planning studies. International Journal of Electrical Power & Energy Systems 1979;1(1):46–56. doi:https://doi.org/10.1016/0142-0615(79)90031-0.

Duane V. Nodal vs zonal pricing for the electricity market. 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.

Dunning I, Huchette J, Lubin M. Jump: A modeling language for mathematical optimization. SIAM Review 2017;59(2):295–320. doi:10.1137/15M1020575.

Egerer J, Grimm V, Kleinert T, Schmidt M, Zöttl G. The impact of neighboring markets on renewable locations, transmission expansion, and generation investment. European Journal of Operational Research 2021;292(2):696–713. URL: https://www.sciencedirect.com/science/article/pii/S0377221720309486. doi:https://doi.org/10.1016/j.ejor.2020.10.055.

Ehrenmann A, Smeers Y. Inefficiencies in European congestion management proposals. Utilities Policy 2005;13(2):135 –52. doi:10.1016/j.jup.2004.12.007.

Ehrenmann A, Smeers Y. Generation capacity expansion in a risky environment: A stochastic equilibrium analysis. Operations Research 2011;59(6):1332–46. doi:10.1287/opre.1110.0992.

Eicke A, Khanna T, Hirth L. Locational investment signals: How to steer the siting of new generation capacity in power systems? Energy Journal 2020;41(1):281–304. doi:10.5547/01956574.41.6.aeic.

ENTSO-E . Overview of Transmission Tariffs in Europe: Synthesis 2018. Technical Report; 2018. Available online: https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/mc-documents/TTO_Synthesis_2018.pdf. Accessed: April 13 2021.

European Commision . Commision Regulation (EU) 2015/1222 of 24 july 2015 establishing a guideline on capacity allocation and congestion management. 2015.

European Parliament, Council of the European Union . Regulation (EU) 2019/943 of the European Parliament and if the Council of 13 july 2009 on conditions for access to the network for cross-border exchanges in electricity. 2009.

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. 2019.

Finon D, Pignon V. Electricity and long-term capacity adequacy: The quest for regulatory mechanism compatible with electricity market. Utilities Policy 2008;16(3):143–58. URL: https://www.sciencedirect.com/science/

article/pii/S0957178708000039. doi:https://doi.org/10.1016/j.jup.2008.01.002; capacity Mechanisms in Imperfect Electricity Markets.

Fraunholz C, Hladik D, Keles D, Möst D, Fichtner W. On the long-term efficiency of market splitting in Germany. Energy Policy 2021;149(February 2020). doi:10.1016/j.enpol.2020.111833.

Grimm V, Kleinert T, Liers F, Schmidt M, Zöttl G. Optimal price zones of electricity markets: a mixed-integer multilevel model and global solution approaches. Optimization Methods and Software 2019;34(2):406–36. URL: https://doi.org/10.1080/10556788.2017.1401069. doi:10.1080/10556788.2017.1401069. arXiv:https://doi.org/10.1080/10556788.2017.1401069.

Grimm V, Martin A, Schmidt M, Weibelzahl M, Zöttl G. Transmission and generation investment in electricity markets: The effects of market splitting and network fee regimes. European Journal of Operational Research 2016a;254(2):493–509. doi:10.1016/j.ejor.2016.03.044.

Grimm V, Martin A, Sölch C, Weibelzahl M, Zöttl G. Market-Based Redispatch May Result in Inefficient Dispatch. SSRN Electronic Journal 2018;(2008):1–27. doi:10.2139/ssrn.3120403.

Grimm V, Martin A, Weibelzahl M, Zöttl G. On the long run effects of market splitting: Why more price zones might decrease welfare. Energy Policy 2016b;94:453–67. doi:10.1016/j.enpol.2015.11.010.

Grimm V, Rückel B, Sölch C, Zöttl G. The impact of market design on transmission and generation investment in electricity markets. Energy Economics 2021;93:104934. URL: https://www.sciencedirect.com/science/article/pii/S0140988320302747. doi:https://doi.org/10.1016/j.eneco.2020.104934.

Hancher L. Delimitation of energy law jurisdiction: the eu and its member states: From organisational to regulatory conflicts. Journal of Energy & Natural Resources Law 1998;16(1):42–67. URL: https://doi.org/10.1080/02646811.1998.11433126. doi:10.1080/02646811.1998.11433126. arXiv:https://doi.org/10.1080/02646811.1998.11433126.

Hary N. 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.

Hirth L, Schlecht I. Market-Based Redispatch in Zonal Electricity Markets. SSRN Electronic Journal 2018;(055). doi:10.2139/ssrn.3286798.

Hirth L, Schlecht I, Maurer C, Tersteegen B. Cost or marketbased? Future redispatch procurement in Germany. Technical Report; 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.

Hogan W. Competitive electricity market design: A wholesale primer. 1998. Available online: https://scholar.harvard.edu/whogan/files/empr1298.pdf. Accessed 26 April 2021.

Hogan WW. Restructuring the Electricity Market : Institutions for Network Systems. Technical Report April; 1999. URL: https://scholar.harvard.edu/whogan/files/hjp0499.pdf.

Hogan WW. Electricity scarcity pricing through operating reserves. Economics of Energy & Environmental Policy 2013;Volume 2(Number 2). URL: https://EconPapers.repec.org/RePEc:aen:eeepjl:2_2_a04.

Hutcheon N, Bialek JW. Updated and validated power flow model of the main continental european transmission network. In: 2013 IEEE Grenoble Conference. 2013. p. 1–5.

International Atomic Energy Agency . Country nuclear power profiles - france. 2020. Available online: https://cnpp.iaea.org/countryprofiles/France/France.htm. Accessed 03/03/21.

Joskow P. The new energy paradigm. 2007. Oxford University Press, Ch. Competitive Electricity Markets and Investment in New Generation Capacities, pp. 76–122.

Joskow PL. Competitive Electricity Markets and Investment in New Generating Capacity. AEI-Brookings Joint Center Working Paper No 06-14 2006;doi:10.2139/ssrn.902005.

Mäntysaari P. Transmission Marketplaces; Cham: Springer International Publishing. p. 347–417. URL: https://doi.org/10.1007/978-3-319-16513-4_5. doi:10.1007/978-3-319-16513-4_5.

Marien A, Luickx P, Tirez A, Woitrin D. Importance of design parameters on flowbased market coupling implementation. In: 2013 10th International Conference on the European Energy Market (EEM). 2013. p. 1–8. doi:10.1109/EEM.2013.6607298.

Milstein I, Tishler A. On the effects of capacity payments in competitive electricity markets: Capacity adequacy, price cap, and reliability. Energy Policy 2019;129:370–85. URL: https://www.sciencedirect.com/science/article/pii/S0301421519301119. doi:https://doi.org/10.1016/j.enpol.2019.02.028.

Neuhoff K, Boyd R. International Experiences of Nodal Pricing Implementation - Frequently Asked Questions. 2011.

Newbery D. Missing money and missing markets: Reliability, capacity auctions and interconnectors. Energy Policy 2016;94:401–10. URL: https://www.sciencedirect.com/science/article/pii/S0301421515301555. doi:https://doi.org/10.1016/j.enpol.2015.10.028.

Open Energy Information . Transparent cost database. 2021. Available online: https://openei.org/apps/TCDB. Accessed 03/03/21.

Open Power System Data . Data package conventional power plants. 2020. doi:https://doi.org/10.25832/conventional_power_plants/2020-10-01; version 2020-10-01.

Ozdemir O. Simulation modeling and optimization of competitive electricity markets and stochastic fluid systems. Ph.D. thesis; Tilburg University; 2013.

Schweppe FC, Caramanis MC, Tabors RD, Bohn RE. Spot Pricing of Electricity. Springer US, 1988. URL: https://doi.org/10.1007/978-1-4613-1683-1. doi:10.1007/978-1-4613-1683-1.

Van den Bergh K, Boury J, Delarue E. The flow-based market coupling in central western europe: Concepts and definitions. The Electricity Journal 2016;29(1):24–9. URL: https://www.sciencedirect.com/science/article/pii/S104061901530004X. doi:https://doi.org/10.1016/j.tej.2015.12.004.

Van den Bergh K, Delarue E. An improved method to calculate injection shift keys. Electric Power Systems Research 2016;134:197–204. URL: https://www.sciencedirect.com/science/article/pii/S0378779616300062. doi:https://doi.org/10.1016/j.epsr.2016.01.020.

Weibelzahl M. Nodal, zonal, or uniform electricity pricing: how to deal with network congestion. Frontiers in Energy 2017;11(2):210–32. doi:10.1007/s11708-017-0460-z.

Appendix 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 Hogan (1992) and formally described in Boucher and Smeers (2001). We repeat these results here using the same notation as in the present paper.

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 $i \in I, n \in N$

$$\begin{aligned}
&\max_{x_{in}} \sum_{t \in T} \Big((\rho_{nt} - MC_i) y_{int} \Big) - IC_i x_{in} \\
&(\mu_{int}) : X_{in} + x_{in} - y_{int} \geq 0 \\
&x_{in} \geq 0, y_{int} \geq 0
\end{aligned} \tag{A.1}$$

KKT system:

$$\begin{aligned}
&0 \leq x_{in} \perp IC_i - \sum_{t \in T} \mu_{int} \geq 0, i \in I, n \in N \\
&0 \leq y_{int} \perp MC_i + \mu_{int} - \rho_{nt} \geq 0, i \in I, n \in N, t \in T \\
&0 \leq \mu_{int} \perp X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T
\end{aligned} \tag{A.2}$$

Consumers. For each $n \in N$

$$\begin{aligned}
&\max_{s_{nt}} \sum_{t \in T} (VOLL - \rho_{nt})(D_{nt} - s_{nt}) \\
&\text{s.t. } (\delta_{nt}) : D_{nt} - s_{nt} \geq 0, t \in T \\
&s_{nt} \geq 0
\end{aligned} \tag{A.3}$$

KKT system:

$$\begin{aligned}
&0 \leq s_{nt} \perp VOLL - \rho_{nt} + \delta_{nt} \geq 0, n \in N, t \in T \\
&0 \leq \delta_{nt} \perp D_{nt} - s_{nt} \geq 0, n \in N, t \in T
\end{aligned} \tag{A.4}$$

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

$$\begin{aligned}
&\max_{r_{nt}} - \sum_{n \in N, t \in T} r_{nt} \rho_{nt} \\
&(\psi_{kt}) : f_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
&(\phi_t) : \sum_{n \in N} r_{nt} = 0, t \in T \\
&(\lambda_{kt}^-, \lambda_{kt}^+) : -TC_k \leq f_{kt} \leq TC_k, k \in K, t \in T
\end{aligned} \tag{A.5}$$

KKT system:

$$
\begin{aligned}
&r_{nt} \text{ free } \perp \rho_{nt} + \phi_t - \sum_{k \in K} PTDF_{kn} \cdot \psi_{kt} = 0 \\
&f_{kt} \text{ free } \perp \psi_{kt} + \lambda^+_{kt} - \lambda^-_{kt} = 0 \\
&\psi_{kt} \text{ free } \perp f_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
&\phi_t \text{ free } \perp \sum_{n \in N} r_{nt} = 0, t \in T \\
&0 \leq \lambda^+_{kt} \perp TC_k - f_{kt} \geq 0, k \in K, t \in T \\
&0 \leq \lambda^-_{kt} \perp TC_k + f_{kt} \geq 0, k \in K, t \in T
\end{aligned}
\tag{A.6}
$$

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

$$
\max_{\rho_{nt}} \rho_{nt}(r_{nt} + D_{nt} - \sum_i y_{int} - s_{zt}) \tag{A.7}
$$

KKT system:

$$
\rho_{nt} \text{ free } \perp r_{nt} + D_{nt} - \sum_i y_{int} - s_{nt} = 0, n \in N, t \in T \tag{A.8}
$$

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):

$$
\begin{aligned}
&0 \leq x_{in} \perp IC_i - \sum_{t \in T} \mu_{int} \geq 0, i \in I, n \in N \\
&0 \leq y_{int} \perp MC_i + \mu_{int} - \rho_{nt} \geq 0, i \in I, n \in N, t \in T \\
&0 \leq \mu_{int} \perp X_{in} + x_{in} - y_{int} \geq 0, i \in I, n \in N, t \in T \\
&0 \leq s_{nt} \perp VOLL - \rho_{nt} + \delta_{nt} \geq 0, n \in N, t \in T \\
&0 \leq \delta_{nt} \perp D_{nt} - s_{nt} \geq 0, n \in N, t \in T \\
&r_{nt} \text{ free } \perp \rho_{nt} + \phi_t - \sum_{k \in K} PTDF_{kn} \cdot \psi_{kt} = 0 \\
&f_{kt} \text{ free } \perp \psi_{kt} + \lambda^+_{kt} - \lambda^-_{kt} = 0 \\
&\psi_{kt} \text{ free } \perp f_{kt} - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
&\phi_t \text{ free } \perp \sum_{n \in N} r_{nt} = 0, t \in T \\
&0 \leq \lambda^+_{kt} \perp TC_k - f_{kt} \geq 0, k \in K, t \in T \\
&0 \leq \lambda^-_{kt} \perp TC_k + f_{kt} \geq 0, k \in K, t \in T \\
&\rho_{nt} \text{ free } \perp r_{nt} + D_{nt} - \sum_i y_{int} - s_{nt} = 0, n \in N, t \in T
\end{aligned}
\tag{A.9}
$$

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 (2). This shows the equivalence between the centralized and decentralized capacity expansion problem in nodal pricing.

Appendix 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 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 $i \in I, z \in Z$

$$\begin{aligned} \max_{x_{iz}} \sum_{t \in T} \Big((\rho_{zt} - MC_i) y_{izt} \Big) - IC_i x_{iz} \\ (\mu_{izt}) : X_{iz} + x_{iz} - y_{izt} \geq 0 \\ x_{iz} \geq 0, y_{izt} \geq 0 \end{aligned} \tag{B.1}$$

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

Consumers. For each $z \in Z$

$$\begin{aligned} \max_{s_{zt}} \sum_{t \in T} (VOLL - \rho_{zt})(D_{zt} - s_{zt}) \\ \text{s.t. } (\delta_{zt}) : D_{zt} - s_{zt} \geq 0, t \in T \\ s_{zt} \geq 0 \end{aligned} \tag{B.2}$$

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.

$$\begin{aligned} \max_{p_{zt}} - \sum_{z \in Z, t \in T} p_{zt} \rho_{zt} \\ (\nu_z) : \sum_{i \in I} x_{iz} - \sum_{n \in N(z)} \tilde{x}_n = 0, z \in Z \\ (\tilde{\rho}_{zt}) : p_{zt} - \sum_{n \in N(z)} \tilde{r}_{nt} = 0, z \in Z, t \in T \\ (\tilde{\rho}_{nt}) : \tilde{r}_{nt} - \tilde{y}_{nt} + D_{nt} = 0, n \in N, t \in T \\ (\tilde{\mu}_{nt}) : \sum_{i \in I} X_{in} + \tilde{x}_n - \tilde{y}_{nt} \geq 0, n \in N, t \in T \\ (\psi_{kt}) : \tilde{f}_{kt} - \sum_{n} PTDF_{nk} \cdot \tilde{r}_{nt} = 0, k \in K, t \in T \\ (\phi_t) : \sum_{n} \tilde{r}_{nt} = 0, t \in T \\ (\lambda^-_{kt}, \lambda^+_{kt}) : -TC_k \leq \tilde{f}_{kt} \leq TC_k, k \in K, t \in T \end{aligned} \tag{B.3}$$

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 $n \in N, t \in T$

$$\max_{\rho_{nt}} \ \rho_{nt}(r_{nt} + D_{nt} - \sum_i y_{int} - s_{zt}) \tag{B.4}$$

Here, we are not in the classical setting of a Nash equilibrium because the set of possible decisions of the TSO depends on the decision of another agent, namely the investment x_{iz} of the producers. We are thus in the setting of a GNE (Harker, 1991), which corresponds to the joint solution of the KKT conditions of all agents. It can thus be formulated as the following MCP:

$$\begin{aligned}
&0 \le x_{iz} \perp IC_i - \sum_{t\in T} \mu_{izt} \ge 0, i \in I, z \in Z \\
&0 \le \mu_{izt} \perp X_{iz} + x_{iz} - y_{izt} \ge 0, i \in I, z \in Z, t \in T \\
&0 \le y_{izt} \perp MC_i + \mu_{izt} - \rho_{zt} \ge 0, i \in I, z \in Z, t \in T \\
&0 \le s_{zt} \perp VOLL - \rho_{zt} + \delta_{zt} \ge 0, z \in Z, t \in T \\
&0 \le \delta_{zt} \perp D_{zt} - s_{zt} \ge 0, z \in Z, t \in T \\
&\nu_z \text{ free } \perp \sum_{i\in I} x_{iz} - \sum_{n\in N(z)} \tilde{x}_n = 0, z \in Z \\
&\rho_{zt} \text{ free } \perp -p_{zt} + \sum_{i\in I} y_{izt} + s_{zt} - D_{zt} = 0, z \in Z, t \in T \\
&\tilde{\rho}_{zt} \text{ free } \perp -p_{zt} + \sum_{n\in N(z)} \tilde{y}_{nt} - D_{zt} = 0, z \in Z, t \in T \\
&0 \le \tilde{y}_{nt} \perp \tilde{\mu}_{nt} - \tilde{\rho}_{Z(n)t} - \tilde{\rho}_{nt} \ge 0, i \in I, n \in N, t \in T \\
&0 \le \tilde{x}_n \perp - \sum_{t\in T} \tilde{\mu}_{nt} + \tilde{\nu}_{Z(n)} \ge 0, n \in N \\
&p_{zt} \text{ free } \perp \rho_{zt} + \tilde{\rho}_{zt} = 0, z \in Z, t \in T \\
&r_{nt} \text{ free } \perp \tilde{\rho}_{nt} - \phi_t + \sum_{k\in K} PTDF_{kn} \cdot \psi_{kt} = 0 \\
&f_{kt} \text{ free } \perp -\psi_{kt} + \lambda^+_{kt} - \lambda^-_{kt} = 0 \\
&0 \le \tilde{\mu}_{int} \perp X_{in} - \tilde{y}_{int} \ge 0, i \in I, n \in N, t \in T \\
&\tilde{\rho}_{nt} \text{ free } \perp -r_{nt} + \tilde{y}_{nt} - D_{nt} = 0, n \in N, t \in T \\
&\phi_t \text{ free } \perp \sum_{n\in N} r_{nt} = 0, t \in T \\
&\psi_{kt} \text{ free } \perp f_{kt} - \sum_{n\in N} PTDF_{kn} \cdot r_{nt} = 0, k \in K, t \in T \\
&0 \le \lambda^+_{kt} \perp TC_k - f_{kt} \ge 0, k \in K, t \in T \\
&0 \le \lambda^-_{kt} \perp TC_k + f_{kt} \ge 0, k \in K, t \in T
\end{aligned} \tag{B.5}$$

Appendix C. Zonal model with price aggregation

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

$$\begin{aligned}
&\min_{y,r,f} \sum_{i\in I, n\in N} MC_i \cdot y_{in} \\
&(\mu_{in}) : X_{in} - y_{in} \geq 0, i \in I, n \in N \\
&(\rho_n) : -r_n + \sum_{i\in I} y_{in} - D_n = 0, n \in N \\
&(\psi_k) : f_k - \sum_{n\in N} PTDF_{kn} \cdot r_n = 0, k \in K \\
&(\phi) : \sum_{n\in N} r_n = 0 \\
&(\lambda_k^-, \lambda_k^+) : -TC_k \leq f_k \leq TC_k, k \in K \\
&y \geq 0
\end{aligned} \tag{C.1}$$

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

$$\begin{aligned}
&\max_{\rho,\mu,\psi,\phi,\lambda} \sum_n D_n \rho_n - \sum_{in} X_{in}\mu_{in} - \sum_k TC_k(\lambda_k^+ + \lambda_k^-) \\
&(r_n) : \rho_n + \sum_k PTDF_{kn}\psi_k - \phi = 0, n \in N \\
&(y_{in}) : MC_i - \rho_n + \mu_{in} \geq 0, i \in I, n \in N \\
&(f_k) : -\psi_k - \lambda_k^- + \lambda_k^+ = 0, k \in K \\
&\mu, \lambda^+, \lambda^- \geq 0
\end{aligned} \tag{C.2}$$

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 (C.2) the following constraints:

$$\rho_{n_1} = \rho_{n_2}, \ \ \forall (n_1, n_2) \in z, \forall z \in Z \tag{C.3}$$

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:

$$\rho_n = \rho_z, \ \ \forall n \in N(z), \forall z \in Z \tag{C.4}$$

This in turn is equivalent to replacing every occurrence of ρ_n by $\rho_{Z(n)}$ in (C.2), yielding:

$$\begin{aligned}
&\max_{\rho,\mu,\psi,\phi,\lambda} \sum_n D_n \rho_{Z(n)} - \sum_{in} X_{in}\mu_{in} - \sum_k TC_k(\lambda_k^+ + \lambda_k^-) \\
&(r_n) : \rho_{Z(n)} + \sum_k PTDF_{kn}\psi_k - \phi = 0, n \in N \\
&(y_{in}) : MC_i - \rho_{Z(n)} + \mu_{in} \geq 0, i \in I, n \in N \\
&(f_k) : -\psi_k - \lambda_k^- + \lambda_k^+ = 0, k \in K \\
&\mu, \lambda^+, \lambda^- \geq 0
\end{aligned} \tag{C.5}$$

Model (C.5) 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:

$$\begin{aligned}
&\min_{y,r,f} \sum_{i\in I, n\in N} MC_i \cdot y_{in} \\
&(\mu_{in}) : X_{in} - y_{in} \geq 0, i \in I, n \in N \\
&(\rho_z) : -\sum_{n\in N(z)} r_n + \sum_{i\in I, n\in N(z)} y_{in} - \sum_{n\in N(z)} D_n = 0, z \in Z \\
&(\psi_k) : f_k - \sum_{n\in N} PTDF_{kn} \cdot r_n = 0, k \in K \\
&(\phi) : \sum_{n\in N} r_n = 0 \\
&(\lambda_k^-, \lambda_k^+) : -TC_k \leq f_k \leq TC_k, k \in K \\
&y \geq 0, s \geq 0
\end{aligned} \tag{C.6}$$

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 (C.5) is eventually equivalent to the following zonal problem:

$$\begin{aligned}
&\min_{y,r,f} \sum_{i\in I, z\in Z} MC_i \cdot y_{iz} \\
&X_{iz} - y_{iz} \geq 0, i \in I, z \in Z \\
&-p_z + \sum_{i\in I} y_{iz} - D_z = 0, z \in Z \\
&p_z - \sum_{n\in N(z)} r_n = 0, z \in Z \\
&f_k - \sum_{n\in N} PTDF_{kn} \cdot r_n = 0, k \in K \\
&\sum_{n\in N} r_n = 0 \\
&-TC_k \leq f_k \leq TC_k, k \in K \\
&y \geq 0, s \geq 0
\end{aligned} \tag{C.7}$$

We now go on to prove the conditions of equivalence between the nodal and zonal models stated in section 3.

Proof of Proposition 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.

$$L^{\text{inter}} = \Big\{ \Big(Z(m(k)), Z(n(k)) \Big) \ \forall k \in L \text{ s.t. } Z(m(k)) \neq Z(n(k)) \Big\}, \tag{C.8}$$

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 (C.2). This in turns implies that $\psi_k = 0 \ \forall k \in L^{\text{intra}}$. Using the first equation of model (C.2), 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 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 (C.5) from (C.2), (C.2) and (C.5) are equivalent, which in turn implies that (C.1) and (C.6) 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. □

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

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

Appendix 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 the decentralized FBMC capacity expansion problem with strategic reserve can be formulated as the following LCP:

$$0 \leq x_{iz} \perp IC_i - \sum_{t \in T} \mu_{izt} \geq 0, i \in I, z \in Z \tag{D.1a}$$

$$0 \leq \mu_{izt} \perp X_{iz} + x_{iz} - y_{izt} \geq 0, i \in I, z \in Z, t \in T \tag{D.1b}$$

$$0 \leq y_{izt} \perp MC_i + \mu_{izt} - \rho^+_{zt} + \rho^-_{zt} \geq 0, i \in I, z \in Z, t \in T \tag{D.1c}$$

$$0 \leq s_{zt} \perp VOLL - \rho^+_{zt} + \rho^-_{zt} + \delta_{zt} \geq 0, z \in Z, t \in T \tag{D.1d}$$

$$0 \leq \delta_{zt} \perp D_{zt} - s_{zt} \geq 0, z \in Z, t \in T \tag{D.1e}$$

$$0 \leq \nu_z \perp \sum_{i \in I} x_{iz} - \sum_{n \in N(z)} \tilde{x}_n \geq 0, z \in Z \tag{D.1f}$$

$$0 \leq \rho^+_{zt} \perp -p^+_{zt} + p^-_{zt} + \sum_{i \in I} y_{izt} + s_{zt} - D_{zt} \geq 0, z \in Z, t \in T \tag{D.1g}$$

$$0 \leq \rho^-_{zt} \perp p^+_{zt} - p^-_{zt} - \sum_{i \in I} y_{izt} - s_{zt} + D_{zt} \geq 0, z \in Z, t \in T \tag{D.1h}$$

$$0 \leq \tilde{\rho}^+_{zt} \perp -p^+_{zt} + p^-_{zt} + \sum_{n \in N(z)} \tilde{y}_{nt} + \sum_{n \in N(z)} z_{nt} - D_{zt} = 0, z \in Z, t \in T \tag{D.1i}$$

$$0 \leq \tilde{\rho}^-_{zt} \perp p^+_{zt} - p^-_{zt} - \sum_{n \in N(z)} \tilde{y}_{nt} - \sum_{n \in N(z)} z_{nt} + D_{zt} = 0, z \in Z, t \in T \tag{D.1j}$$

$$0 \leq \tilde{y}_{nt} \perp \tilde{\mu}_{nt} - \tilde{\rho}^+_{Z(n)t} + \tilde{\rho}^-_{Z(n)t} - \tilde{\rho}^+_{nt} + \tilde{\rho}^-_{nt} \geq 0, n \in N, t \in T \tag{D.1k}$$

$$0 \leq \tilde{x}_n \perp - \sum_{t \in T} \tilde{\mu}_{nt} + \tilde{\nu}_{Z(n)} \geq 0, n \in N \tag{D.1l}$$

$$0 \leq p^+_{zt} \perp \rho^+_{zt} - \rho^-_{zt} + \tilde{\rho}^+_{zt} - \tilde{\rho}^-_{zt} \geq 0, z \in Z, t \in T \tag{D.1m}$$

$$0 \leq p^-_{zt} \perp -\rho^+_{zt} + \rho^-_{zt} - \tilde{\rho}^+_{zt} + \tilde{\rho}^-_{zt} \geq 0, z \in Z, t \in T \tag{D.1n}$$

$$0 \leq \tilde{r}^+_{nt} \perp \tilde{\rho}^+_{nt} - \tilde{\rho}^-_{nt} - \phi^+_t + \phi^-_t + \sum_{k \in K} PTDF_{kn} \cdot (\psi^+_{kt} - \psi^-_{kt}) \geq 0 \tag{D.1o}$$

$$0 \leq \tilde{r}^-_{nt} \perp -\tilde{\rho}^+_{nt} + \tilde{\rho}^-_{nt} + \phi^+_t - \phi^-_t - \sum_{k \in K} PTDF_{kn} \cdot (\psi^+_{kt} - \psi^-_{kt}) \geq 0 \tag{D.1p}$$

$$0 \leq f^+_{kt} \perp -\psi^+_{kt} + \psi^-_{kt} + \lambda^+_{kt} - \lambda^-_{kt} \geq 0 \tag{D.1q}$$

$$0 \leq f^-_{kt} \perp \psi^+_{kt} - \psi^-_{kt} - \lambda^+_{kt} + \lambda^-_{kt} \geq 0 \tag{D.1r}$$

$$0 \leq \tilde{\mu}_{nt} \perp \sum_{i \in I} X_{in} - \tilde{y}_{nt} \geq 0, n \in N, t \in T \tag{D.1s}$$

$$0 \leq \tilde{\rho}^+_{nt} \perp -\tilde{r}^+_{nt} + \tilde{r}^-_{nt} + \tilde{y}_{nt} + z_{nt} - D_{nt} \geq 0, n \in N, t \in T \tag{D.1t}$$

$$0 \leq \tilde{\rho}^-_{nt} \perp \tilde{r}^+_{nt} - \tilde{r}^-_{nt} - \tilde{y}_{nt} - z_{nt} + D_{nt} \geq 0, n \in N, t \in T \tag{D.1u}$$

$$0 \leq \phi_t^+ \perp \sum_{n \in N} \tilde{r}_{nt}^+ - \tilde{r}_{nt}^- \geq 0, t \in T \tag{D.1v}$$

$$0 \leq \phi_t^- \perp \sum_{n \in N} \tilde{r}_{nt}^- - \tilde{r}_{nt}^+ \geq 0, t \in T \tag{D.1w}$$

$$0 \leq \psi_{kt}^+ \perp f_{kt}^+ - f_{kt}^- - \sum_{n \in N} PTDF_{kn} \cdot r_{nt} \geq 0, k \in K, t \in T \tag{D.1x}$$

$$0 \leq \psi_{kt}^- \perp -f_{kt}^+ + f_{kt}^- + \sum_{n \in N} PTDF_{kn} \cdot r_{nt} \geq 0, k \in K, t \in T \tag{D.1y}$$

$$0 \leq \lambda_{kt}^+ \perp TC_k - f_{kt}^+ + f_{kt}^- \geq 0, k \in K, t \in T \tag{D.1z}$$

$$0 \leq \lambda_{kt}^- \perp TC_k + f_{kt}^+ - f_{kt}^- \geq 0, k \in K, t \in T \tag{D.1aa}$$

$$0 \leq Z_n \perp \tilde{IC} - \sum_{t \in T} \gamma_{nt} \geq 0, n \in N \tag{D.1ab}$$

$$0 \leq z_{nt} \perp \tilde{MC} + \gamma_{nt} - \tilde{\rho}_{nt}^+ + \tilde{\rho}_{nt}^- - \tilde{\rho}_{Z(n)t}^+ + \tilde{\rho}_{Z(n)t}^- \geq 0, n \in N, t \in T \tag{D.1ac}$$

$$0 \leq \gamma_{nt} \perp Z_n - z_{nt} \geq 0, n \in N, t \in T \tag{D.1ad}$$

This is essentially problem (B.5) 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.

Our reference for the proof of existence is Cottle et al. (2009). We use the following result:

Definition 1. A matrix $M \in \mathbb{R}^{n \times n}$ is said to be **copositive** if $x^\top Mx \geq 0$ for all $x \in \mathbb{R}_+^n$

Theorem 1. *(Theorem 3.8.6 of Cottle et al. (2009)) Let $M \in \mathbb{R}^{n \times n}$ be copositive and let $q \in \mathbb{R}^n$ be given. If the implication*

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

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

In other words, Theorem 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.

Let us apply this theorem to our problem. We first show that the matrix M of our LCP (D.1) 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):

$$\nu_z \tag{D.2}$$

$$\tilde{N} = \begin{matrix} x_{iz} \\ \\ \\ \\ \end{matrix} \begin{pmatrix} 0 & \dots & I & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \tag{D.3}$$

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 non-negative if each x_i and $\tilde{\nu}_i$ are non-negative.

Now, let $v^* = (x^*_{iz}, y^*_{izt}, \dots, \gamma^*_{nt})$ be a solution to the homogeneous version of LCP (D.1). Using (D.1ab) and (D.1ad), we have that $\gamma^*_{nt} = 0$. From (D.1ac), we deduce that $\tilde{\rho}^{*+}_{nt} - \tilde{\rho}^{*-}_{nt} \leq -\tilde{\rho}^{*+}_{Z(n)t} + \tilde{\rho}^{*-}_{Z(n)t}$. (D.1a), (D.1b) and (D.1c) imply that $\rho^{*+}_{zt} - \rho^{*-}_{zt} \leq 0$ which in turns leads to $\tilde{\rho}^{*+}_{zt} - \tilde{\rho}^{*-}_{zt} \geq 0$ by (D.1m) and (D.1n) 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 line are non-negative, as they correspond to the product of non-negative quantities. In the second line, the first term is zero by equations (D.1m) and (D.1n) and the second term is non-negative by the non-positivity of $\tilde{\rho}^{*+}_{nt} - \tilde{\rho}^{*-}_{nt}$ that we have shown above. By Theorem 1, we can conclude that the LCP (D.1) 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.

Appendix D.2. Unicity

In order to investigate the unicity of solutions of 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 (D.1), different from the one reported in section 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 and thus that the solution set of our problem is not 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 paper, 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 paper, 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 paper. A more formal treatment of this multiplicity of equilibria and their implication is thus relegated to future research.

Appendix E. Dimensionality reduction

We perform a dimensionality reduction on both the load duration curve and the network.

Appendix E.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. We use the dynamic programming algorithm presented in Konno and Kuno (1988) to solve for the best approximation.

Appendix E.2. Network reduction

For the network reduction, we also fix a priori the targeted number of nodes to 100. We apply a hierarchical clustering algorithm with a distance measure that is the weighted sum of two distance measures: the Euclidean Commute Time distance (Yen et al., 2005) and the shortest path distance. The weights of the ECT distance are set to the Euclidean norm of the differences in nodal prices, obtained with fixed capacity, on the 20 periods. We choose to use these two distances in order to have clusters that at the same time are moderate in size and reflect the structural congestion of the network. The effect of the ECT distance weighted by nodal prices is to favor clusters that have similar nodal prices across periods and thus low internal congestion. The unweighted shortest path distance favors clusters that are strongly connected. Note that, in the shortest path distance, we do not consider cross-zonal lines. This ensures that clusters will only contain nodes of the same bidding zone.

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 Fortenbacher et al. (2018). 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 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.

$$\bar{\rho}_{mt} = \frac{\sum_{n \in N(m)} \bar{\rho}_{nt}}{|N(m)|}, \quad \forall m \in M$$

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 $\bar{\rho}_{mt}$:

$$\min_{\substack{TC \\ y,s,r,f \\ \rho,\mu,\psi,\phi,\lambda}} \sum_{m\in M, t\in T} (\rho_{mt} - \bar{\rho}_{mt})^2 \tag{E.1a}$$

$$\sum_{i\in I, m\in M} MC_i \cdot y_{imt} + \sum_{m\in M} VOLL \cdot s_{mt} = \sum_{n} D_{mt}\rho_{mt} - \sum_{im} X_{im}\mu_{imt} - \sum_{k} TC_k(\lambda^+_{kt} + \lambda^-_{kt}), t \in T \tag{E.1b}$$

$$X_{im} - y_{imt} \geq 0, i \in I, m \in M, t \in T \tag{E.1c}$$

$$D_{mt} - s_{mt} \geq 0, m \in M, t \in T \tag{E.1d}$$

$$- r_{mt} + \sum_{i\in I} y_{imt} + s_{mt} - D_{mt} = 0, m \in M, t \in T \tag{E.1e}$$

$$f_{kt} - \sum_{m\in M} PTDF_{km} \cdot r_{mt} = 0, k \in K, t \in T \tag{E.1f}$$

$$\sum_{m\in M} r_{mt} = 0, t \in T \tag{E.1g}$$

$$- TC_k \leq f_{kt} \leq TC_k, k \in K, t \in T \tag{E.1h}$$

$$\rho_{mt} + \sum_{k} PTDF_{km}\psi_{kt} - \phi = 0, m \in M, t \in T \tag{E.1i}$$

$$MC_i - \rho_{mt} + \mu_{imt} \geq 0, i \in I, m \in M, t \in T \tag{E.1j}$$

$$VOLL - \rho_{mt} + \delta_{mt} \geq 0, m \in M, t \in T \tag{E.1k}$$

$$- \psi_{kt} - \lambda^-_{kt} + \lambda^+_{kt} = 0, k \in K, t \in T \tag{E.1l}$$

$$y \geq 0, s \geq 0, \mu, \lambda^+, \lambda^- \geq 0 \tag{E.1m}$$

In this problem, constraint (E.1b) represents strong duality of the market clearing problem, constraints (E.1c)-(E.1h) are the primal constraints and constraints (E.1i)-(E.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 (E.1b) make the problem nonconvex. These nonconvexities and the large size of the problem make it complicated to solve directly. Therefore, we do not pass it directly to the solver, but instead we solve it with the alternating direction method. That is, we start by fixing the capacities TC_k to an initial value and we solve (E.1) by alternatively fixing TC_k and λ^+_k, λ^-_k until no more progress can be made, which yields a suboptimal solution.

Appendix F. Splitting-based algorithm for solving the decentralized FBMC problem.

We use a modified version of the basic splitting algorithm to solve the LCP corresponding to the decentralized FBMC problem. This corresponds to Algorithm 5.2.1 of Cottle et al. (2009). Let us denote this problem by $LCP(q, M)$, where M is the matrix of the LCP and q is the vector of independent terms. The splitting approach consists in separating M into two matrices B and C, with $M = B+C$, and such that the $LCP(\cdot, B)$ is easier to solve. A natural split for our problem is to define B as the matrix associated to the LCP formulation of the centralized problem. This LCP can be formulated equivalently as a linear program (LP) and is thus solved efficiently. Based on this splitting, the algorithm can be written as follows (Cottle et al., 2009):

Step 0. Initialization. Let z_0 be an arbitrary nonnegative vector, set $\nu = 0$.

Step 1. General iteration. Given $z^{\nu} \geq 0$, solve the $LCP(q^{\nu}, B)$ where

$$q^{\nu} = q + Cz^{\nu}$$

and let $z^{\nu+1}$ be an arbitrary solution.

Step 2. Test for termination. If $z^{\nu+1}$ satisfies a prescribed stopping rule, terminate. Otherwise, return to Step 1 with ν replaced by $\nu + 1$.

These iterations have the important property that, if a fixed point is reached, then it is a solution of $LCP(q, M)$. However, the converse is not necessarily true if the $LCP(q, M)$ can have multiple solutions: a solution of the $LCP(q, M)$ may not be a fixed point of the iterations. As the $LCP(q^{\nu}, B)$ can be solved as an LP, we can remedy this by selecting $z^{\nu+1}$ as the closest point to z^{ν} in the solution set of that LP, which restores the property that a solution of the $LCP(q, M)$ is a fixed point of the iterations. $LCP(q^{\nu}, B)$ can thus be solved with the following convex quadratic program:

$$\min \|z^{\nu} - z^{\nu-1}\|_2^2 \tag{F.1a}$$

$$\sum_{i\in I, z\in Z} \Big(IC_i + \sum_{m\in\{1,\ldots,M\}} U_{miz}\gamma_m^{\nu-1}\Big) \cdot x_{iz} + \sum_{i\in I, z\in N, t\in T} MC_i \cdot y_{izt} + \sum_{n\in N, t\in T} VOLL \cdot s_{zt} = \sum_{zt} p_{zt} D_{zt} - \sum_{izt} X_{iz}\mu_{izt} + \sum_{m} \gamma_m W_m \tag{F.1b}$$

$$X_{iz} + x_{iz} \geq y_{izt}, i \in I, z \in Z, t \in T \tag{F.1c}$$

$$-p_{zt} + \sum_{i\in I} y_{izt} + s_{zt} - D_{zt} = 0, z \in Z, t \in T \tag{F.1d}$$

$$\sum_{z\in Z} V_{mz} p_z + \sum_{i\in I, z\in Z} U_{miz} x_{iz} + W_m \geq 0, \;\; \forall m \in \{1, ..., M\}, t \in T \tag{F.1e}$$

$$IC_i + \sum_{m\in\{1,\ldots,M\}} U_{miz}\gamma_m^{\nu-1} - \sum_{t\in T} \mu_{izt} - \sum_{m\in\{1,\ldots,M\}} U_{miz}\gamma_m \geq 0, i \in I, z \in Z \tag{F.1f}$$

$$MC_i - \rho_z + \mu_{iz} \geq 0, i \in I, z \in Z \tag{F.1g}$$

$$VOLL - \rho_z \geq 0, z \in Z \tag{F.1h}$$

$$\rho_{zt} - \sum_{m\in\{1,\ldots,M\}} V_{mz}\gamma_m \geq 0 \tag{F.1i}$$

$$x \geq 0, y \geq 0, s \geq 0, \rho \geq 0, \mu \geq 0, \gamma \geq 0 \tag{F.1j}$$

where $\gamma^{\nu-1}$ represents the optimal value of variable γ at the previous iteration, z^{ν} is the full vector of variables at the current iteration and $z^{\nu-1}$ is the full vector of the optimal values of variables at the previous iteration. In this formulation, equations (F.1b) to (F.1i) correspond to the optimality conditions of the centralized market clearing problem, where the investment cost IC_i has been replaced by $IC_i + \sum_{m\in\{1,\ldots,M\}} U_{miz}\gamma_m^{\nu-1}$. In particular, (F.1b) is the strong duality constraint, equations (F.1c)-(F.1e) are the primal constraints and equations (F.1f)-(F.1i) are the dual constraints. Objective (F.1a) minimizes the distance to the previous iterate among all solutions to the augmented LCP, i.e. $LCP(q^{\nu}, B)$.

Solving the $LCP(q^\nu, B)$ through problem (F.1) at each iteration, we now have the property that a vector z^* is a solution to $LCP(q, M)$ if and only if it is a fixed point of the iterations of the splitting algorithm. We can therefore have the following stopping criterion: $\|z^\nu - z^{\nu-1}\|_2^2 \leq \epsilon$, with a given ϵ, that we set in our experiments to $1E-2$. We have observed in practice that these iterates converge indeed to a solution, while we have not been able to obtain convergence in the case where we solve only the primal problem at each iteration. We have, however, no proof of convergence for this algorithm.

References for appendices

Boucher J, Smeers Y. Alternative models of restructured electricity systems, Part 1: No market power. Operations Research 2001;49(6):821–386. doi:10.1287/opre.49.6.821.10017.

Cottle R, Pang JS, Stone R. The Linear Complementarity Problem. SIAM, 2009. doi:10.1137/1.9780898719000.

Fortenbacher P, Demiray T, Schaffner C. Transmission network reduction method using nonlinear optimization. In: 2018 Power Systems Computation Conference (PSCC). 2018. p. 1–7. doi:10.23919/PSCC.2018.8442974.

Harker PT. Generalized Nash games and quasi-variational inequalities. European Journal of Operational Research 1991;54(1):81–94. doi:10.1016/0377-2217(91)90325-P.

Hogan WW. Contract networks for electricity power transmission: technical reference. Journal of Regulatory Economics 1992;242(September 1990):211–42.

Konno H, Kuno T. Best piecewise constant approximation of a function of single variable. Operations Research Letters 1988;7(4):205–10. URL: https://www.sciencedirect.com/science/article/pii/0167637788900302. doi:https://doi.org/10.1016/0167-6377(88)90030-2.

Yen L, Vanvyve D, Wouters F, Fouss F, Verleysen M, Saerens M. Clustering using a random walk based distance measure. In: ESANN 2005, European Symposium on Artificial Neural Networks. Bruges (Belgium); 2005. p. 317–24.