Energy Strategy Reviews 66 (2026) 102325

Contents lists available at ScienceDirect

Energy Strategy Reviews

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

Full length articles with a review element

Incentive design for demand-side flexibility in transmission system operation: A bi-level market equilibrium approach

Jannis Eichenberg [a,*], Hannes Hobbie [a], Tizian Schug [b]

[a] *Chair of Energy Economics, TUD Dresden University of Technology, Münchner Platz 3, Dresden, 01062, Germany*
[b] *Institute for Operations Research and Information Systems, Hamburg University of Technology, Am Schwarzenberg-Campus 4, Hamburg, 21073, Germany*

ARTICLE INFO

Handling Editor: Henrik Lund

Keywords:
Flexibility premium
Grid tariff
Bi-level optimization
Congestion management
Transmission grid
Market intervention

ABSTRACT

Increasing renewable electricity generation and the electrification of industry, mobility, and heating through sector coupling pose significant challenges to grid operators in maintaining secure and reliable system operations. Demand-side sector coupling applications increase electricity demands and stress electricity grids, but they also offer transmission system operators increased flexibility for congestion management. Due to the complexity of directly controlling decentralized demand-side technologies, incentive mechanisms present a promising solution for harnessing demand-side flexibility.

This study investigates various incentive schemes that encourage grid-supportive demand-side behavior, focusing on heat pumps by developing a bi-level programming framework. The framework models the decision-making processes of key stakeholders, including a TSO, an aggregator, and a market clearing agent, considering model-endogenous wholesale market equilibrium formation and congestion management optimization. The economic efficiency of different design options for grid congestion management is evaluated using an extended IEEE test system adjusted to an exemplary German market configuration.

The findings highlight the critical importance of time-dynamic premium design concepts due to the variability of renewable generation. While incentive-based market interventions increase electricity market costs and thereby shift consumer rents to producers, the reduced transmission system operation cost leads to overall gains in total system welfare.

1. Introduction

1.1. Motivation

The ongoing decarbonization of electricity systems is fundamentally transforming power systems. While the gradual replacement of conventional, centrally dispatched power plants with distributed and weather-dependent renewables reduces carbon emissions, it also poses significant challenges to market and grid operations. From a theoretical perspective, economically efficient pricing requires energy prices to reflect marginal costs at each location and time, i.e., nodal pricing. However, in systems that rely on zonal pricing, such as Germany, wholesale market prices do not accurately reflect scarcity. As a result, the continued integration of renewables exacerbates grid congestion and increases the need for costly remedial actions.

Additionally, renewables introduce regional mismatches between generation and consumption, driven, for example, by the expansion of offshore wind and the location of industrial centers far inland. At the same time, the further electrification of heating, mobility, and industrial processes is increasing electricity demand beyond traditional levels. These factors intensify the need for interventions in the transmission network. For example, Germany has seen a 75% increase in congestion management volumes and a near tripling of related costs over the past five years [1]. These levels are projected to persist as long as grid reinforcement and expansion do not keep pace with renewable deployment.

Demand-side flexibility is widely considered a promising means of providing system operators with operational support. Although flexible loads are connected to grids at different voltage levels, aggregated decentralized loads in distribution systems offer significant potential to contribute to congestion management in the transmission system. For example, [2] show that coordinated charging of two million electric vehicles could reduce Austria's congestion management costs by approximately 14%. Similarly, [3] analyze the grid-supportive use of various decentralized flexible loads in a case study for Germany in 2030, finding a reduction in cumulative grid congestion of 8%–16%.

* Corresponding author.
E-mail address: jannis.eichenberg@tu-dresden.de (J. Eichenberg).

https://doi.org/10.1016/j.esr.2026.102325
Received 2 March 2026; Received in revised form 30 June 2026; Accepted 3 July 2026
Available online 15 July 2026

Unlocking this demand-side potential requires efficient market mechanisms that reflect scarcity and coordinate deployment. However, Germany, like many other European countries, faces regulatory, political, and institutional constraints that currently preclude the implementation of fully locational pricing. Some countries, including Germany, have therefore begun to implement time-dynamic tariff schemes at the distribution level as an alternative to directly pricing scarcity. However, these efforts rarely extend to the transmission grid, where large-scale congestion occurs and redispatch is particularly costly. Without addressing transmission-level dynamics, such tariffs have limited impact on relieving high-voltage grid constraints or substituting conventional redispatch actions.

Moreover, congestion management in transmission networks is inherently spatiotemporal: it is not only important when load is shifted, but also where. Unlike uniform tariffs or wholesale market prices, redispatch actions are targeted to relieve specific network bottlenecks. As such, location-dependent and time-dynamic incentives are essential to activate flexible loads in a way that effectively supports transmission system operation.

In this constrained market environment, incentive-based mechanisms that reward flexible demand for load shifting can improve welfare without requiring full price differentiation in the short run. These incentives can be interpreted as a side-payment to flexible consumers through reductions in grid tariffs. Compared to the status quo of costly redispatch, such schemes enable ex-ante adjustments of demand at lower system cost, providing a readily implementable means to enhance flexibility in transmission system operation, which is urgently needed. Ultimately, they constitute a pragmatic second-best-type mechanism relative approaches in which market prices fully reflect or approximate scarcity (i.e. nodal pricing or market splitting) and thereby directly coordinate flexibility.

Additionally, they reduce implementation and transaction costs while facilitating broader participation of decentralized flexibility by relying on relatively simple and transparent incentives rather than fundamental changes to market design, local flexibility markets, or direct control mechanisms, which typically involve long implementation times. As a result, even though such mechanisms are not as economically efficient as pricing locational scarcity, they can yield welfare gains by lowering total congestion management costs, in the short run, until grid investments alleviate congestion. Moreover, price-based incentives not only support grid operations but also enable flexible end users to benefit from cost reductions, resulting in a win-win outcome [4].

The design of such incentives in a constrained market environment introduces a methodological challenge: tariff signals must provide effective economic incentives to flexible consumers by accounting for both time-varying zonal market prices and spatiotemporal transmission grid dynamics, while ensuring that the resulting load adjustments remain technically feasible and operationally beneficial within the physical constraints of the transmission system. Additionally, because transmission-connected flexible loads (including aggregated loads from distribution levels) are typically large, their behavior can significantly affect wholesale market outcomes, including zone-wide marginal prices and dispatch schedules. Hence, evaluating incentive-based mechanisms in this context requires an integrated, multi-level modeling framework that endogenously captures interactions between tariff setting, wholesale market clearing, and grid constraints.

This paper addresses these challenges through a bi-level optimization framework that reflects the hierarchical nature of tariff-based coordination. At the upper level, the regulated TSO designs dynamic transmission-grid flexibility premia. These premia are offered to flexible consumers to incentivize shifting demand toward periods that are expected to alleviate network congestion before market clearing. The TSO determines the premia to minimize the overall congestion management cost, defined as the sum of flexibility costs (i.e., premium payments) and any remaining redispatch costs. Since these costs are ultimately recovered through regulated network tariffs, inefficient incentive design can increase network tariffs, whereas well-designed flexibility premia reduce the total cost of congestion management. The lower level models the market clearing process and the aggregated electricity procurement decisions of flexible consumers, who respond optimally to the imposed tariffs under the assumption of perfect foresight and in a deterministic setting.

By explicitly capturing the interaction between tariff design and market behavior, the model enables the principled design of spatiotemporal dynamic pricing schemes. The framework contributes to the energy systems literature by offering a structured approach to co-optimizing regulatory interventions, market outcomes, and infrastructure operations and to investigating their interactions in systems characterized by high flexibility and diminishing controllability.

1.2. Economic background

Zonal pricing is widely used across European electricity markets. A well-known drawback of this pricing mechanism is that market prices do not reflect local congestion costs. This issue is particularly pronounced in systems characterized by large geographic disparities between (renewable) electricity generation and electricity demand, where substantial congestion occurs at the transmission grid level—as is the case in Germany.

In such contexts, incentive-based mechanisms, such as time- and location-varying grid charges, aiming at alleviating grid congestion through load shifting prior to market clearing can be economically interpreted as side payments. Typically, these side payments are designed to reward large electricity consumers or demand-side aggregators for grid-supportive load dispatch or procurement behavior.

As market participants incorporate these side payments into their procurement decisions, the resulting incentives alter bidding behavior and shift the market equilibrium. In contrast to nodal pricing, where congestion costs are internalized locally in prices, this shift in the market equilibrium under zonal pricing affects all market participants within the zone. Fig. 1 illustrates how a side payment to flexible loads shifts the market equilibrium and how it affects the distribution of welfare. Note, the figure presents the supply and demand curves from two market snapshots back-to-back, allowing both equilibria to be displayed in a single diagram.

Without incentive: Consider a single zone with supply (green) and demand (blue/purple) curves, separated into fix (blue) and flexible (purple) demand. For two time steps t_x (expensive) and t_y (cheap) the market clears at prices $p_x > p_y$, respectively. Because flexible loads are indifferent to the timing of consumption, they are dispatched entirely in the cheap hour t_y.

With incentive: Assume a local congestion is predicted in the cheap hour t_y. Flexible demand is incentivized by a monetary flexibility premium at time t_x, so that the effective price for flexible loads becomes $\hat{p}_x$. If the incentive is sufficiently large ($\hat{p}_x \le p'_y$), flexible loads are incentivized to shift their consumption to the expensive hour t_x, relieving the congestion in t_y. The additional load in t_x raises the equilibrium price to $p'_x > p_x$, while the reduced load in t_y lowers the price to $p'_y < p_y$. The first price effect shifts consumer surplus toward producers, while the second operates in the opposite direction. Flexible consumers receive a flexibility premium (side-payment) of $p'_x - p'_y$ (or larger), which compensates them for the higher purchase price in t_x and also lets them indirectly benefit from the price drop in t_y in terms of additional consumer surplus generated ($p_y - p'_y$).

Comparing the gains and losses in surplus, the net welfare impact is governed by the slope of the supply curve. In the schematic illustration the producers capture a larger share of the surplus at the expense of the (fixed) consumers. Because an incentive scheme always induces a shift

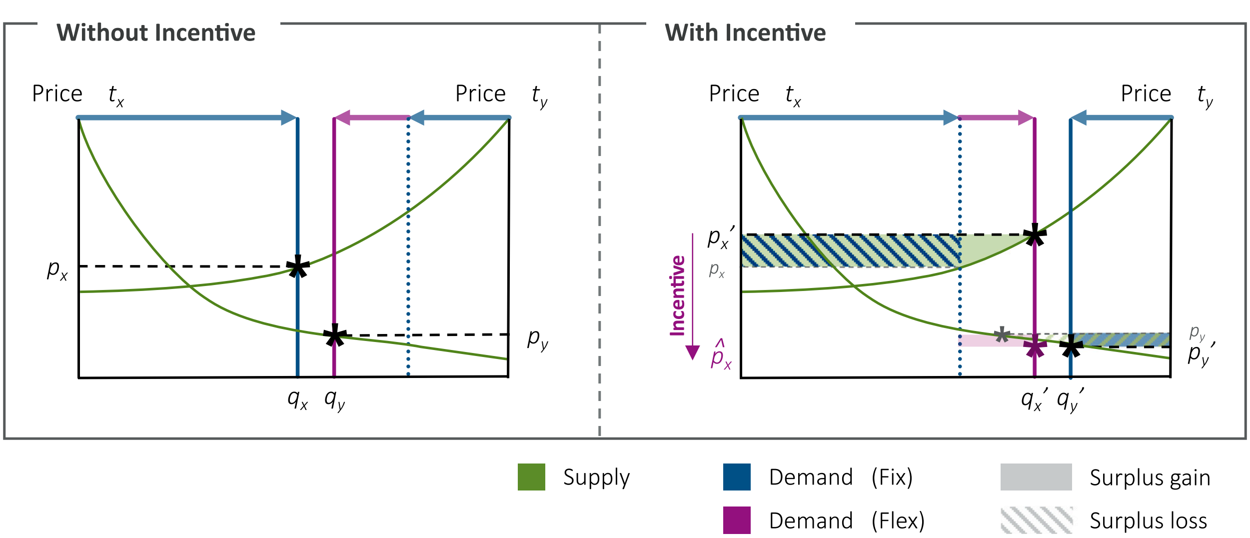


Fig. 1. Illustrative example. Incentivizing load-shifts toward a higher-price hour transfers consumer surplus to producers, highlighting a potential welfare-distribution problem. *Left panel:* market-clearing without any incentive. *Right panel:* market equilibrium after the incentive has shifted flexible demand. The vertical axis on the left of each panel corresponds to the expensive hour t_x (read from left-to-right in the figure). The vertical axis on the right corresponds to the cheap hour t_y and is plotted in reverse order (read from right-to-left).

of flexible demand toward a higher market price,[1] and because supply curves typically become steeper at higher price levels, this pattern is not accidental: it reflects a systematic consequence of mitigating local congestion within a zonal market design.

In Germany, most transmission congestion stems from a north–south imbalance, wherein renewable generation in the north and load centers in the south imply high stress on the lines. Consequently, a local side-payment that targets flexible loads in the south during congested periods deserves careful consideration since it may affect (fix) consumers throughout the entire zone, including the north. In contrast, a nodal design would price the congestion locally, charging only the south-side participants.

A holistic distribution analysis, however, cannot be based solely on (shifted) market outcomes, as it must also capture the benefits of cheaper grid operation through extended congestion management.[2] Consequently, a model-based analysis necessitates co-optimized market clearing and extended congestion management through incentive instruments.

2. Literature review on incentive-based mechanisms for congestion management

Flexibility for congestion management can be enabled through different approaches. One strand of research focuses on flexibility markets, which provide a framework for trading flexibility [5–8]. In contrast, this study focuses on incentive-based mechanisms to activate demand-side flexibility. Research on monetary incentives for grid-supportive load behavior can be further divided into three categories according to the recent literature, see Fig. 2. The first category focuses on regulatory and policy-related aspects, with particular attention to the German implementation of dynamic grid charges. The second category emphasizes operational perspectives solely and examines the role of such mechanisms in distribution grid operation and congestion management. The third category analyzes market outcomes, with a particular focus on welfare and distributional effects, but neglects the specifics of grid operation. While these strands provide valuable insights, studies that explicitly examine the intersection between regulatory design, grid operation, and market outcomes remain relatively scarce. This intersection constitutes the main focus of the present study.

The following sections briefly summarize the main contributions within each category of the literature and identify potential gaps in the existing research landscape. A more descriptive overview on the scientific literature is provided in Table 3 in Appendix A.

Regulatory and policy perspectives

Several countries, including Sweden, Norway, France, and Germany, have implemented early-stage incentive schemes at the distribution level to promote grid-serving flexibility [9]. These efforts remain limited, largely due to insufficient grid digitization and lack of real-time system data. For example, the German regulator explicitly describes its approach as a first step toward more dynamic tariffs. This underscores the continued need for methodological and practical advancement, as reflected in the growing research on incentive-based demand-side flexibility. Focusing on the transmission level, the German Monopolies Commission proposes dynamic grid fees as the third-best option[3] for pricing transmission-grid constraints, while the existing static fee regime is currently under revision [10]. In a discussion paper [11], the four German transmission system operators outline what a possible dynamic grid fee could look like as part of a future grid fee system. The paper focuses on implementation issues and discusses various grid fee designs in relation to practical challenges. Industry stakeholders emphasize that high temporal resolution goes hand in hand with high requirements for data platforms. However, they prefer release dates close to market clearing, hourly resolution and distance from further temporal updates. Spatial resolution must outweigh congestion requirements, implementation effort, and the effects of load shifting.

Operational perspectives on grid congestion management

A great variety of literature explores model-based indirect congestion management instruments, such as incentive mechanisms to

[1] Assuming cost-minimal flexibility optimizes against wholesale market clearing prices.

[2] All congestion management costs (including cost for demand-side incentivization) are passed on to consumers via grid fees.

[3] The first- and second-best options are nodal price systems and market zone splitting, respectively.

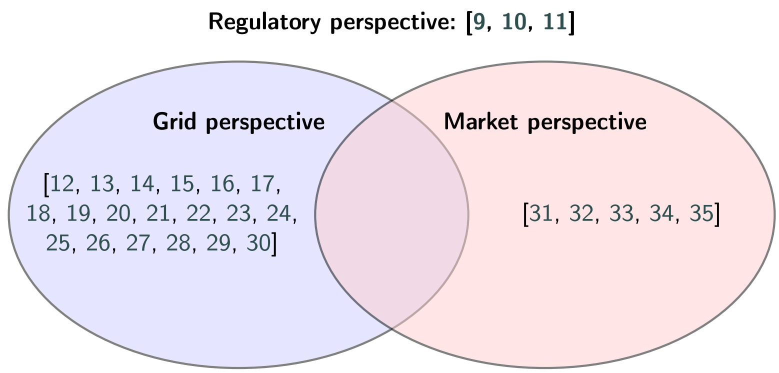


Fig. 2. Classification of literature on incentive-based congestion management.

mitigate infeasible demand-side dispatch due to (distribution) grid constraints. For example, [12] introduce a congestion price in distribution systems to manage grid constraints and steer demand response. The procedure for calculating the congestion price is evaluated with spot-market prices from NordPool Spot and verified through a case study on a Danish distribution system. Infeasible congestion situations, however, may request unlimited congestion prices. To ensure consumer participation, [13] provide a relaxed model, where congestion prices are limited. Moreover, they incorporate additional congestion instruments, such as *direct control* of DSOs over flexible assets for congestion resolution. [14] extend congestion management by an incentive based emergency demand-response program. They compare different demand elasticities in a IEEE-30-bus system and show, that the demand-response program can reduce congestion costs depending on consumers elasticities. [15] employ a bi-level program where the DSO gives bonuses to agents who can boost demand or cut back on power production. Since these models usually need a lot of computing power, they use an IEEE-37-bus test system. The results show that the game-theoretical approach is suitable to activate local flexibilities and decreases RES curtailments. [16] study three different charging strategies for electric vehicles concerning their implications for the operation of distribution grids. Their findings indicate that charging strategies optimized against time-variable grid charges outperform other strategies in terms of grid load management. Similarly, [17] investigate various electricity price and grid charging models, focusing on residential consumer electricity procurement and grid reinforcement costs. They showed that flexible loads can be effectively steered using time-varying grid charges, particularly when both temporal and locational components are considered. [18] highlight the necessity of frequently updating dynamic price signals that incentivize grid-serving behavior. Static time-of-use structures spanning long intervals fail to reflect real-time grid utilization, leading to inefficient and misguided incentives, whereas highly dynamic incentives track the true state of the grid. [19] introduce multidimensional dynamic electricity prices that incorporate grid signals. These signals are stacks of limited price series, that allow wholesale market orientation while incentivizing for peak load reduction. Key evaluation indicators include economic results for households, peak load in the distribution grid, and the power during the highest residual load in the market zone. The results show that, without power limits, wholesale prices concentrate demand,[4] whereas stacked prices with power limits reduce load peaks. Another case study on incentive instruments indicates that using deterministic methods might not work well because of the DSO's forecast error [21]. The authors provide a robust extension that handles congestion well in uncertain conditions, like when there are differences between the DSO's forecasts and the aggregators accurate energy requirements. Complementing this, [22] evaluates various dynamic tariff structures to determine how effectively they steer household flexibility toward grid-serving behavior. The results suggest that while time-variable tariffs can inadvertently create rebound peaks, linking charges directly to real-time grid load offers a promising alternative to direct load control for maintaining system efficiency. The authors stress, that inappropriate designs may cause unnecessary demand response activations and welfare losses.

In a market-integration context, [23] embed grid-serving tariff designs into locational energy markets to reflect distribution-level constraints. The influence of tariff designs on energy community performance is examined in [24], whose results reveal that coupling energy sharing arrangements with heterogeneous tariff structures can erode overall grid efficiency.

[25] use a bi-level program to examine how electrolyzer flexibility can contribute to congestion management in distribution grids. Their case study indicates that congestion pricing effectively encourages investments in electrolyzers while reducing total distribution grid operation and reinforcement costs. [26] argue that any grid friendly dynamic pricing scheme must differentiate between residential and commercial areas. They conclude that incentive instruments have to capture the distinct consumption typologies and patterns that appear across distribution networks. While [27] provide the only known analysis of incentive mechanisms across different voltage levels, their study is limited to transformer-level impacts and omits an explicit grid model. [28] emphasize that considering rebound effects is substantial for incentive-based demand-side management. If physical properties, such as catch-up effects are not considered, incentives for load reduction might not resolve a congestion, but shift it to another time step. To capture the impact of customers' demand response behavior, [29] incorporated psychological data into a machine learning model. While the study does not address congestion management, it highlights the critical need to evaluate the risks that non-participation poses to grid operation. There is no grid model, but it does have a TSOs perspective because of the focus on generation adequacy.

Pilot projects such as *OctoFlexBW* [30] and *Grids and Benefits* [31] allow German transmission system operators to experiment with electric mobility flexibility for transmission grid operation. While OctoFlexBW provides flexibility via market-based platforms, Grids and Benefits steers flexibility indirectly via dynamic grid fees. Findings indicate that dynamic grid fees cause load-shifts of up to 20%. Since the topology of the transmission grid and the temporal generation mix differ markedly across regions, the project underlines the need for region-specific grid-fee designs that are calibrated to the characteristics of each grid area.

Market outcomes and welfare distribution

[32] formulate a bi level model that illustrates how a strategic aggregator's bidding behavior reshapes the residual demand curve during

[4] [20] argued that wholesale prices generally do not cause congestion since most studies are highly simplified and assume a single dynamic price for all consumers. In reality, however, a variety of dynamic prices are offered, which slightly differ in their ability to steer flexible demand.

endogenous market clearing when exogenous grid fees are present. The analysis, however, focuses primarily on the aggregator's procurement costs, while macroeconomic implications on residual demand are addressed downstream. A recent study examines how grid serving tariff designs interact with industrial battery storage operation [33]. While it provides detailed modeling of individual company load behavior using more than 800 real industrial profiles, it does not include a physical network model and therefore cannot assess grid congestion, voltage issues, or transformer loading. The results show that tariff incentives can oversteer load shifting and even create new peak loads, particularly under atypical usage provisions or tariff reforms without capacity components. The authors also highlight regressive distributional effects: grid fee reductions often exceed the actual cost savings for the system, allowing some firms to reduce their grid payments almost entirely. As a result, inflexible consumers effectively cross subsidize more flexible industrial users. [34] also shows distributional impacts of cost-reflective grid tariffs. The authors highlight, that efficient pricing designs may unintentionally burden passive, low-usage consumers by charging them similarly to high-usage flexible users, who often have greater financial means. They also show that when dynamic price components or side payments are introduced at the market clearing stage, the resulting shifts in flexible demand can alter wholesale market outcomes. If adopted at scale, such behavior can significantly affect clearing prices, increasing electricity procurement costs for less flexible consumers. Building on this, [35] examine the interplay between market interventions and system-level outcomes, emphasizing that effective incentive design must carefully balance potential savings in grid operation and reinforcement costs against possible increases in wholesale market expenditures. Beyond the implications of a single market zone, [36] analyze spillover effects of grid fee designs in coupled market zones. They argue that grid tariffs that incentivize inefficient storage investments (regarding wholesale market clearing), affect arbitrage and net imports. The study builds on [37], who set up a stylized equilibrium model in which consumers strategically shift grid costs to one another. Neglecting grid constraints, [36] compare consumer costs and investments and found that welfare-related spillover effects can be either positive or negative depending on the tariff combination in both market zones.

Literature synthesis and research contribution

The existing literature provides extensive techno-economic analyses of incentive-based mechanisms for congestion management, particularly in distribution networks, highlighting their potential to enable grid-supportive behavior from flexible consumers. However, most studies focus on distribution-level applications and pay limited attention to transmission systems, where the economic implications of load shifting can be substantially larger. At this level, flexible demand involves higher volumes and interacts directly with wholesale electricity markets, implying that demand shifts may significantly affect market clearing prices and dispatch outcomes.

Capturing these interactions requires analytical approaches that move beyond models with exogenous prices and instead explicitly account for the endogenous interaction between incentive design and market responses. Unlike many market models that abstract from grid constraints, this paper contributes to the literature at the intersection of incentive design, techno-economic system analysis, and macroeconomic market modeling. It co-optimizes alternative designs for dynamic grid charges, transmission grid operation, and market clearing within a bi-level framework. This framework allows for a quantitative assessment of welfare effects across supply, demand, and transmission system operators.

The study addresses the following research questions:

(i) How effective are incentive-based mechanisms in supporting transmission system operation and alleviating grid congestion?
(ii) How economically efficient are alternative incentive design options?
(iii) How do grid incentives interact with wholesale market outcomes from a welfare perspective?
(iv) What policy implications arise for future congestion management, particularly in the German transmission system?

The remainder of the paper is structured as follows. Section 3 introduces the modeling framework. Section 4 presents a stylized case study based on an IEEE test system adapted to reflect the future German grid. Section 5 analyzes system-level and stakeholder-specific outcomes of different incentive schemes. Section 6 concludes with regulatory insights for the design of future congestion management strategies.

3. Methodology

Modeling electricity systems can be approached in numerous ways, taking into account the various regulatory alternatives and market clearing principles used across wholesale electricity markets. The modeling framework described below presents a stylized setting aligned with German market principles.

Nomenclature

Sets and Indices	
$n \in N$	Grid nodes
$reg \in REG$	Regions
$t \in T$	Time steps
$n' \in N'$	Nodes with heat pumps
$p \in P$	Plants
$l \in L$	Power lines
Parameters	
$h^{mca,*}_{n',t}$	Heat pump dispatch (market equil.)
$slp^{H}_{n',t}$	Standard load profile
$c^{G}_{p,t}$	Generation capacity
oc_p	Operational cost
c^{th}_{l}	Thermal line capacity
M	Large positive constant
$g^{mca,*}_{p,t}$	Power generation (market equil.)
$dem^{H}_{n',t}$	Heat pump energy demand
$c^{H}_{n'}$	Heat pump limit
$d_{n,t}$	Inflexible electricity demand
$ptdf_{l,n}$	Power transfer distr. factors
ϵ	Flexibility range
Primal variables	
$FP^{tso}_{reg,t}$	Flexibility premium
C^{agg}	Aggregator costs
C^{rd}	Redispatch cost
$G^{mca}_{p,t}$	Power generation
$H^{agg}_{n',t}$	Heat pump dispatch (agg.)
$H^{tso,+}_{n',t}$	Heat pump increase
$G^{tso,+}_{p,t}$	Redispatch up
$NI_{n,t}$	Nodal injection
C^{d}	Dispatch cost
C^{flex}	Flexibility cost
C^{cm}	Congestion management cost
$H^{mca}_{n',t}$	Heat pump dispatch (mca)
$DEM^{H^{mca}}_{n',t}$	Heat pump endog. demand (mca)

$H_{n',t}^{tso,-}$	Heat pump decrease
$G_{p,t}^{tso,-}$	Redispatch down
$LF_{l,t}$	Line flow
Dual variables	
Π_t^{mc}	Market clearing price
$\underline{\eta}_{n',t}, \bar{\eta}_{n',t}$	Heat pump limits
$\bar{o}_{p,t}, \underline{o}_{p,t}$	Generation constraints
$\underline{\theta}_{n',t}, \bar{\theta}_{n',t}$	Flexibility constraints
Binary variables	
$B_{n',t}^{\eta}$	Heat pump (lower bound)
$B_{n',t}^{\theta}$	Heat Pump (lower b. corridor)
$B_{p,t}^{o}$	Generator (lower bound)
$B_t^{\Pi^{mc}}$	Market clearing
$B^{\bar{\eta}}$	Heat pump (upper bound)
$B_{n',t}^{\bar{\theta}}$	Heat Pump (upper b. corridor)
$B_{p,t}^{\bar{o}}$	Generator (upper bound)

3.1. Modeling framework

First, a zonal market clearing and associated redispatch problem is formulated and sequentially solved (Section 3.2). This model simulates market clearing and congestion management based on the current market design in Germany, providing insights into how alternative, incentive-based market designs can impact market outcomes. In the model, a central market clearing agent aims to minimize total system costs by controlling both generation dispatch and the dispatch of flexible demands. The resulting allocation corresponds to a competitive equilibrium and is thus Pareto efficient, consistent with the First Welfare Theorem. It is equivalent to the outcome that would emerge if individual aggregators and generators independently optimized their own costs within a competitive market framework. With the future electricity system expected to feature significant amounts of flexible demand-side technologies due to sector coupling, Section 3.3 details the modeling of these flexible demand-side applications used in the mentioned sequential optimization approach and explains how flexibility is used to minimize corresponding procurement costs.

Given the objective of this work to explore incentive mechanisms for grid-serving demand behavior, the modeling framework is extended to include the decision-making processes for designing efficient grid load incentives. Generally, incentive mechanisms can be implemented at different energy trading stages: before, during, or after day-ahead market clearing. This work assumes incentives are provided to market stakeholders before the market clearing stage, close to current policies, e.g., dynamic electricity tariff components. To enable incentivized flexible dispatching, which diverges from a purely market-optimized approach, an aggregator is introduced as a new actor. The aggregator can be conceptualized as a collective of competitors operating in a purely rational manner, which allows for its implementation as a single actor. It can respond to given incentives while adhering to flexibility constraints introduced in Section 3.3. This extended framework (Section 3.4) enhances the market clearing and redispatch problem by integrating regulatory decisions and is described using a game-theoretic Stackelberg framework formulated as a bi-level program.

Capital letters denote model-endogenous decisions in all mathematical expressions, while parametric inputs are represented in lower case. Superscripts identify the actor controlling specific variables or further specify model entities. Greek letters are used for dual variables.

3.2. Model I - Sequential linear optimization

Zonal market clearing problem

The following model represents a market clearing agent (mca) responsible for determining the least-cost generation dispatch to satisfy all demands. For enhanced comprehension, the model's equations and inequalities have been simplified to their essentials, following standard textbook formulations:

$$min \quad C^d = \sum_{p,t} G_{p,t}^{mca} \cdot oc_p \tag{i.1}$$

$s.t.$

$$\sum_n d_{n,t} + \sum_{n'} H_{n',t}^{mca} - \sum_p G_{p,t}^{mca} = 0 \qquad : \Pi_t^{mc} \qquad \forall t \tag{i.2}$$

$$0 \le G_{p,t}^{mca} \le c_{p,t}^G \qquad \forall p,t \tag{i.3}$$

$$0 \le H_{n',t}^{mca} \le c_{n'}^H \qquad \forall n',t \tag{i.4}$$

$$dem_{n',t}^H - \epsilon \le \sum_{\hat{t} \le t} H_{n',\hat{t}}^{mca} \le dem_{n',t}^H + \epsilon \qquad \forall n',t \tag{i.5}$$

The objective function (i.1) determines the cost-minimizing dispatch $G_{p,t}^{mca}$, over all time steps $t \in T$ and power generating units $p \in P$, considering their corresponding generation costs oc_p. The energy balance (i.2) ensures that, for all $t \in T$, the inflexible demand $d_{n,t}$ of all nodes $n \in N$ plus the flexible demand $H_{n',t}^{mca}$ of all nodes $n' \in N' \subset N$ is met in total. Constraint (i.3) limits the generated electricity of each power plant in each time step to its available capacity $c_{p,t}^G$. Eqs. (i.4) and (i.5) describe the modeling of the characteristics of flexible demands (see Section 3.3). The (zonal) market clearing price in each time step, Π_t^{mc}, is derived from the dual variable of the energy balance (i.2).

Grid constraints are not considered when computing the market equilibrium according to German market rules. Consequently, the resulting dispatch may cause grid congestion, which is resolved in the subsequent congestion management optimization.

Congestion management problem

The outputs of the market clearing problem serve as inputs for the congestion management optimization. The following system of equations represents the optimization problem of a transmission system operator (tso), who adjusts the market clearing dispatch considering load flow constraints in a cost-minimizing manner subsequently to the market clearing, commonly referred to as redispatch:

$$min \quad C^{rd} = \sum_{p,t} \left(G_{p,t}^{tso,+} - G_{p,t}^{tso,-} \right) \cdot oc_p \tag{i.6}$$

$s.t.$

$$\sum_{p(n)} \left(g_{p,t}^{mca,*} + G_{p,t}^{tso,+} - G_{p,t}^{tso,-} \right) - h_{n'(n),t}^{mca,*} - H_{n'(n),t}^{tso,+} + H_{n'(n),t}^{tso,-} - d_{n,t} - NI_{n,t} = 0 \qquad \forall n,t \tag{i.7}$$

$$\sum_n NI_{n,t} = 0 \qquad \forall t \tag{i.8}$$

$$LF_{l,t} = \sum_n \left(NI_{n,t} \cdot ptdf_{l,n} \right) \qquad \forall l,t \tag{i.9}$$

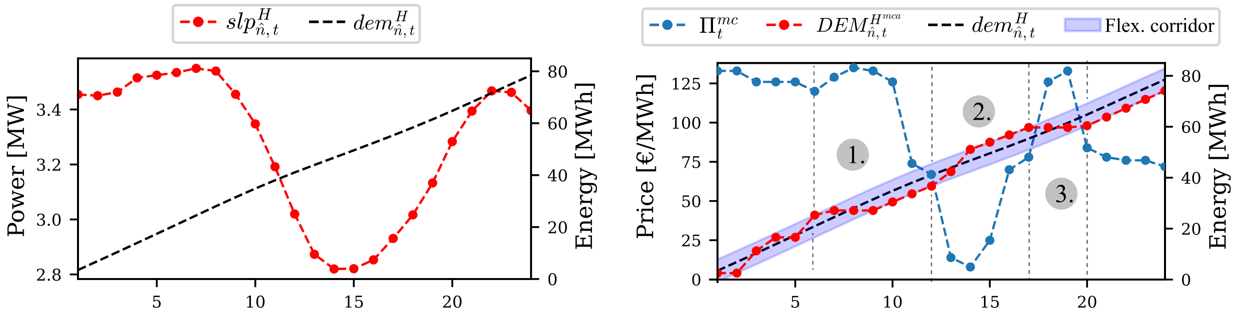
$$-c_l^{th} \le LF_{l,t} \le c_l^{th} \qquad \forall l,t \tag{i.10}$$

$$0 \le G_{p,t}^{tso,+} \le c_{p,t}^G - g_{p,t}^{mca,*} \qquad \forall p,t \tag{i.11}$$

$$0 \le G_{p,t}^{tso,-} \le g_{p,t}^{mca,*} \qquad \forall p,t \tag{i.12}$$

$$0 \le h_{n',t}^{mca,*} + H_{n',t}^{tso,+} - H_{n',t}^{tso,-} \le c_{n'}^H \qquad \forall n',t \tag{i.13}$$

$$dem_{n',t}^H - \epsilon \le \sum_{\hat{t} \le t} \left(h_{n',\hat{t}}^{mca,*} + H_{n',\hat{t}}^{tso,+} - H_{n',\hat{t}}^{tso,-} \right) \le dem_{n',t}^H + \epsilon \qquad \forall n',t \tag{i.14}$$



(a) Standard load profile and total electricity consumption for the aggregated pool of heat pumps located at node $\hat{n}$ (24 h extraction).

(b) Flexibility corridor utilization vs. market clearing price at node $\hat{n}$ (24 h extraction).

Fig. 3. Derivation of the heat-pump-related flexibility corridor at node $\hat{n}$.

Power flow constraints are imposed by incorporating a transmission topology network and adding power flows resulting from power injection at each node using a DC load flow approximation. The nodal injection $NI_{n,t}$ is computed by subtracting the demands $d_{n,t}$ and $h^{mca,*}_{n,t}$ from the market clearing dispatch $g^{mca,*}_{p,t}$ assigned to each node n and in each time step t (i.7). Generation dispatch can be adjusted through simultaneous positive redispatch ($G^{tso,+}_{p,t}$) and negative redispatch ($G^{tso,-}_{p,t}$) of generation quantities or by temporal load shifting ($H^{tso,+}_{n',t}, H^{tso,-}_{n',t}$). Constraint (i.8) ensures a balanced system-wide energy balance. Power transfer distribution factors ($ptdf$) are used to compute line flows $LF_{l,t}$ on each line l and in each time step t, resulting from the nodal injections (i.9). Line flows are constrained by line-specific thermal transfer capacities (i.10).

Constraints (i.11), (i.12), (i.13), and (i.14) restrict the use of redispatch and load shifting capacities according to their market dispatch. For a detailed description of the modeling of load shifting (i.14), refer to Section 3.3. The objective function (i.6) aims to minimize only redispatch costs based on the incremental fuel cost of the power generating units and does not include explicit cost components for load-shifting. However, these are indirectly included via inter-temporal redispatch considerations. For enhanced comprehension, the following example is provided.

Example 3.1. Any load shift from time step t_i to t_j necessitates a corresponding generation redispatch to maintain the system-wide energy balance (i.8). Specifically, if the TSO reduces the load in t_i ($H^{tso,-}_{n',t_i} > 0$), a simultaneous negative redispatch ($G^{tso,-}_{p,t_i} > 0$) is required to preserve power equality. Conversely, if the TSO increases the load at t_j ($H^{tso,+}_{n',t_j} > 0$), a positive redispatch ($G^{tso,+}_{p,t_j} > 0$) is necessary to meet the additional load in t_j. Thus, load shifting indirectly contributes to the objective function value through the incremental fuel cost of conventional generators associated with the inter-temporal redispatch. The TSO evaluates the economic implications of both conventional redispatch and load shifting, selecting the most economically advantageous alternative.

Note that in current real world market designs, flexible loads cannot be controlled by the grid operator and flexibility is only leveraged during the market clearing stage, which can be represented by fixing $H^{tso,+-}_{n',\hat{t}}$ to zero. However, in Section 4, $H^{tso,+-}_{n',\hat{t}}$ are also treated as free variables to benchmark the market design expansions discussed therein.

3.3. Modeling of demand side-flexibility

This study explores incentive mechanisms for enabling flexibility in load applications connected to lower-voltage grids (Eqs. (i.4)–(i.5)), which are then made available to transmission system operations (Eqs. (i.13)–(i.14)). These applications are aggregated into a technological bundle and managed collectively by aggregators. To illustrate the impact of such flexibility on transmission system operations, we focus on heat pumps, which feature a high operational simultaneity allowing them to be represented without significant aggregation error. Moreover heat pumps are expected to become a major source of demand-side flexibility in future German energy markets in the light of ongoing decarbonization and sector coupling initiatives.

The following sections detail the modeling of demand-side flexibility used in this research. Flexible load applications cannot be dispatched entirely flexibly. For example, interruptions in operating heat supply technologies, such as heat pumps, may result in comfort losses. The model-based representation of load applications considers operational constraints to account for these limitations, utilizing a flexibility corridor derived from the application's standard load profile. While heat pumps are the primary example of a flexible load application in this work, the methodology introduced here can be easily applied to other flexible demand-side technologies. However, for reasons of clarity, here the analysis is limited to heat pumps.

Definition of standard load profiles

The grid model distinguishes between various heat pump pools across different nodes. The subset $N' \subset N$ represents all nodes where demand can be partially flexibilized. The standard load profile for a heat pump, $slp^H_{n',t}$, is based on [38] and is used to calculate the power consumption of heat pumps aggregated as a pool per node. The profile is generic and derived based on assumptions about the ambient temperature over the course of a day. To create a consistent scenario, To create a consistent scenario, we scaled them to match projected heat pump deployment in Germany see the case study in Section 4. Each aggregated pool consists of an exogenous demand component and a flexible share, the latter of which is determined by model-endogenous optimization. The exogenous components are calculated by summing the aggregated standard load profile over all time steps $\hat{t}$ until each time step t at each node:

$$dem^H_{n',t} := \sum_{\hat{t} \le t} slp^H_{n',t} \qquad \forall n', t. \tag{1}$$

Fig. 3(a) illustrates the operational characteristics of the heat pump pool at an exemplary transmission grid node. The red line displays the aggregated standard load profile, while the dashed black line represents the total electricity consumption (1).

Introduction of flexibility constraints

In the following, constraints are introduced that describe the flexible components of the heat pump demand tied to the standard load profile. The dispatch of each pool of heat pumps, $H^{mca}_{n',t}$, is constrained by its

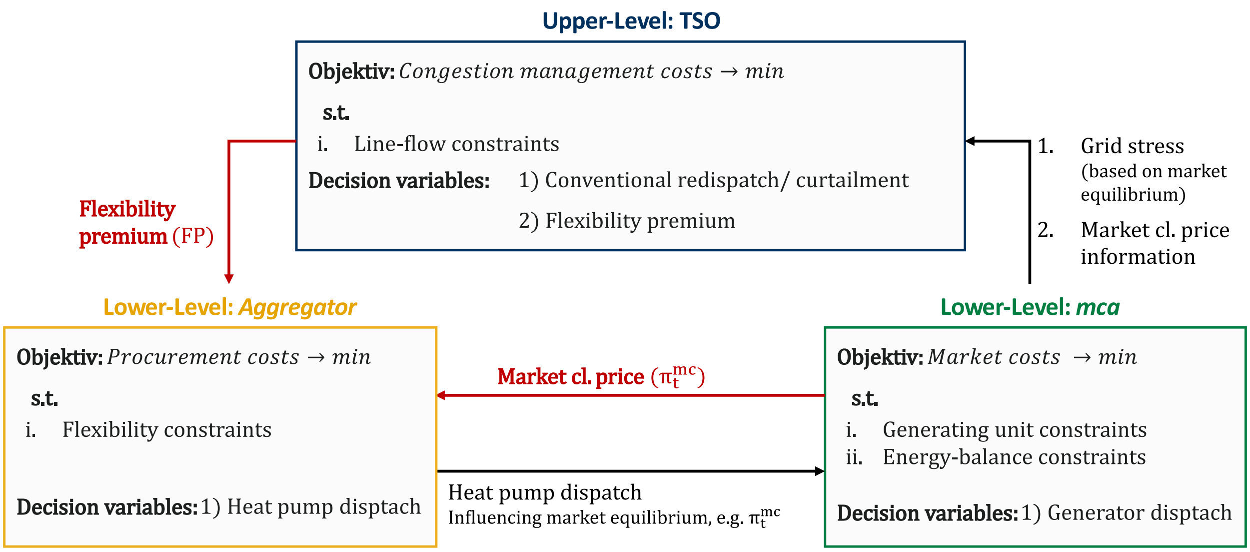


Fig. 4. Bi-level modeling framework for incentive-based congestion management via flexibility premiums: Financial incentives intersect with market clearing prices (red).

electrical power capacity $c^H_{n'} \in \mathbb{R}^+$ at each node and in each time step:

$$0 \leq H^{mca}_{n',t} \leq c^H_{n'} \qquad \forall n', t \tag{2}$$

To allow for flexible dispatch within a defined range, a corridor based on the summation of the standard load profile (1) is introduced. Within these boundaries, the heat pump can be dispatched flexibly

$$dem^H_{n',t} - \epsilon \leq \sum_{\hat{t} \leq t} H^{mca}_{n',\hat{t}} \leq dem^H_{n',t} + \epsilon \qquad \forall n', t \tag{3}$$

where the scalar $\epsilon \in \mathbb{R}^+$ denotes the boundaries of the corridor. In simple terms, these boundaries determine the maximal duration of operation interruption that is considered feasible for avoiding any reduction in the utility resulting from applying the technical application, e.g., comfort losses.

Fig. 3(b) displays the model-endogenous dispatch of the aggregated heat pump pool after solving the market clearing problem. This illustrates how the market's driving forces affect the dispatch of flexible applications. The blue area surrounding the total electricity consumption (black dashed line) describes the flexibility corridor. The dashed blue line represents the model-endogenous market clearing price (dual value of energy balance). The dashed red line refers to the flexible dispatch of the heat pump pool constrained by the flexibility corridor. Note that $DEM^{H^{mca}}_{n',t} := \sum_{\hat{t} \leq t} H^{mca}_{n',\hat{t}}$ in Fig. 3(b) for improved readability.

The following three observations explain how the flexibility inherent in the heat pump pool can minimize the aggregator's electricity purchase costs:

1. Prices are at a local maximum → electricity consumption is at the lower bound.
2. Prices are low → electricity consumption increases toward the upper bound.
3. Prices rise to a local maximum → electricity consumption decreases again.

The model-endogenous optimization of the heat pump dispatch yields a deviation of the dispatch profile from the standard load profile. The wholesale electricity price formation drives this deviation which minimizes the aggregator's electricity purchase costs to supply his customers with electricity.

This study, however, explores the potential for utilizing the flexibility of the aggregated pool in a grid-serving capacity, potentially diverging from the cost-minimal market dispatch. This deviation may be initiated through a direct control signal as seen in *Model I* (Section 3.2) or by an offered incentive, reducing the purchasing costs in otherwise more expensive hours (*Model II*, Section 3.4). In both cases, the aforementioned flexibility constraints must still be observed.

3.4. *Model II - Bi-level modeling framework*

The following model extends the sequential market clearing and congestion management problem (Section 3.2) by incorporating model-endogenous aggregator optimization and incentive design. This model enables the quantification of the interactions between the market clearing and redispatch stage, assuming the zonal market design (Model I) is expanded to include an incentive-based market intervention to promote grid-serving demand-side behavior from flexible loads. The modeling framework includes three players whose decision levels feature a nested structure. At the upper-level, the TSO optimizes all expected grid congestion resulting from the market clearing dispatch. The lower-level players comprise the market clearing agent and the aggregator. This bi-level framework allows for studying different incentive mechanism design options, considering the aggregator's reactions, including its impact on market equilibrium and its implications for grid operation. Fig. 4 illustrates the decision levels and players involved.

3.4.1. *Upper-level problem*

The upper-level represents the decision-making process of a regulated TSO responsible for grid operation. To resolve grid congestion, the TSO has two options:

1. Adjust the market clearing dispatch of conventional generators through simultaneous redispatch. This action only impacts the supply side and does not alter the dispatch of heat pumps. The cost associated with conventional redispatch, C^{rd}, is similar to that in *Model I* (see Eq. (i.6)):

$$C^{rd} := \sum_{p,t} \left(G^{tso,+}_{p,t} - G^{tso,-}_{p,t} \right) \cdot oc_p \tag{ii.1}$$

2. Pay a premium to the aggregator to incentivize load-shifting in a manner that serves the grid. The cost for this flexibility provision, C^{flex}, depends on the incentivized load-shifting quantities and the corresponding premium:

$$C^{flex} := \sum_{n',t} H^{agg}_{n',t} \cdot FP^{tso}_{reg(n'),t}. \tag{ii.2}$$

The premium paid can be considered a time- and region-varying price component that reduces the aggregator's wholesale cost

$$FP^{tso}_{reg,t} \geq 0 \tag{ii.3}$$

which is defined over all regions $reg \in REG$ and time steps $t \in T$, with each node mapped to a specific region

$$reg : N \to REG; \qquad n \mapsto reg(n). \tag{ii.4}$$

The objective function of the upper-level problem (ii.5) minimizes the total congestion management costs resulting from conventional redispatch and flexibility provision. This objective reflects the role of a cost-regulated TSO, which determines flexibility premia to minimize the total congestion management cost, defined as the sum of flexibility costs (premium payments) and any remaining redispatch costs. Since these expenditures are ultimately recovered through regulated network tariffs, minimizing the congestion management cost is equivalent to minimizing the corresponding network tariff burden.

$$min \quad C^{cm} := C^{rd} + C^{flex} \tag{ii.5}$$

$$s.t.$$

$$\sum_{p(n)} \left(G^{mca}_{p,t} + G^{tso,+}_{p,t} - G^{tso,-}_{p,t} \right) - H^{agg}_{n'(n),t} - d_{n,t} - NI_{n,t} = 0 \quad \forall n,t \tag{ii.6}$$

$$\sum_{n} NI_{n,t} = 0 \quad \forall t \tag{ii.7}$$

$$LF_{l,t} = \sum_{n} \left(NI_{n,t} \cdot ptdf_{l,n} \right) \quad \forall l,t \tag{ii.8}$$

$$-c^{th}_{l} \leq LF_{l,t} \leq c^{th}_{l} \quad \forall l,t \tag{ii.9}$$

$$0 \leq G^{tso,+}_{p,t} \leq c^{G}_{p,t} - G^{mca}_{p,t} \quad \forall p,t \tag{ii.10}$$

$$0 \leq G^{tso,-}_{p,t} \leq G^{mca}_{p,t} \quad \forall p,t \tag{ii.11}$$

Constraints (ii.7), (ii.8), (ii.9), (ii.10), and (ii.11) remain unchanged compared to the congestion management problem formulated in *Model I* (see Section 3.2). The main difference in this model version is that generation and heat pump dispatch are no longer parameters but variables controlled by the associated lower-level players. Since the grid-serving operation of the heat pumps is now incentivized by the TSO during the market clearing stage, the load-shifting variables in the nodal energy balance (ii.6) are no longer controlled by the TSO. Therefore, this model version has no positive and negative adjustment variables for the heat pump redispatch.

3.4.2. Lower-level problems

Aggregator: The following system of equations and inequalities represents the decision process of the aggregator who minimizes his energy procurement costs, whereby dual variables are presented at the right hand side of the constraints in greek letters:

$$\min \quad C^{agg} = \sum_{n',t} H^{agg}_{n',t} \cdot \left(\Pi^{mc}_{t} - FP^{tso}_{reg(n'),t} \right) \tag{ii.12}$$

$$\text{s.t.}$$

$$0 \leq H^{agg}_{n',t} \leq c^{H}_{n'} \qquad : \underline{\eta}_{n',t}, \bar{\eta}_{n',t} \forall n',t \tag{ii.13}$$

$$dem^{H}_{n',t} - \epsilon \leq \sum_{\hat{t} \leq t} H^{agg}_{n',\hat{t}} \leq dem^{H}_{n',t} + \epsilon \qquad : \underline{\theta}_{n',t}, \bar{\theta}_{n',t} \forall n',t \tag{ii.14}$$

The aggregator's objective function includes the wholesale market procurement cost based on the time-varying market clearing price Π^{mc}_{t} and flexibility provision revenues based on the flexibility premium $FP^{tso}_{reg(n'),t}$ (ii.12). These revenues reduce the total procurement cost, whilst incentivizing a grid-serving dispatch of flexible load applications, assuming rational behavior that leads to a cost-minimizing purchase strategy. However, the dispatch of the aggregated pool of heat pumps is still subject to the flexibility constraints outlined in Section 3.3, specifically Eqs. (ii.13)–(ii.14).

Market clearing agent: During market clearing, the operator aims to dispatch generators at minimal costs to meet all inflexible electricity demands and align with the aggregator's procurement strategy. The following system of equations and inequalities represents the optimization problem of the clearing agent:

$$\min \quad C^{d} = \sum_{p,t} G^{mca}_{p,t} \cdot oc_{p} \tag{ii.15}$$

$$\text{s.t.}$$

$$0 \leq G^{mca}_{p,t} \leq c^{G}_{p,t} \qquad : \underline{o}_{p,t}, \bar{o}_{p,t}, \quad \forall p,t \tag{ii.16}$$

$$\sum_{n} d_{n,t} + \sum_{n'} H^{agg}_{n',t} - \sum_{p} G^{mca}_{p,t} = 0 \qquad : \Pi^{mc}_{t} \quad \forall t \tag{ii.17}$$

The generation dispatch is constrained by the available generation capacity (ii.16). The total generation must satisfy all demands in the power balance (ii.17). Since the operation of heat pumps across the nodes, $\sum_{n'} H^{agg}_{n',t}$, is now a decision variable in the aggregator's optimization problem, affected by the incentives provided by the TSO, the flexible demand levels may differ from those observed in *Model I* (see Section 3.2). Depending on the TSO's incentivization strategy to handle grid congestion, wholesale market outcomes may be altered, influencing market prices Π^{mc}_{t}.

3.4.3. Overcoming nested structures, non-linearities, and the ill-posed nature of bi-level problems

The bi-level problem cannot be solved with ordinary MIP solvers due to its nested structure and several non-linearities in its current form. For these reasons, Appendix B illustrates how the bi-level program is reformulated into a single-level mixed-integer linear program using a twofold application of the theorem of strong duality and a big-M approach.

However, the bi-level problem as formulated is ill-posed, as the aggregator's optimization problem does not yield a unique solution set (see Appendix C for further details). Thus, solving the reformulated single-level problem yields optimal lower-level solutions concerning the upper-level objective function, commonly referred to as *optimistic approach* (see, e.g., [39] for a detailed description). This may alter the comparability of the results computed in different model runs regarding the incentive mechanism design options investigated in the case study conducted here (see Section 4). Regularization techniques applied to mitigate such optimistic bias artifacts are outlined in Appendix C.

4. Case study

The presented models are applied to an extended IEEE test system within a market scenario framework to represent various incentive design alternatives. The following section describes the techno-economic key parameters of the system configuration and each grid management concept investigated.

4.1. System configuration

The transmission topology consists of a 118-bus high-voltage grid with high renewable penetration. The test system, adapted from [40], does not represent the actual German transmission grid, but captures its characteristic north-south imbalances between wind generation in the north and load centers in the south. It was adjusted to match Germany's generation capacity projected in the TYNDP 2022 *distributed energy scenario* for 2030. However, it was scaled down based on the ratio between the demand levels in the test system and the scenario report to align with the capacity of the transmission system. Fig. 5 provides a graphical representation of the test system including the grid topology and the generation structure, and Table 1 lists all key parameters.

The system also assumes significant electrification of the residential heating sector, with heat pumps being the primary heat source. These heat pumps are allocated across grid nodes according to nodal electricity demands. The model assumes a conservative share of heat pumps in

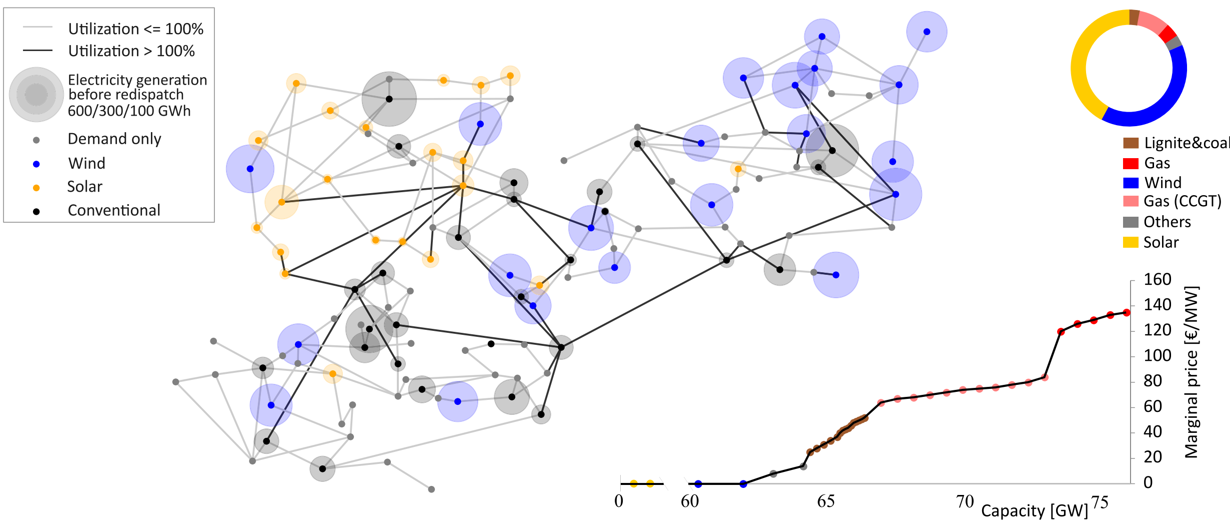


Fig. 5. Techno-economic system configuration.

Table 1
Key technological system configuration parameter.

Grid	Generation	Demand
card(N) = 118	Capacity = 76 050 MW	$\varnothing_{n,t} dem_{n,t}$ = 158.0 MW
card(N') = 114	Share RES cap. = 82%	$\varnothing_{n',t} H^{mca,agg}_{n',t}$ = 4.6 MW
card(L) = 317	Share conv. cap. = 18%	$c^H_{n'}$ = 8.74 MW
–	–	$\epsilon = \frac{c^H_{n'}}{2}$ = 4.37 MW
–	–	Share flexible load = 2.9%

Table 2
Overview on grid management concepts' key characteristics.

Concept	Model	Flex. utilization	CM-Instruments	Prem. characteristics	
				Spatial	Temporal
UC	I	–	RD	–	–
PO	I	Market	RD	–	–
PO.DC	I	Market + CM	RD + direct control	–	–
PO.P– (a)	II	Market + CM	RD + premium (a)	Nodal	Time-varying
PO.P– (b)	II	Market + CM	RD + premium (b)	Regional	Time-varying
PO.P– (c)	II	Market + CM	RD + premium (c)	Nodal	Time-static

the total electricity demand of 2.9%, which corresponds to a total electrical capacity of 8.74 MW. This is slightly lower than the 5% forecast for Germany (2030) in studies such as [41] or [42]. This lower estimate reflects the assumption that demand-side flexibility is not fully utilized in every household. The heat pump dispatch is parameterized to allow interruption for up to 1 h under full load (see the flexibility corridor ϵ introduced in Section 3.3). The model's temporal horizon spans 28 days, with an hourly resolution. To simplify the solving procedure, the model employs a rolling planning approach with additional foresight to handle inter-temporal constraints such as load-shifting (local time-coupling). Rolling planning has been widely used in previous studies to solve congestion management problems similar to the model developed in this work that does not have global time-coupling constraints (For a detailed discussion on different solving procedures for congestion management problems see [43]). *Model I*, being a linear program (LP), is solved using CPLEX, while *Model II* utilizes ODH-CPLEX. Both models are implemented in the General Algebraic Modeling Language (GAMS).

4.2. Control instruments and premium concepts

Instruments incentivizing load-shifting can be designed with varying spatial and temporal degrees of granularity. Given the complexity associated with each design in real-world market applications, this study examines various design concepts and compares their implications to benchmark cases. While instruments with high spatial and temporal resolution may be too complex to implement, reducing granularity increases their practicality from an operational perspective. Table 2 summarizes the key features of the grid management concepts distinguished in this work.

The first set of concepts assumes that aggregators do not face any incentives for grid-serving load behavior, which is correlated with *Model I*. This set helps interpret general market interactions resulting from grid management practices and enables comparing them with the design alternatives including incentive instruments. The first concept considers that aggregators do not respond to wholesale market prices and the flexibility of heat pumps is not utilized by TSOs for grid optimization ($\epsilon = 0$). This market framework can be modeled through sequential market clearing and redispatch optimization (*Model I*), with heat pump operation set as fixed parameters. The framework yields the highest market costs, due to the economically inefficient dispatch of the heat pump. Since heat pump operation is not optimized during market clearing and congestion management, it is referred to as *uncontrolled* (**UC**).

Relaxing heat pump operation constraints during market clearing reduces electricity procurement costs resulting from the more efficient dispatch associated with the flexibility. A second concept considers heat pump operation as flexible within the boundaries outlined in Section 3.3 during the market clearing stage, assuming profit-maximizing aggregator behavior. However, grid operators cannot access heat pump flexibility to optimize grid operation, resulting in high congestion management costs and serving as an upper bound against which all incentive instruments can be benchmarked. This scenario is referred to as *price-optimized* (**PO**). Additionally, concept **PO.DC** assumes a flexible operation of heat pumps during the congestion management stage, controlled by TSOs. This concept is expected to yield the lowest congestion management costs, serving as a lower bound to benchmark incentive instruments, but involves very high information and communication requirements and legal implications. Since heat pump operation is considered *price-optimized* and TSOs are assumed to have *direct control*, this concept is referred to as **PO.DC**.

A second set of grid management concepts is explored, where aggregators receive a side payment known to them prior to the market clearing stage. This payment incentivizes grid-serving load behavior and reduces wholesale energy procurement costs. Due to the endogenous optimization by aggregators in response to these incentives, *Model II*, as described previously, is employed for their study. Three grid

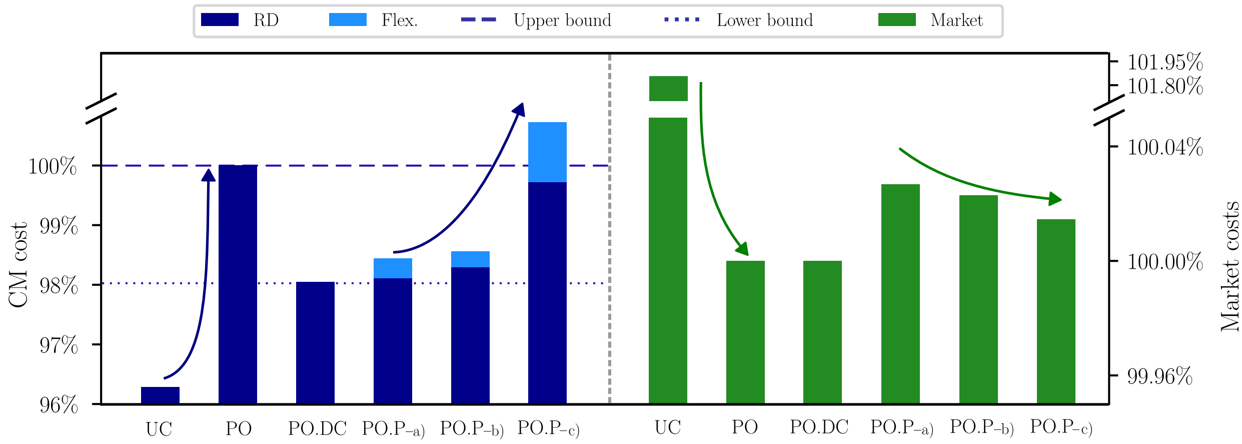


Fig. 6. *Left panel*: Congestion management costs. **PO** (no TSO control) and **PO.DC** (direct TSO control) establish upper and lower bounds, providing a benchmark for the incentive-based (indirect TSO control) measures. *Right panel*: Market costs indicate the degree of market intervention due to the incentive.

management concepts that incentivize aggregator load application are investigated, denoted as **(a)**, **(b)**, and **(c)**.

In concept **PO.P– (a)**, aggregators are offered a time- and location-varying premium, i.e. each region corresponds to a single node (cf. (ii.3) and (ii.4)). This premium, a model-endogenous result of transmission system optimization, balances expected redispatch and load-shifting costs. Alternatively, concept **PO.P– (b)** merges adjacent transmission grid nodes into four regions, each receiving the same premium to simplify legal requirements like non-discriminatory grid and market access. These regions are defined based on hierarchical clustering of nodal prices, resulting in regions with similar grid load characteristics. See Fig. 11 in the annex for a graphical representation of the four grid regions.

Due to the time-varying characteristics, the previously described grid management concepts require regular premium updates. This complexity may hinder their practical application in real-world market practices. Therefore, concept **PO.P– (c)** assumes a time-static premium design. From a modeling perspective this means, that the offered premium is a parametric input rather than a decision variable. This premium is determined using a simple yet pragmatic heuristic based on the differences between nodal and zonal prices. A clustering algorithm identifies critical time windows during a day that consistently exhibit significantly higher nodal prices than zonal prices. During these windows, increased load would further stress the electricity grids beyond the levels induced by market clearing. The premium is calculated to cover the additional costs aggregators would incur when shifting their loads to adjacent time steps next to these critical time windows, whereby the time windows are kept constant over the model's planning horizon. To avoid excessive flexibility management expenses during single time steps, the premium considers a cap of 30 €/MWh. Though this concept relies on model-endogenous market clearing at a nodal and zonal clearing auction, it can be easily applied to historical market results.

5. Results

The presented grid management concepts are investigated using the previously introduced modeling frameworks within the IEEE test system. The following analysis focuses on market outcomes, beginning with dispatch, congestion management, and total system costs. From the perspective of the transmission system operator, the reported economic efficiency gains should be interpreted with respect to the regulated TSO's objective of minimizing the total congestion management cost, rather than maximizing social welfare. The congestion management cost comprises conventional redispatch costs and flexibility costs, i.e., the premia paid to incentivize consumer flexibility. Since these expenditures are ultimately recovered through regulated network tariffs, reductions in congestion management cost directly translate into a lower tariff burden for network users. However, unlike redispatch costs, which arise from additional fuel consumption, these premiums constitute transfer payments to participating consumers and therefore do not represent net costs from a system-wide welfare perspective. Consequently, flexibility premiums are considered only when evaluating congestion management from the TSO perspective, as discussed in Section 5.1.[5] Section 5.2 subsequently evaluates the grid management concepts from a system-wide perspective, where system costs are determined exclusively by fuel-related dispatch and redispatch costs. Finally, Section 5.3 provides a detailed analysis of incentives across time and location, before distributional effects and their implications are discussed in Section 5.4.

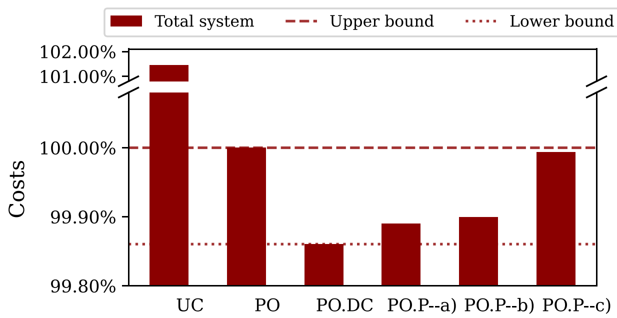
5.1. Market clearing dispatch, redispatch and flexibility costs

Incentivizing grid-serving load dispatch impacts both power plant dispatch and associated expenses, expressed as market clearing costs and congestion management costs. Fig. 6 summarizes the market outcomes observed in each investigated grid management concept. Note that market and congestion management costs are displayed on two different axes for better readability of the observed volumes. Due to the small flexible share of total demand, results are given in relative terms. The absolute values are listed in Fig. 7.

In concept **UC**, aggregators do not respond to market signals, leading to economically inefficient dispatch of heat pumps with respect to wholesale market outcomes. This concept results in the highest total market costs. However, the absence of a price signal reduces simultaneity, defined here as the number of heat pumps operating at full load in a given time step. Consequently, load peaks are comparatively low, thereby resulting in the lowest grid congestion and associated congestion management costs.

The flexible dispatch of heat pumps within the boundaries defined in Section 3.3 assumed in concept **PO** reduces market costs by 1.82%. Despite only 2.9% of the load being flexible and constrained by technical limitations, this decrease is significant. However, the price response assumed in this concept increases the simultaneity of

[5] Moreover, the TSO's cost structures *must* be disclosed to determine regulated network charges, through which the operating costs of the transmission system are ultimately allocated among network users.



	RD	Market	Sys.
UC	15.65	215.47	231.12
PO	16.26	211.55	227.80
PO.DC	15.92	211.55	227.49
PO.P– a)	15.95	211.60	227.61
PO.P– b)	15.98	211.60	227.62
PO.P– c)	16.21	211.58	227.95

Fig. 7. Total system costs: *relative* (left panel) and *absolute* (right panel) values.

heat pump dispatch, causing load peaks and grid congestion, thereby raising congestion management requirements and costs. Incentivizing aggregators to dispatch their loads in a grid-serving manner should decrease congestion management costs below the level observed in concept **PO**, serving as an upper benchmark for incentive instruments. In concept **PO.DC**, relaxing heat pump operation constraints during congestion management optimization reduces associated costs by 1.96%, making it the most economically efficient alternative from a grid operation cost perspective and thus presenting a lower bound to compare results with.

When aggregators are offered a premium to incentivize load-shifting before the market clearing stage, market expenses and load-shifting revenues influence the aggregator's cost-optimal electricity procurement strategy. The time- and location-varying premium assumed in concept **PO.P– (a)** most efficiently resolves grid congestion, yielding the lowest congestion management costs among all the alternatives in which incentive strategies are investigated. However, incentivizing load-shifting strongly intervenes in the resulting wholesale market equilibrium, increasing wholesale market costs beyond those observed in a purely price-optimized market dispatch. Concept **PO.P– (b)** results in slightly higher congestion management costs due to less efficient locational management of flexible quantities. Consequently, the reduced level of market intervention in favor of congestion management results in a slight decline in market costs. In summary, it is an efficient practice-oriented simplification of **PO.P– (a)**, given its proximity to **PO.P– (a)** itself and to the lower bound. Lastly, concept **PO.P– (c)** exhibits the least efficient management of load-shifting quantities for resolving grid congestion. Here, aggregators receive premiums even during time-steps when load-shifting has little or no value for grid operation, significantly decreasing economic grid management efficiency.

5.2. Total costs at a system level

Summarizing the results on the wholesale market and redispatch costs reveals an intricate opposing relationship regarding total fuel costs: the more efficiently grid congestion is handled by incentivizing grid-serving load behavior, the more market equilibria are affected, leading to increased power plant dispatch costs. From a regulatory perspective, any policy must be evaluated in terms of its overall system effect. Combining wholesale market and redispatch costs for each grid management concept, Fig. 7 displays the total system cost for supplying the total electricity demand in the IEEE test system. While direct control of load applications via a TSO yields the lowest system costs, concepts **PO.P– (a)** and **PO.P– (b)** still exhibit more economically efficient power supply compared to not utilizing flexibility during congestion management (concept **PO**), with only marginal cost increases in concept **PO.P– (b)**. However, a time-static premium, as assumed in concept **PO.P– (c)**, cannot be recommended based on the resulting system costs, which are only slightly lower than those of **PO.P– (b)**.

5.3. Incentive structure and premium volumes at a grid node level

The IEEE test system used in this study assumes high levels of renewable energy penetration, aligning with Germany's 2030 energy transition goals. It is thus crucial to consider not only the costs of congestion management and market operation but also the role of load-shifting as a congestion management strategy and its effects on renewable integration to interpret market outcomes under different grid management concepts. For concept **PO.P– (a)**, Fig. 8 illustrates statistical relationships between load-shifting activities and renewable generation across various hours within the modeling horizon displayed as daily values. A strong positive linear correlation (Pearson coefficient $r = 0.65$) between RES generation and flexibility costs, and a moderate correlation ($r = 0.48$) with incentivized hours is observed.

The results highlight an notable dependency: Aggregators are more strongly incentivized to dispatch their flexible loads grid-supportively on days with higher renewable generation. This observation holds also for the grid management concepts **PO.P– (b)** and **PO.P– (c)** (see Figs. 12 and 13 in the annex). Generally, renewable curtailment incurs very high congestion management costs due to the significant difference between the marginal costs of renewable generators and conventional power plants for corresponding upward redispatch. This relationship makes load-shifting particularly economically efficient in systems dominated by renewables, especially during hours with high renewable curtailment volumes. In fact, renewable curtailment decreased from *264.6 GWh* in concept **PO** to *259.4 GWh* in **PO.P– (a)** and *259.8 GWh* in **PO.P– (b)**. In summary, load shifting can be regarded as a dominant congestion management alternative, particularly in the context of reducing curtailment. However, given the relatively low flexible shares, it cannot be considered a complete substitute.

Renewable generation is a fluctuating electricity source due to its weather dependency, caused by variations in wind velocity and solar radiation over time and across regions. This variability necessitates congestion management strategies with precise temporal and locational effects. Fig. 9 shows the magnitude of successful flexibility premium offers accepted by aggregators across transmission grid nodes and time steps for the three grid management concepts that rely on incentivizing load shifting, as considered in this work.

The premium offers observed in concept **PO.P– (a)** highlight the significance of location- and time-varying premium designs (top plot of Fig. 9). The structure of premium payments is highly heterogeneous across nodes and over the modeling horizon, with no systematic patterns evident upon visual inspection. Aggregating nodes into regions results in less frequent premium payments (center plot). This approach, which causes incentivization costs at numerous grid nodes (merged into regions), leads to undesirable responses that increase overall flexibility management costs and reduce efficiency. Consequently, load-shifting is utilized less frequently as a congestion management strategy. Lastly, the results of concept **PO.P– (c)** show a repeating pattern due to time-static premium payments. Note that a 30 €/MWh price cap was applied

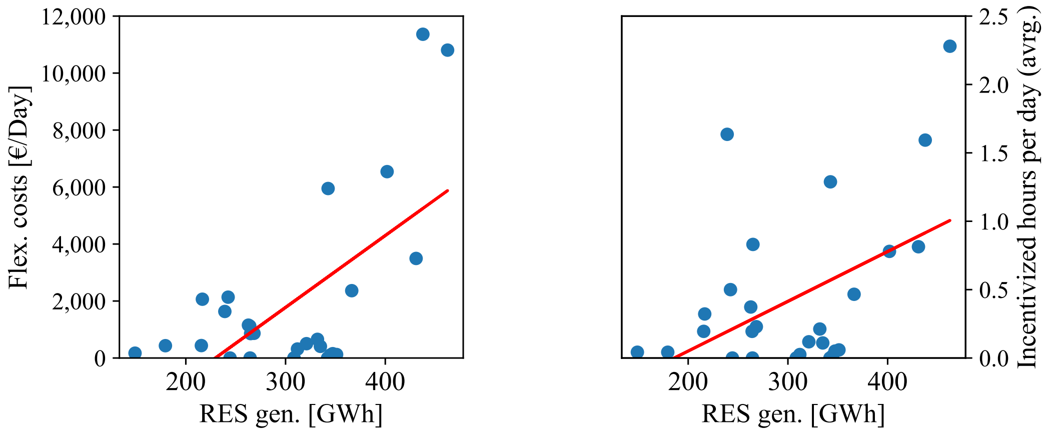


Fig. 8. Statistical relationship between load-shifting activity and renewable generation.

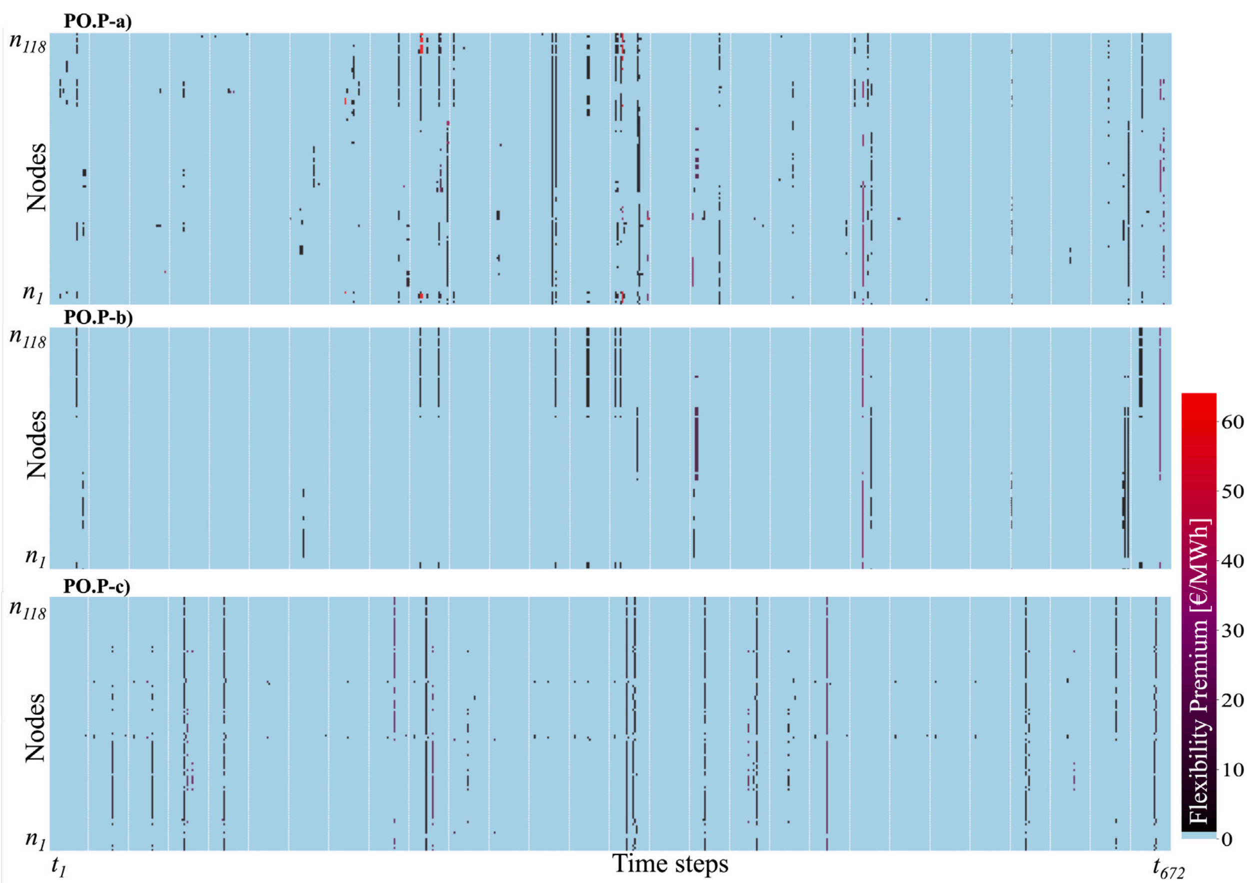


Fig. 9. Successful premium offers across time steps and nodes. Each square represents the acceptance decision of a premium offering in each time step (left to right) for each node (bottom to top). The color indicates the offered premium amount.

to avoid excessive payment volumes. When aggregators reject capped offers due to higher opportunity costs for load shifting, the structure of payments displayed in the bottom plot in Fig. 9 exhibits some interruptions.

5.4. Distributional effects

Incentivizing load applications before the market clearing stage means intervening in wholesale market equilibria. Fig. 10 shows the resulting distributional effects, i.e., consumer and producer rents, as an absolute difference compared to the rents in grid management concept **PO**. Since grid operation costs are typically reallocated to final electricity-consuming customers, consumer rents are displayed both with and without consideration of effects on resulting grid operation costs.

Since concept **PO.DC** assumes direct control of flexible load applications by the TSO, market outcomes are the same as in concept **PO**. Consequently, rents for producers and consumers are identical. However, utilizing flexible loads for congestion management reduces grid operation costs, leading to a net increase in consumer rents. According to market principles, electricity supply and demand determine wholesale prices and equilibrium quantities. The extent of incentivized load-shifting can alter wholesale price formation, resulting in a price

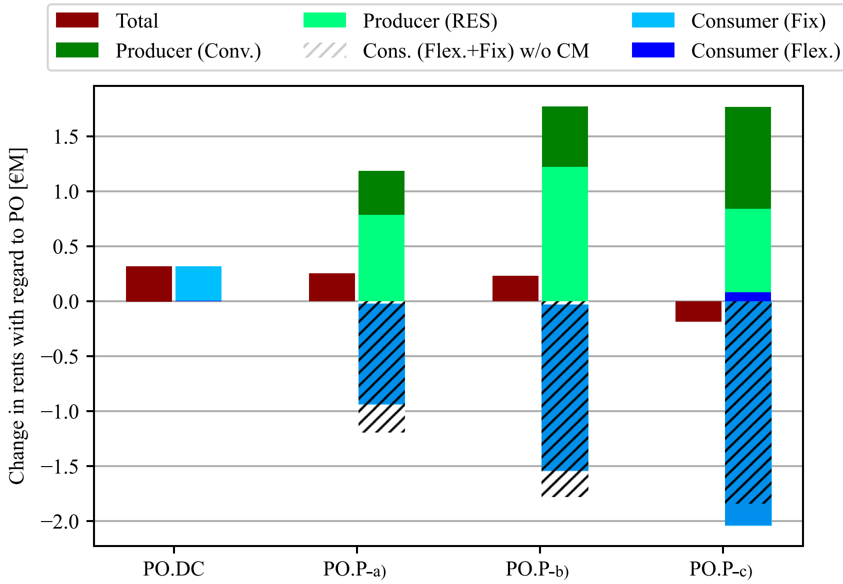


Fig. 10. Absolute difference of producer and consumer rents to concept **PO**. Hatched areas display the difference when congestion management costs are not allocated to consumers.

effect. In concept **PO.P-** **(a)**, incentivized load-shifting for grid management purposes increases prices during certain hours of the modeling horizon. Although the aggregator is compensated via *regional* premium payments, higher *zonal* wholesale prices negatively impact the rents of all other consumers (including flexible demands at nodes that are not incentivized), resulting in a net negative effect on total consumer rent. However, reductions in grid operation costs slightly counterbalance these losses.[6] Since concept **PO.P-** **(b)** incentivizes load applications across entire grid regions rather than single nodes, the price effect is more pronounced, leading to greater losses in consumer rents. The incentivization structure in concept **PO.P-** **(c)** results in very inefficient market outcomes. Consumers face higher market prices and increased grid operation costs.

The price effect observed in the concepts using incentivization as a congestion management strategy also affects producer rents. Concept **PO.P-** **(a)** shows an increase in producer rents, which is even more significant in concept **PO.P-** **(b)**. Intriguingly, conventional producers profit more in concept **PO.P-** **(c)** than renewable producers.

Calculating the difference between producer and consumer rents indicates the total welfare effect of the different congestion management concepts examined in this study. Direct control of load applications by TSOs results in the highest net positive welfare effect, though it requires stringent legal measures and additional costs for communication and information infrastructures, which are not accounted for in the calculations. Incentivizing load applications through premiums, which provide aggregators with additional revenue streams, also yields net positive welfare effects. Interestingly, these effects are similar for concepts **PO.P-** **(a)** and **PO.P-** **(b)**, where rents shift from consumers to producers. In contrast, concept **PO.P-** **(c)** results in net negative welfare effects due to the inefficient use of flexible load applications. The increase in consumer rents yields a net positive welfare effect in concept **PO.DC**.

[6] Note that this results in the difference between the blue bars, which show consumer-surplus values including all congestion management costs (the flexibility premium and redispatch costs) that are socialized through the grid-fees, and the hatched bars, which represent outcomes without such socialization, i.e., where rent-shifts are only driven by price effects during market clearing.

6. Conclusions

Integrating large amounts of renewable energy into electricity systems increasingly challenges power system operations. The gradual phase-out of large thermal-based generators to decarbonize the German energy supply significantly reduces system controllability. Incentive-based instruments constitute a valuable means of replacing traditional congestion management practices. This study investigates three distinct incentive-based congestion management strategies, which vary in premium design's temporal and regional granularity, and compares their effects on market outcomes.

More general findings from this work highlight the significant economic value of load-shifting in renewable-dominated electricity systems. Renewable curtailment is an inefficient method for resolving grid congestion from an environmental perspective and considering economic implications, as it requires redispatching generators with the greatest marginal cost differences. Therefore, designing incentive instruments for load-shifting should focus on avoiding renewable curtailment and encouraging the dispatch of load applications during these hours, especially when market signals alone are insufficient to attract such behavior.

Comparing the different grid management concepts studied in this work reveals a central relationship between market and grid efficiency. Efficiently managing grid congestion through incentivizing grid-serving dispatch of flexible load applications, such as temporal and locational varying premium payments, results in higher market costs. This result is due to the price effect associated with increased demand when incentivized flexible demand quantities reach a certain magnitude. In simple terms, while local grid congestion is mitigated through demand shifting, the resulting price increments raise electricity costs for all consumers due to zonal market clearing and associated wholesale prices accounting for the entire market zone. Alongside associated losses in consumer rents, producers will profit, particularly when the zonal market clearing is combined with a uniform pricing auction, as in Germany, from these price increments. Regulators, thus, should consider distributional effects when designing efficient demand-side incentive instruments for grid operation purposes and thoroughly outweigh who will bear the additional burden connected with such a congestion management strategy. Especially in energy systems where substantial amounts of flexibility are shifted, and distributional effects become pronounced, policymakers may need to address this issue. Importantly, our results do not reveal a simple redistribution of welfare; instead, when flexibility is incentivized efficiently, total welfare actually increases, due to reduced congestion management costs. Consequently, incentivizing flexibility can be recommended even when transmission system operation is regulated, provided that redistribution concerns are addressed and when zonal market clearing remains in place. Although this mechanism does not fully internalize scarcity, it represents a valuable short-term approach until grid reinforcement alleviates congestion.

The findings of this study further highlight the critical importance of time-specific premium design when developing incentives for grid-serving load application behavior. The inherent variability of renewable generation, such as wind and solar power, necessitates this focus on timing. While aggregating nodes into regions with a uniform premium across nodes for flexible electricity demand increases consumer rent losses, the total welfare effects are similar to those of time- and location-specific premium designs examined in this research, suggesting that regional incentive designs offer a viable alternative to nodal ones. However, reducing the temporal granularity of premium designs leads to undesirable incentives, as it overlooks short-term fluctuations in renewable generation and grid congestion. Therefore, incentive instruments should be seamlessly integrated into daily market practices, allowing for regular updates to premium offers to accommodate the variability in generation within systems with high renewable penetration.

From a practical implementation perspective, these results imply a trade-off between how precisely the incentive mechanism governs load shifting and what cost are associated with information and communications technology (ICT). While simpler premium designs may reduce computational burdens, they often fail to reflect the required granularity, whereas highly precise designs require significant investment in communication and control technologies and impose other transaction cost. Regional incentive designs seem to balance implementation effort and improved grid operation fairly well. Furthermore, policy-makers must evaluate alternatives such as direct control by the transmission system operator over flexible resources. Although direct control could theoretically alleviate grid congestion more efficiently, it entails substantial regulatory challenges, particularly regarding non-discrimination, and demands even greater ICT costs than highly precise price-based incentives. Consequently, direct control appears less practical to implement compared to flexibility premiums.

Given the time constraints imposed by market clearing processes and their interaction with congestion management practices, computational challenges may limit the herein developed framework for determining efficient premiums in real-world market practices. We leave this aspect for future research.

CRediT authorship contribution statement

Jannis Eichenberg: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Visualization. **Hannes Hobbie:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision. **Tizian Schug:** Methodology, Validation, Writing – original draft, Visualization.

Funding sources

This work was supported by the BMWK project *Effiziente Einbindung hoher Anteile Erneuerbarer Energien in technisch-wirtschaftlich integrierte Energiesysteme - EffiziEntEE* (fund number 03EI1050B). Moreover, funding by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) via the “Responsible Electronics in the Climate Change Era – REC^2” Cluster of Excellence (EXC 3035, Project-ID 533607596) is gratefully acknowledged.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge Professor Marcel Thum and Professor Kathrin Fischer for their valuable support in this work. Furthermore, the authors would like to thank Professor Dominik Möst and all members of the BMWK project *EffiziEntEE* for their insightful input and helpful discussions during the research and development phase.

Appendix A. Overview of reviewed scientific literature

Table 3 provides an overview of the classification of the reviewed literature. The table summarizes the key characteristics of the included studies along several dimensions, including their methodological approach, modeling perspective (grid- or market-oriented), and representation of flexibility sources such as distributed energy resources (DERs), battery electric vehicles (BEVs), and prosumers flexibility. In addition, it highlights the underlying optimization or equilibrium frameworks employed, ranging from linear (LP) and mixed-integer linear programming (MILP) formulations to bi-level and game-theoretic equilibrium models. This structured comparison facilitates the identification of modeling trends, methodological gaps, and differences in how flexibility and flexibility activation are represented across the literature.

Appendix B. Reformulation of bi-level problem into MILP

The reformulation breaks down into two essential steps:

1. **MPEC formulation:** First, both lower-level problems are reformulated using their corresponding Karush-Kuhn–Tucker (KKT) conditions, rendering two mixed complementary problems (MCP). By incorporating the MCPs into the upper-level constraints, a mathematical problem with equilibrium constraints (MPEC) is obtained. Due to the convexity of the two lower-level objectives and the linearity of their corresponding constraints, their constraint qualifications hold. Consequently, the KKTs are necessary and sufficient, ensuring optimality of the lower-level problems. For a detailed description, refer to [44, Chapter 2.8].
2. **MILP formulation:** Then, the complementary slackness conditions (associated with the MCPs) are linearized using a big-M approach, and the strong duality theorem is applied to express the remaining bilinear terms through a linear combination. See [44, Chapter 6.4] for a comprehensive understanding.

B.1. MPEC formulation

The lower-level problems expressed as MCPs are obtained by taking the first-order conditions of their corresponding Lagrangian functions together with their associated complementary slackness conditions, as follows:

agg :
The Lagrangian function of the aggregator’s optimization problem has the following form:

$$\begin{aligned}\mathcal{L}^{agg} =& C^{agg} + \sum_{n',t}\Bigg(-H^{agg}_{n',t}\cdot \underline{\eta}_{n',t} - \left(c^{H}_{n'} - H^{agg}_{n',t}\right)\cdot \bar{\eta}_{n',t} \\ &- \epsilon\cdot\left(\bar{\theta}_{n',t} + \underline{\theta}_{n',t}\right) + \left(dem^{H}_{n',t} - \sum_{\hat{t}\le t} H^{agg}_{n',\hat{t}}\right)\cdot\left(\underline{\theta}_{n',t} - \bar{\theta}_{n',t}\right)\Bigg)\end{aligned} \tag{B.1}$$

Table 3
Overview of reviewed studies.

Ref.	Journal	Year	Grid persp.	Market persp.	Grid Lvl.	Flex. activation	Flex. source	Model
[12]	IEEE Transactions on Smart Grid	2014	✓	✗	D	Con. price	Household	ACOPF
[13]	IEEE Transactions on Smart Grid	2019	✓	✗	D	Con. price	DERs	Seq. optimization
[14]	Int. J. of Advanced Trends in Computer Science and Engineering	2021	✓	✗	T	Incentive	Dem. elasticity	OPF
[15]	IEEE Transactions on Smart Grid	2021	✓	✗	D	Incentive	DERs	Bi-level
[16]	Power System Integration Symposium	2022	✓	✗	D	Grid fee	BEVs	SpiceEV
[17]	Energy Policy	2024	✓	✗	D	Grid fee	Prosumer	EVaTar
[18]	–	2023	✓	✗	D	Grid fee	Prosumer	Optimization
[19]	ACM SIGEnergy Energy Informatics Review	2024	✓	✗	D	Mul. dim. el. pric	Prosumer	Optimization
[21]	Int. J. of Electrical Power & Energy Systems	2022	✓	✗	D	Incentive	BEVs	Rob. optimization
[22]	(Preprint)	2022	✓	✗	D	Grid fee	Prosumer	LP
[23]	Electr. Power Syst. Res.	2024	✓	✗	D	Grid fee	Prosumer	Optimization
[24]	Sustainable Energy, Grids and Networks	2024	✓	✗	D	Grid fee	DERs	MILP
[25]	Appl. Energy	2023	✓	✗	D	Con. price	Electrolyzer	Bi-level
[26]	IEEE Access	2024	✓	✗	D	Grid fee	BEVs	Gaussian mixture
[27]	Energy Policy	2025	✓	✗	D + T	grid fee	Agg. load	Optimization
[28]	IEEE Transactions on Power Syst.	2019	✓	✗	D	Block offer	DERs	OPF
[29]	IEEE Transactions on Sustainable Energy	2020	(✓)	✗	(T)	Incentive	Abstract	ML
[32]	International Conference on the European Energy Mark	2018	✗	✓	D	Grid fee	Prosumer	Bi-Level
[33]	Energy Policy	2026	✗	✓	D	Grid fee	Industry	MILP
[34]	The Energy J.	2020	✗	✓	D	Grid fee	Prosumer	Bi-Level
[35]	Energy Policy	2021	✗	✓	D	Grid fee	Prosumer	LP
[36]	Energy Economics	2019	✗	✓	D	Grid fee	Prosumer	Equilibrium mod.
[37]	Energy Economics	2018	✗	✓	D	Grid fee	Prosumer	Equilibrium mod.
This paper	Energy Strategy Rev.	2026	✓	✓	T	Incentive	Heat pumps	Bi-Level

(a) with first-order condition, $\forall n', t$

$$\frac{\partial \mathcal{L}^{agg}}{\partial H^{agg}_{n',t}} = \Pi^{mc}_t - FP^{tso}_{reg,t} - \underline{\eta}_{n',t} + \bar{\eta}_{n',t} + \sum_{\hat{t}\geq t}\left(\bar{\theta}_{n',\hat{t}} - \underline{\theta}_{n',\hat{t}}\right) = 0 \quad \text{(B.2)}$$

(b) and complementary slackness conditions

$$0 \leq H^{agg}_{n',t} \qquad \perp \underline{\eta}_{n',t} \geq 0 \qquad \forall n', t \quad \text{(B.3)}$$

$$0 \leq c^H_{n'} - H^{agg}_{n',t} \qquad \perp \bar{\eta}_{n',t} \geq 0 \forall n', t \quad \text{(B.4)}$$

$$0 \leq -dem^H_{n',t} + \epsilon + \sum_{\hat{t}\leq t} H^{agg}_{n',\hat{t}} \qquad \perp \underline{\theta}_{n',t} \geq 0 \forall n', t \quad \text{(B.5)}$$

$$0 \leq dem^H_{n',t} + \epsilon - \sum_{\hat{t}\leq t} H^{agg}_{n',\hat{t}} \qquad \perp \bar{\theta}_{n',t} \geq 0 \forall n', t \quad \text{(B.6)}$$

mca :
The Lagrangian function of the market clearing agent's optimization problem is as follows:

$$\mathcal{L}^{mca} = C^d + \sum_{p,t}\left(-G^{mca}_{p,t}\cdot \underline{o}_{p,t} - \left(c^G_{p,t} - G^{mca}_{p,t}\right)\cdot \bar{o}_{p,t}\right) + \sum_t \left(\sum_n d_{n,t} + \sum_{n'} H^{agg}_{n',t} - \sum_p G^{mca}_{p,t}\right)\cdot \Pi^{mc}_t \quad \text{(B.7)}$$

(a) with first-order condition, $\forall p, t$

$$\frac{\partial \mathcal{L}^{mca}}{\partial G^{mca}_{p,t}} = oc_p - \Pi^{mc}_t - \underline{o}_{p,t} + \bar{o}_{p,t} = 0 \quad \text{(B.8)}$$

(b) and complementary slackness conditions

$$0 \leq G^{mca}_{p,t} \qquad \perp \underline{o}_{p,t} \geq 0 \forall p, t \quad \text{(B.9)}$$

$$0 \leq c^G_{p,t} - G^{mca}_{p,t} \qquad \perp \bar{o}_{p,t} \geq 0 \forall p, t \quad \text{(B.10)}$$

$$0 = \sum_n d_{n,t} + \sum_{n'} H^{agg}_{n',t} - \sum_p G^{mca}_{p,t} \qquad \perp \Pi^{mc}_t \geq 0 \forall t \quad \text{(B.11)}$$

Combining the upper-level constraints (Section 3.4.1) with the lower-level problems, after substituting them with their corresponding MCPs (B.2)–(B.6) and (B.8)–(B.11), results in a single-level MPEC.

B.2. MILP formulation

The MPEC can be reformulated as a MILP by linearizing all nonlinearities to provide improved computational efficiency. These include all complementary slackness conditions and the product of the flexibility premium with the day-ahead dispatch of heat pumps in the objective function of the aggregator (see (ii.2), resp. (ii.5)).

The complementarity conditions ((B.3)–(B.6) and (B.9)–(B.11)) are reformulated using a big-M[7] approach, as illustrated below:

agg :

$$H^{agg}_{n',t} \leq m \cdot B^{\underline{\eta}}_{n',t} \qquad \forall n', t \quad \text{(B.3a)}$$

$$\underline{\eta}_{n',t} \leq m \cdot \left(1 - B^{\underline{\eta}}_{n',t}\right) \qquad \forall n', t \quad \text{(B.3b)}$$

$$c^H_{n'} - H^{agg}_{n',t} \leq m \cdot B^{\bar{\eta}}_{n',t} \qquad \forall n', t \quad \text{(B.4a)}$$

$$\bar{\eta}_{n',t} \leq m \cdot \left(1 - B^{\bar{\eta}}_{n',t}\right) \qquad \forall n', t, \quad \text{(B.4b)}$$

$$-dem^H_{n',t} + \epsilon + \sum_{\hat{t}\leq t} H^{agg}_{n',\hat{t}} \leq m \cdot B^{\underline{\theta}}_{n',t} \qquad \forall n', t \quad \text{(B.5a)}$$

$$\underline{\theta}_{n',t} \leq m \cdot \left(1 - B^{\underline{\theta}}_{n',t}\right) \qquad \forall n', t, \quad \text{(B.5b)}$$

$$dem^H_{n',t} + \epsilon - \sum_{\hat{t}\leq t} H^{agg}_{n',\hat{t}} \leq m \cdot B^{\bar{\theta}}_{n',t} \qquad \forall n', t \quad \text{(B.6a)}$$

$$\bar{\theta}_{n',t} \leq m \cdot \left(1 - B^{\bar{\theta}}_{n',t}\right) \qquad \forall n', t, \quad \text{(B.6b)}$$

mca :

$$G^{mca}_{p,t} \leq m \cdot B^{\underline{o}}_{p,t} \qquad \forall p, t \quad \text{(B.9a)}$$

$$\underline{o}_{p,t} \leq m \cdot \left(1 - B^{\underline{o}}_{p,t}\right) \qquad \forall p, t, \quad \text{(B.9b)}$$

$$c^G_{p,t} - G^{mca}_{p,t} \leq m \cdot B^{\bar{o}}_{p,t} \qquad \forall p, t \quad \text{(B.10a)}$$

$$\bar{o}_{p,t} \leq m \cdot \left(1 - B^{\bar{o}}_{p,t}\right) \qquad \forall p, t, \quad \text{(B.10b)}$$

[7] All big-M values are set to maximum generation capacity plus a small offset.

$$\sum_n d_{n,t} + \sum_{n'} H^{agg}_{n',t} - \sum_p G^{mca}_{p,t} \le m \cdot B^{\Pi^{mc}}_t \qquad \forall t \tag{B.11a}$$

$$\Pi^{mc}_t \le m \cdot \left(1 - B^{\Pi^{mc}}_t\right) \qquad \forall t, \tag{B.11b}$$

where $B^{\cdot}_{n',t} \in \{0,1\}$ represents a newly introduced auxiliary variable and $m \in \mathbb{R}^+$ is a sufficiently large constant.

Recalling the bilinear term (ii.2) included in Eq. (ii.5), Lemma 1 is used to express it exactly as a linear combination.

Lemma 1. *Given an optimal solution of the lower-level problems of the bi-level problem formulated in Section 3.4, the following equality holds:*

$$\sum_{n',t} H^{agg,*}_{n',t} \cdot FP^{tso}_{reg(n'),t} = \sum_{p,t} \left(G^{mca,*}_{p,t} \cdot oc_p + c^G_p \cdot \bar{o}^*_{p,t}\right) - \sum_{n,t} d_{n,t} \cdot \Pi^{mc,*}_t$$

$$- \sum_{n',t} \left(-c^H_{n'} \cdot \bar{\eta}^*_{n',t} + \left(dem^H_{n',t} - \epsilon\right) \cdot \underline{\theta}^*_{n',t} + \left(-dem^H_{n',t} - \epsilon\right) \cdot \bar{\theta}^*_{n',t}\right) \tag{ii.2a}$$

where "" denotes optimality for the decision problems.*

Proof. The left-hand side of Eq. (ii.2a) represents the bilinear term of interest and is also present in the objective function of the aggregator's decision problem in the lower-level (ii.12). As previously mentioned, the constraint qualifications hold for both lower-level problems. Consequently, the strong duality theorem applies, allowing the primal objectives to be equated with their respective Lagrangians at the optimum.[8] Appendix D demonstrates the derivation of the optimal values for the dual problems (right-hand side of (B.12) and (B.15)).

For the aggregator, the following holds:

$$\sum_{n',t} H^{agg,*}_{n',t} \cdot \left(\Pi^{mc}_t - FP^{tso}_{reg(n'),t}\right) =$$
$$\sum_{n',t} \left(-c^H_{n'} \cdot \bar{\eta}^*_{n',t} + \left(dem^H_{n',t} - \epsilon\right) \cdot \underline{\theta}^*_{n',t} + \left(-dem^H_{n',t} - \epsilon\right) \cdot \bar{\theta}^*_{n',t}\right), \tag{B.12}$$

which can be reformulated to

$$\sum_{n',t} H^{agg,*}_{n',t} \cdot FP^{tso}_{reg(n'),t} = \sum_{n',t} H^{agg,*}_{n',t} \cdot \Pi^{mc}_t$$
$$- \sum_{n',t} \left(-c^H_{n'} \cdot \bar{\eta}^*_{n',t} + \left(dem^H_{n',t} - \epsilon\right) \cdot \underline{\theta}^*_{n',t} + \left(-dem^H_{n',t} - \epsilon\right) \cdot \bar{\theta}^*_{n',t}\right). \tag{B.13}$$

The right-hand side of Eq. (B.13) now exhibits a second bilinear term:

$$\sum_{n',t} H^{agg,*}_{n',t} \cdot \Pi^{mc}_t. \tag{B.14}$$

The theorem of strong duality is applied again to linearize this non-linearity. Since (B.14) is part of the Lagrangian (B.7) representing the market clearing, the market clearing agent's decision problem is utilized:

$$\sum_{p,t} G^{mca,*}_{p,t} \cdot oc_p = \sum_{p,t} -c^G_{p,t} \cdot \bar{o}^*_{p,t} + \sum_{n,t} \left(d_{n,t} + H^{agg}_{n'(n),t}\right) \cdot \Pi^{mc,*}_t, \tag{B.15}$$

which can be reformulated to

$$\sum_{n,t} H^{agg}_{n'(n),t} \cdot \Pi^{mc,*}_t = \sum_{p,t} \left(G^{mca,*}_{p,t} \cdot oc_p + c^G_{p,t} \cdot \bar{o}^*_{p,t}\right) - \sum_{n,t} d_{n,t} \cdot \Pi^{mc,*}_t. \tag{B.16}$$

Finally, Lemma 1 can be obtained by substituting Eq. (B.16) into Eq. (B.13). □

[8] $H^{agg,*}_{n',t}, \bar{\eta}^*_{n',t}, \underline{\theta}^*_{n',t}, \bar{\theta}^*_{n',t}$ denote optimality for the aggregator's decision problem. $FP^{tso}_{n',t}$ and Π^{mc}_t are controlled by the TSO and the day-ahead market clearing agent, respectively, and can be treated as parameters in the aggregator's problem. Similarly, $\Pi^{mc,*}_t, G^{mca,*}_{p,t}$, and $\bar{o}^*_{p,t}$ denote optimality for the market clearing agent's decision problem, allowing $H^{agg}_{n',t}$ to be treated as a parametric input.

Since the application of Lemma 1 results in a mathematically exact linear reformulation of the bilinear term in the TSO's objective function (ii.5), the MPEC can be expressed as a MILP, enhancing solvability.

Appendix C. Regularization techniques to mitigate optimistic bias artifacts for the ill-posed Model II

Well-posed optimization problems differ from ill-posed problems i. a. by the existence of a unique solution. However, the lower-level problem of the applied bi-level framework - in particular the aggregator - is generally not uniquely solvable in terms of what we call *temporal indifference*. Temporal indifference may occur in two situations:

1. If several time steps are close together and equally expensive (in terms of market clearing prices), or
2. each time an incentive is set in the model.

In the first situation, the solution set for each node is not necessarily unique, since the temporal choice of the dispatch decision does not affect the aggregator's purchasing costs (as long as feasibility is provided). The second situation implies the first one not in terms of market clearing prices, but purchasing costs: The aggregators' opportunity costs for grid-serving load shifts are determined by price gaps between the time steps during which load-shifting occurs. As the smallest possible successful incentive, a premium must reduce this price gap to zero to pay out exactly those opportunity costs. However, the aggregators' objective remains unchanged, regardless of whether it responds to the incentive or not.

In conclusion, the bi-level problem is ill-posed. Thus, the upper-level only draws those solutions from the (not unique) lower-level solution set that enable cost-optimized congestion management. This is known as the *optimistic approach* and may lead to bias artifacts, disturbing the comparability of incentive effects regarding different incentive designs. In the following, it is described which premium designs may are affected by the optimistic approach and which regularization techniques are developed to mitigate corresponding bias artifacts:

Since **PO.P– (a)** is characterized by a premium where temporal and spatial resolution correspond to the model resolution, a non-optimistic formulation of the model could take control of temporal indifference, by setting an (additional) $\hat{\epsilon}$-small incentive to the same nodes and time steps, the optimistic approach would draw. If $\hat{\epsilon} \to 0$, the upper-level objective function is not affected and both approaches, optimistic and non-optimistic yield equivalent solutions. This implies that there is no requirement for the application of regularization techniques.

Suppose a premium is set for a whole region, because of situation 2, the upper-level in **PO.P– (b)** may draw solutions where only nodes useful for congestion management are reacting to the incentive, while others do not take part in the congestion management measure. These nodes are subject to the incentive, but neither generate flexibility costs nor do they exert an overriding influence on load flows. In conclusion, the optimistic character transforms the regional design into a nodal one. This necessitates an actual implementation of a regional additional $\hat{\epsilon}$-small incentive:

1. Since variable costs of all power plants and thus all market clearing prices in the case study are integers, defining Π^{mc} as an integer variable does not impact model results:

$$\Pi^{mc}_t \in \mathbb{Z} \quad (\forall t \in T). \tag{C.17}$$

2. Because of 1., all price gaps and thus all incentives in our optimistic approach will be integers. Thus, defining premiums as integer variable does not impact model results, too:

$$FP^{tso}_{reg(n'),t} \in \mathbb{Z} \quad \left(\forall n' \in N', t \in T\right). \tag{C.18}$$

3. Expanding the objective function of the aggregator about $\hat{\epsilon} = 1.01$ results in the modified objective,

$$min \quad C^{agg} = \sum_{n',t} H^{agg}_{n',t} \cdot \left(\Pi^{mc}_t - FP^{tso}_{reg(n'),t} \cdot \hat{\epsilon} \right). \tag{ii.12a}$$

Thus the first-order condition modifies to

$$\Pi^{mc}_t - FP^{tso}_{reg(n'),t} \cdot \hat{\epsilon} - \underline{\eta}_{n',t} + \bar{\eta}_{n',t} + \sum_{\hat{t} \geq t} \left(\bar{\theta}_{n',\hat{t}} - \underline{\theta}_{n',\hat{t}} \right) = 0, \quad \forall n', t. \tag{B.2a}$$

Since 1., 2. and $\hat{\epsilon}$ imply that an incentive overwhelms opportunity costs, we denote $\hat{\epsilon}$ as *incentive enhancement factor*. Consequently, the aforementioned concerns are addressed by enforcing a unique incentive-response mechanism, whereby each node within an incentivized region will react to.

4. Dividing the right-hand side of Eq. (ii.2a) by $\hat{\epsilon}$ adjusts the flexibility costs to the original model version while maintaining the unique characteristic of the regional incentive.

Since **PO.P– (c)** disables $\hat{\epsilon}$-small premiums during time steps $t^c_n \in T^{inc\,C}_n$ not element of an incentive window T^{inc}_n, the argument formulated in **PO.P– (a)** does not hold again. To mitigate the resulting bias, normal distributed noise is added to the aggregator's objective, so the first-order condition modifies to

$$\Pi^{mc}_t + u_{n',t} - FP^{tso}_{reg(n'),t} - \underline{\eta}_{n',t} + \bar{\eta}_{n',t} + \sum_{\hat{t} \geq t} \left(\bar{\theta}_{n',\hat{t}} - \underline{\theta}_{n',\hat{t}} \right) = 0 \quad \forall n', t, \tag{B.2b}$$

where $u_{n',t} \sim \mathcal{U}(-0.1, 0.1)$. Consequently, in all time steps t^c_n, a unique random solution is enforced, eliminating any optimistic bias. Similar regularization techniques are discussed in [39].

Appendix D. Estimating optimal values of dual problems

For optimal values λ^*, μ^* of the corresponding dual problem of the optimization problem

$$min_x f(x) \qquad s.t. \quad g(x) \leq 0, \quad h(x) = 0, \tag{D.1}$$

estimating the optimal value of the dual, means to minimize its Lagrangian with respect to x:

$$\mathcal{L}^*_{x^*,\lambda^*,\mu} = min_x \mathcal{L}^*_{x,\lambda^*,\mu}$$
$$\iff \mathcal{L}^*_{x,\lambda^*,\mu} \quad s.t. \quad \frac{\partial \mathcal{L}^*_{x,\lambda^*,\mu}}{\partial x} = 0. \tag{D.2}$$

In the following, we want to apply the relationships mentioned to the lower-level problems.

Aggregator

Regarding the Lagrangian (B.1) and its derivative (B.2) of the aggregator, $\frac{\partial \mathcal{L}^{agg}}{\partial H^{agg}_{n',t}} = 0$ is equivalent to

$$\Pi^{mc}_t - FP^{tso}_{reg(n'),t} - \underline{\eta}_{n',t} + \bar{\eta}_{n',t} + \sum_{\hat{t} \geq t} \left(\bar{\theta}_{n',\hat{t}} - \underline{\theta}_{n',\hat{t}} \right) = 0 \quad \forall n', t. \tag{D.3}$$

With the equivalency in (D.2) we receive the optimal value of the dual problem denoted by $\mathcal{L}^{agg,*}$ as

$$\mathcal{L}^{agg,*} = \sum_{n',t} \left(-c^H_{n'} \cdot \bar{\eta}^*_{n',t} + \left(dem^H_{n',t} - \epsilon \right) \cdot \underline{\theta}^*_{n',t} + \left(-dem^H_{n',t} - \epsilon \right) \cdot \bar{\theta}^*_{n',t} \right). \tag{D.4}$$

Day-ahead market clearing operator

Regarding the Lagrangian (B.7) and its derivative (B.8) of the day-ahead market clearing operator, $\frac{\partial \mathcal{L}^{mca}}{\partial G^{mca}_{p,t}} = 0$ is equivalent to

$$oc_p - \underline{o}_{p,t} + \bar{o}_{p,t} - \Pi^{mc}_t = 0 \quad \forall p, t. \tag{D.5}$$

With the equivalency in (D.2) we receive the optimal value of the dual problem denoted by $\mathcal{L}^{mca,*}$ as

$$\mathcal{L}^{mca,*} = \sum_{p,t} -c^G_p \cdot \bar{o}^*_{p,t} + \sum_{n,t} \left(d_{n,t} + H^{agg}_{n'(n),t} \right) \cdot \Pi^{mc,*}_t \tag{D.6}$$

Appendix E. Figures

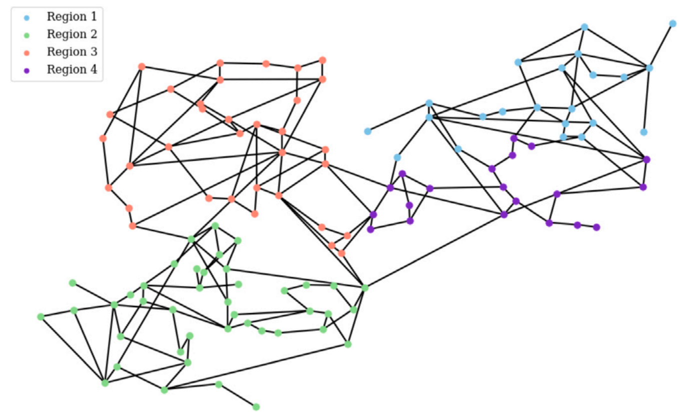


Fig. 11. Grid regions.

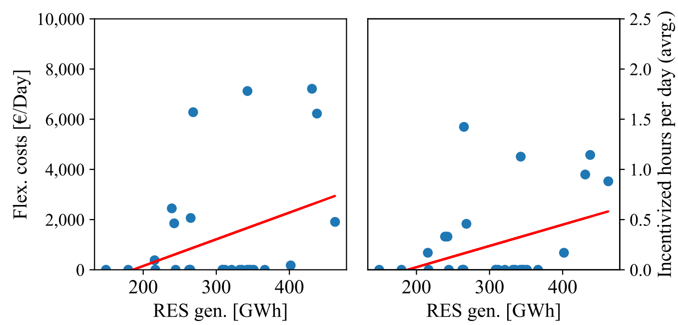


Fig. 12. correlation RES vs. **PO.P– (b)**.

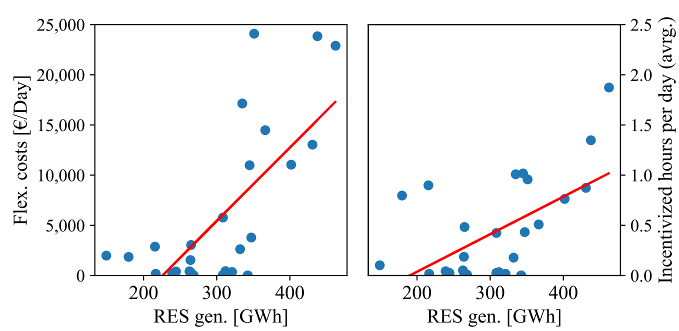


Fig. 13. correlation RES vs. **PO.P– (c)**.

See Figs. 11, 12 and 13.

Data availability

Data will be made available on request.

References

[1] Bundesnetzagentur (Federal Network Agency), Smard: Electricity and gas market data for Germany, 2024, https://www.smard.de/home. (Accessed 13 October 2025).

[2] C. Loschan, D. Schwabeneder, G. Lettner, H. Auer, Flexibility potential of aggregated electric vehicle fleets to reduce transmission congestions and redispatch needs: A case study in Austria, Int. J. Electr. Power Energy Syst. 146 (2023) 108802, http://dx.doi.org/10.1016/j.ijepes.2022.108802.

[3] D. Bauknecht, F. Flachsbarth, M. Koch, M. Vogel, The role of decentralised flexibility options for managing transmission grid congestions in Germany, Electr. J. 37 (1) (2024) 107363, http://dx.doi.org/10.1016/j.tej.2023.107363, URL https://www.sciencedirect.com/science/article/pii/S1040619023001306.

[4] H. Tang, S. Wang, Game-theoretic optimization of demand-side flexibility engagement considering the perspectives of different stakeholders and multiple flexibility services, Appl. Energy 332 (2023) 120550, http://dx.doi.org/10.1016/j.apenergy.2022.120550.

[5] A. Abdollahi, S.C. Khavar, Enhancing the flexibility of decentralized energy resources through bi-level optimization in intra-day regional markets, Renew. Energy Focus. 56 (2026) 100778, http://dx.doi.org/10.1016/j.ref.2025.100778.

[6] R. Khodabakhsh, M. Haghifam, M.S.E. Eslami, Designing a bi-level flexibility market for transmission system congestion management considering distribution system performance improvement, Sustain. Energy, Grids Netw. 34 (2023) 101000, http://dx.doi.org/10.1016/j.segan.2023.101000.

[7] X. Jin, Q. Wu, H. Jia, Local flexibility markets: Literature review on concepts, models and clearing methods, 2020, http://dx.doi.org/10.1016/j.apenergy.2019.114387.

[8] O. Rebenaque, C. Schmitt, K. Schumann, T. Dronne, F. Roques, Success of local flexibility market implementation: A review of current projects, Util. Policy 80 (2023) http://dx.doi.org/10.1016/j.jup.2023.101491.

[9] S. Häseler, A.J. Wulf, Promoting real-time electricity tariffs for more demand response from german households: A review of four policy options, Energy, Sustain. Soc. 14 (2024) 59, http://dx.doi.org/10.1186/s13705-024-00490-z.

[10] Monopolkommission (German Monopolies Commission), 10. Sector report on energy: Competition and efficiency for a sustainable energy system, Bonn (2025) URL https://monopolkommission.de/de/gutachten/sektorgutachten/sektorgutachten-energie/477-10-sektorgutachten-energie-2025.html.

[11] TransnetBW, Amprion, TenneT TSO, 50Hertz Transmission, Dynamic grid fees: Discussion paper by the four transmission system operators, Discuss. Pap. (2025) URL https://www.transnetbw.de/_Resources/Persistent/6/4/7/3/64735d3bc909da39efaae4d38008dd65bbd7e6d2/4%C3%9CNB-Diskussionspapier%20Dynamische%20Netzentgelte_202511.pdf.

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

[13] F. Shen, S. Huang, Q. Wu, S. Repo, Y. Xu, J. Ostergaard, Comprehensive congestion management for distribution networks based on dynamic tariff, reconfiguration, and re-profiling product, IEEE Trans. Smart Grid 10 (2019) 4795–4805, http://dx.doi.org/10.1109/TSG.2018.2868755.

[14] A. Erita, D. Citra, D.T.P. Yanto, Modelling demand response to alleviate congestion in active power market, Int. J. Adv. Trends Comput. Sci. Eng. 10 (2021) , 1277–1283, http://dx.doi.org/10.30534/ijatcse/2021/1111022021.

[15] S. Fattaheian-Dehkordi, M. Tavakkoli, A. Abbaspour, M. Fotuhi-Firuzabad, M. Lehtonen, An incentive-based mechanism to alleviate active power congestion in a multi-agent distribution system, IEEE Trans. Smart Grid 12 (2021) 1978–1988, http://dx.doi.org/10.1109/TSG.2020.3037560.

[16] J.T. Meyer, C. Daam, J. Gemassmer, Realise flexibility potential of ev fleets through grid-serving charging strategies, in: 6th E-Mobil. Power Syst. Integr. Symposium, EMOB 2022, vol. 6 (2022) http://dx.doi.org/10.1049/icp.2022.2716.

[17] J. Stute, M. Klobasa, How do dynamic electricity tariffs and different grid charge designs interact? - Implications for residential consumers and grid reinforcement requirements, Energy Policy 189 (2024) 114062, http://dx.doi.org/10.1016/j.enpol.2024.114062.

[18] P. Godron, M. Herrndorff, S. Müller, K. Schaber, M. Zackariat, N. Jooss, S. Köppl, M. Müller, J. Reinhard, S. von Roon, A. Weiss, Haushaltsnahe flexibilitäten nutzen. Wie elektrofahrzeuge, wärmepumpen und co. die stromkosten für alle senken können (using household-related flexibility. How electric vehicles, heat pumps and the like can reduce electricity costs for everyone), 2023, URL https://www.agora-energiewende.de/publikationen/haushaltsnahe-flexibilitaeten-nutzen.

[19] T. Riedel, C. Hauschke, H. Schmeck, Power-dependent price profiles - defining grid- and market-oriented incentives for building energy management systems, ACM SIGEnergy Energy Inform. Rev. 4 (2024) 78–87, http://dx.doi.org/10.1145/3717413.3717420.

[20] J. Stute, M. Kühnbach, Dynamic pricing and the flexible consumer – investigating grid and financial implications: A case study for Germany, Energy Strat. Rev. 45 (2023) 100987, http://dx.doi.org/10.1016/j.esr.2022.100987.

[21] F. Shen, Q. Wu, Robust dynamic tariff method for day-ahead congestion management of distribution networks, Int. J. Electr. Power & Energy Syst. 134 (2022) 107366, http://dx.doi.org/10.1016/j.ijepes.2021.107366.

[22] C. Winzer, P. Ludwig, Optimal design of dynamic grid tariffs (preprint), 2022, http://dx.doi.org/10.2139/ssrn.4130379, SSRN.

[23] O. Banovic, K. Schumann, J. Zocher, A. Ulbig, On grid-serving grid tariff design in local energy markets, Electr. Power Syst. Res. 234 (2024) 110655, http://dx.doi.org/10.1016/j.epsr.2024.110655.

[24] B. Velkovski, V.Z. Gjorgievski, D. Kothona, A.S. Bouhouras, S. Cundeva, N. Markovska, Impact of tariff structures on energy community and grid operational parameters, Sustain. Energy, Grids Netw. 38 (2024) 101382, http://dx.doi.org/10.1016/j.segan.2024.101382.

[25] S. Ghaemi, X. Li, M. Mulder, Economic feasibility of green hydrogen in providing flexibility to medium-voltage distribution grids in the presence of local-heat systems, Appl. Energy 331 (2023) 120408, http://dx.doi.org/10.1016/j.apenergy.2022.120408.

[26] L.J. Lepolesa, K.E. Adetunji, K. Ouahada, Z. Liu, L. Cheng, Optimal EV charging strategy for distribution networks load balancing in a smart grid using dynamic charging price, IEEE Access 12 (2024) 47421–47432, http://dx.doi.org/10.1109/ACCESS.2024.3382124.

[27] M. Hofmann, S. Bjarghov, H. Sæle, K.B. Lindberg, Grid tariff design and peak demand shaving: A comparative tariff analysis with simulated demand response, Energy Policy 198 (2025) 114475, http://dx.doi.org/10.1016/j.enpol.2024.114475.

[28] A. Hermann, J. Kazempour, S. Huang, J. Ostergaard, Congestion management in distribution networks with asymmetric block offers, IEEE Trans. Power Syst. 34 (2019) 4382–4392, http://dx.doi.org/10.1109/TPWRS.2019.2912386.

[29] J. Gao, Z. Ma, Y. Yang, F. Gao, G. Guo, Y. Lang, The impact of customers' demand response behaviors on power system with renewable energy sources, IEEE Trans. Sustain. Energy 11 (2020) 2581–2592, http://dx.doi.org/10.1109/TSTE.2020.2966906.

[30] TransnetBW, Octopus Energy, Mobility4Grid: OctoFlexBW, 2026, https://www.transnetbw.de/de/unternehmen/portraet/innovationen/mobility4grid/octoflexbw-1. (Accessed 27 February 2026).

[31] J. Bronisch, A. Herrmann, Grids and benefits: Final review, 2026, https://ut-um.files.svdcdn.com/production/media/downloads/Grids-and-Benefits_Final-Review_public.pdf?dm=1770024456. (Accessed 27 February 2026).

[32] N. Govaerts, K. Bruninx, E. Delarue, Impact of distribution tariff design on the profitability of aggregators of distributed energy storage systems, in: 2018 15th International Conference on the European Energy Market, EEM, IEEE, Lodz, 2018, pp. 1–5, http://dx.doi.org/10.1109/EEM.2018.8469793.

[33] J. Beranek, N. Niesler, P. Jochem, A. Ardone, W. Fichtner, Flexibility at a cost: Industrial battery storage and the breakdown of grid fee fairness in Germany, Energy Policy 212 (2026) 115156, http://dx.doi.org/10.1016/j.enpol.2026.115156.

[34] T. Schittekatte, L. Meeus, Least-cost distribution network tariff design in theory and practice, Energy J. 41 (2020) 119–156, http://dx.doi.org/10.5547/01956574.41.5.tsch.

[35] M. Avau, N. Govaerts, E. Delarue, Impact of distribution tariffs on prosumer demand response, Energy Policy 151 (2021) http://dx.doi.org/10.1016/j.enpol.2020.112116.

[36] N. Govaerts, K. Bruninx, H.L. Cadre, L. Meeus, E. Delarue, Spillover effects of distribution grid tariffs in the internal electricity market: An argument for harmonization? Energy Econ. 84 (2019) 104459, http://dx.doi.org/10.1016/j.eneco.2019.07.019.

[37] T. Schittekatte, I. Momber, L. Meeus, Future-proof tariff design: Recovering sunk grid costs in a world where consumers are pushing back, Energy Econ. 70 (2018) 484–498, http://dx.doi.org/10.1016/j.eneco.2018.01.028.

[38] T. Hartmann, T. Mühlhaus, H. Neumann, H. Reuter, H. Röschmann, Lastprofile für unterbrechbare verbrauchseinheiten (load profiles for interruptible consumption units), 2003, URL https://www.bdew.de/media/documents/LPuVe-Praxisleitfaden.pdf.

[39] A.B. Zemkoho, Solving ill-posed bilevel programs, Set-Valued Var. Anal. 24 (2016) 423–448, http://dx.doi.org/10.1007/s11228-016-0371-x.

[40] H. Barrios, A. Roehder, H. Natemeyer, A. Schnettler, A benchmark case for network expansion methods, in: 2015 IEEE Eindhoven PowerTech, IEEE, Eindhoven, 2015, http://dx.doi.org/10.1109/PTC.2015.7232601.

[41] Entsog, Entso-e, TYNDP 2022 scenario report, 2022, https://2022.entsos-tyndp-scenarios.eu/. (Accessed 27 February 2026).

[42] F. Sensfuß, C. Maurer, S. Blömer, S.B. De, M. Pehnt, M.P. De, P. Mellwig, P.M. De, J. Müller-Kirchenbauer, J.B.D. Auftraggeber, Langfristszenarien für die transformation des energiesystems in Deutschland 3 (long-term scenarios for the transformation of the energy system in Germany 3), 2021, URL https://www.langfristszenarien.de/enertile-explorer-wAssets/docs/LFS-Gebaeude.pdf.

[43] H. Hobbie, J. Mehlem, C. Wolff, L. Weber, F. Flachsbarth, D. Möst, A. Moser, Impact of model parametrization and formulation on the explorative power of electricity network congestion management models - insights from a grid model comparison experiment, Renew. Sustain. Energy Rev. 159 (2022) 112163, http://dx.doi.org/10.1016/j.rser.2022.112163.

[44] S.A. Gabriel, A.J. Conejo, J.D. Fuller, B.F. Hobbs, C. Ruiz, Complementarity Modeling in Energy Markets, vol. 180, Springer New York, New York, NY, 2013, http://dx.doi.org/10.1007/978-1-4419-6123-5.