The Cost of Decoupling Trade and Transport in the European Entry-Exit Gas Market with Linear Physics Modeling

Tom Böttger, Veronika Grimm,
Thomas Kleinert, Martin Schmidt*

Abstract. Liberalized gas markets in Europe are organized as entry-exit regimes so that gas trade and transport are decoupled. The decoupling is achieved via the announcement of technical capacities by the transmission system operator (TSO) at all entry and exit points of the network. These capacities can be booked by gas suppliers and customers in long-term contracts. Only traders who have booked capacities up-front can "nominate" quantities for injection or withdrawal of gas via a day-ahead market. To ensure feasibility of the nominations for the physical network, the TSO must only announce technical capacities for which all possibly nominated quantities are transportable. In this paper, we use a four-level model of the entry-exit gas market to analyze possible welfare losses associated with the decoupling of gas trade and transport. In addition to the multilevel structure, the model contains robust aspects to cover the conservative nature of the European entry-exit system. We provide several reformulations to obtain a single-level mixed-integer quadratic problem. The overall model of the considered market regime is extremely challenging and we thus have to make the main assumption that gas flows are modeled as potential-based linear flows. Using the derived single-level reformulation of the problem, we show that the feasibility requirements for technical capacities imply significant welfare losses due to unused network capacity. Furthermore, we find that the specific structure of the network has a considerable influence on the optimal choice of technical capacities. Our results thus show that trade and transport are not decoupled in the long term. As a further source of welfare losses and discrimination against individual actors, we identify the minimum prices for booking capacity at the individual nodes.

Key words and phrases. OR in Energy, Entry-Exit Gas Market, Gas Market Design, Multilevel Optimization, Robust Optimization

1. Introduction

Starting in the 1990s, the European gas market has been liberalized step by step over the last decades. The First and the Second Gas Directive [15, 17] paved the way for the Third Energy Package [16], which was introduced in 2009. This package essentially prescribes the decoupling of gas trade and transport via an appropriate entry-exit market design in all member states. Today, European transmission system operators (TSOs) usually operate under variants of such an entry-exit regime, in which traders sign long-term capacity contracts—so-called bookings—at entry and exit points of the network. Only traders who have booked capacities can afterward "nominate" quantities to feed-in (in the case of suppliers) or withdraw (in the case of consumers) on a daily basis. These nominations are determined by the trade of gas (up to the individually booked capacity) on a day-ahead market. To ensure that all nominations are indeed feasible w.r.t. the given network, the TSO announces so-called technical capacities up-front that limit the long-term capacity contracts (i.e., the bookings) per entry or exit point. While the entry-exit market design effectively decouples trade and transport of gas and thus might achieve to make market interaction more transparent, it clearly comes at a cost. In particular, the

Date: April 26, 2021.
2010 *Mathematics Subject Classification.* 90-XX, 90C35, 91B15, 91B16, 91B24.
*Corresponding author: Martin Schmidt (martin.schmidt@uni-trier.de)

technical capacity requirement that any possible market outcome must be feasible w.r.t. the network implies that the usable network capacity significantly lags behind the actual capacity of the network. Consequently, as compared to a nodal-pricing design, it is likely that the network capacity is not used optimally.

Our contribution addresses exactly this issue: Does the decoupling of trade and transport leave network capacity unused and ultimately results in economic inefficiencies? To the best of our knowledge, a detailed quantitative analysis of this effect is missing in the literature up to date. The first paper that provides a rigorous modeling of entry-exit-like systems is [24]. The four-level model developed in that paper comprises a two-stage market model that captures the booking and the nomination decisions and also includes an accurate representation of the network. We build on this model to compare three scenarios: (i) A nodal-pricing regime, (ii) an entry-exit regime, and (iii) a hybrid model in which technical capacities restrict the trade at the daily gas market, but the requirement is that only the market solution (instead of all possibly occurring market solutions) needs to be feasible. We interpret this model as a second-best benchmark, which would be relevant if the network operator could rely on experience when determining the technical capacities and therefore would not have to ensure that all possible nominations are feasible.

Note that the considered multilevel problems are a generalization of bilevel optimization problems that are known to be NP-hard [27, 34]. Thus, to provide an insightful quantitative analysis we have to make several assumptions. First, the market model in [24], which we build on, is based on the assumption of perfect competition among gas traders. As a matter of fact, the consideration of strategic interaction in markets in combination with complex physical flow models that restrict decisions makes an insightful analysis extremely challenging in our context. Our setup, on the contrary, allows for clear insights into the interplay of market interaction and physical gas transport. Furthermore, note that even in a perfect competition setup, the interaction of economic agents on a network, which is characterized by complex physical flow constraints, challenges standard economic analysis. In [23], for example, it is shown that the usual equivalence between the allocation in the welfare optimum and under perfect competition no longer holds. Besides the challenges arising from the multilevel structure of the market model, a correct modeling of the technical capacities set by the TSO results in an adjustable-robustness constraint in the upper level of the problem: The TSO must choose technical capacities such that every possible feasible nomination is transportable. Adjustable-robust optimization problems can be considered as two-stage optimization problems and are therefore challenging by themselves [4, 5, 57]. From another perspective, in [39] it is shown, that the problem of validating the feasibility of a booking—and technical capacities can be interpreted as bookings—is in coNP for the case of nonlinear (but algebraic) pressure loss functions and general networks. Going even further, a detailed modeling of gas physics introduces additional nonlinearities up to differential equations to accurately model pressure losses in pipes. Thus, and in line with the arguments in [24], we assume some reasonable simplifications for the physical flow modeling to keep the problem computationally tractable: (i) We consider only stationary gas flow, (ii) we do not include controllable elements like compressors or (control) valves, and (iii) we consider a potential-based gas flow model in which pressure losses linearly depend on the flow. We are aware of the fact that these physical assumptions, especially the linearity of gas flows, are strong assumptions. However, we are convinced that nonlinear gas flow models are far out of reach due to theoretical complexity reasons. We will discuss this issue later on in more detail. Nevertheless, we are convinced that our results shed some interesting light on the economic implications of the entry-exit gas market system in Europe.

The contribution of this paper is the following. We use the bilevel reformulation of the four-level entry-exit model presented in [24] and develop an exact single-level reformulation. To this end, we analyze the primal-dual optimality conditions of the bilevel problem's lower level in detail to derive bounds that can be used as problem-specific big-Ms. Moreover, we develop problem-tailored linearizations for the occurring nonsmooth and nonconvex constraints of the single-level reformulation. Finally, we re-state the challenging robustness constraint, which is needed to specify technical capacities, as a system of linear constraints by exploiting the characterization given in [39]. By doing so, we arrive at a model reformulation that can be tackled with state-of-the-art solvers. In a case study, we compare the results of the entry-exit regime described by this model to a first-best benchmark as well as a "second-best" regime in which technical capacities are determined as to make the resulting market allocation feasible, but not all allocations that might occur given the capacity restrictions at the nodes. Our case study yields several insights. First, we identify the cost of decoupling trade and transport by showing that the robustness constraint in an entry-exit market design accounts for significant welfare losses as compared to first-best and second-best models. Second, the design of the network has a significant impact on the specification of technical capacities at the individual nodes. Thus, our results show that, from a long-run perspective, trade and transport are not decoupled at all. Third, the booking price floors established by the TSO to collect payments from the traders to reimburse transportation costs may further reduce welfare and, depending on the used pricing regime, can lead to discrimination against individual actors. Fourth, the computational effort required to solve the problem mostly stems from the robustness constraint. Finally, our second-best benchmark captures the situation in which the network operator sets the technical capacities based on experience, and therefore, the robustness constraint does not need to be imposed in its strictest form. As one would expect, the welfare losses are lower than under the strict robustness constraint. Less restrictive feasibility constraints might also be possible in practice if interruptible contracts exist for some market participants. However, this case is too complex to be modeled and solved in the context of our formal analysis and case study.

Up to date, the literature on gas markets has mainly focused on strategic interaction of suppliers. Typically, stylized two-node networks are used to get insights on the potential effects of market power in network based industries; see, e.g., Cremer and Laffont [12], Ikonnikova and Zwart [32], Jansen et al. [33], Meran et al. [42], Oliver et al. [46], and Yang et al. [56]. Other contributions to the gas market literature analyze strategic interaction in gas markets using complementarity problems that allow to computationally derive equilibrium predictions. Those contributions typically rely on less restrictive assumptions regarding the analyzed network structure. However, they do not provide general analytical solutions of the market interaction. Examples are Baltensperger et al. [3], Boots et al. [7], Boucher and Smeers [8, 9], Chyong and Hobbs [11], Egging et al. [13, 14], Gabriel et al. [21], Holz et al. [29], Huppmann [31], Siddiqui and Gabriel [50], and Zwart and Mulder [59]. Some papers go beyond the classical linear network flow and account for the influence of pressure gradients between nodes of the network, e.g. Midthun et al. [43], Midthun et al. [44], and Rømo et al. [49]. Those papers focus on the assessment of infrastructure and dispatch for gas networks based on the so-called Weymouth equation [54]. The inefficiency of decoupling gas transport and trade due to unused network capacity has been argued by some authors; see, e.g., Smeers [51]. By now, however, a detailed analysis is missing. Still, there are some illustrative examples, e.g., in, Alonso et al. [1], Glachant et al. [22], Hallack and Vazquez [26], Hirschhausen [28], Hunt [30], and Vazquez et al. [52].

The paper is structured as follows. In Section 2, we review each level of the four-level model from [24] in detail and also state the bilevel reformulation given in [24]. Then, we develop an equivalent single-level reformulation of the model in Section 3 and propose further reformulations in Section 4. In Section 5, we tackle the robustness constraint. Finally, in Section 6 we measure inefficiencies arising in entry-exit-like systems and conclude in Section 7.

2. A Multilevel Model of the Entry-Exit Gas Market

In this section, we review the four-level model of the European entry-exit gas market as it has been developed in [24]. We briefly introduce each level and state an equivalent bilevel reformulation of the four-level model. All other details and the rationale of the multilevel modeling are given in [24].

In a nutshell, the four levels of the considered market environment model the following sequence of actions:

(i) Specification of technical capacities and booking price floors by the TSO.
(ii) Booking of capacity rights by gas buying and selling firms.
(iii) Day-ahead nomination by gas buying and selling firms.
(iv) Cost-optimal transport of the realized nominations by the TSO.

Before we formally state the particular optimization problem of each level we introduce some notation. Gas transport networks are modeled as a directed graph $G = (V, A)$ with node set V and arc set A. The node set is split up into the set of entry nodes $V_+ \subseteq V$ at which gas is supplied, the set of exit nodes $V_- \subseteq V$ at which gas is discharged, and the set of inner nodes $V_0 \subseteq V$ without gas supply or withdrawal. Thus, $V = V_+ \cup V_- \cup V_0$. The model allows for multiple gas selling or gas buying firms $i \in \mathcal{P}_u$ for $u \in V_+$ or $u \in V_-$, respectively.

Due to the general hardness of the four-level model, we need some important simplifying assumptions. In particular, we do not consider controllable elements such as compressors or control valves, we consider stationary gas flow, and we assume a linear pressure loss function. We will discuss these assumptions in more detail later, when they are formally introduced. We now describe every level in detail.

2.1. Level 1: Specification of Technical Capacities and Booking Price Floors.

In the first of the four levels of the model, the TSO specifies technical capacities and price floors in order to maximize total social welfare obtained in the market:

$$\max_{q^{\mathrm{TC}}, \underline{\pi}^{\mathrm{book}}} \quad \sum_{t \in T} \left(\sum_{u \in V_-} \sum_{i \in \mathcal{P}_u} \int_0^{q_{i,t}^{\mathrm{nom}}} P_{i,t}(s) \, \mathrm{d}s - \sum_{u \in V_+} \sum_{i \in \mathcal{P}_u} c_i^{\mathrm{var}} q_{i,t}^{\mathrm{nom}} \right) - \varphi^4(q^{\mathrm{nom}}) - C \tag{1a}$$

$$\text{s.t.} \quad 0 \le q_u^{\mathrm{TC}}, \ 0 \le \underline{\pi}_u^{\mathrm{book}} \quad \text{for all } u \in V_+ \cup V_-, \tag{1b}$$

$$\sum_{u \in V_+ \cup V_-} \sum_{i \in \mathcal{P}_u} \underline{\pi}_u^{\mathrm{book}} q_i^{\mathrm{book}} = \varphi^4(q^{\mathrm{nom}}) + C, \tag{1c}$$

$$\forall \hat{q}^{\mathrm{nom}} \in \mathcal{N}(q^{\mathrm{TC}}) : \mathcal{F}(\hat{q}^{\mathrm{nom}}) \neq \emptyset, \tag{1d}$$

$$q_i^{\mathrm{book}} \in \arg\max\,(3) \quad \text{for all } i \in \mathcal{P}_u, \ u \in V_+ \cup V_-, \tag{1e}$$

$$q_{i,t}^{\mathrm{nom}} \in \arg\max\,(4) \quad \text{for all } i \in \mathcal{P}_u, \ u \in V_+ \cup V_-, \ t \in T. \tag{1f}$$

Since we consider multiple time periods of gas trading and transport, total social welfare is aggregated over all time periods $t \in T$ with $|T| < \infty$. Gas buying firms are modeled by strictly decreasing inverse market demand functions $P_{i,t}$, $i \in \mathcal{P}_u$, $u \in V_-$, and gas selling firms are characterized by pairwise distinct variable costs of production c_i^{var}, $i \in \mathcal{P}_u$, $u \in V_+$. Throughout the paper, φ^k denotes the optimal value of the kth level of the multilevel model, i.e., $k \in \{1, 2, 3, 4\}$. For instance, φ^4 denotes the minimum costs of gas transport obtained in the fourth level. The decision variables of the first level include the technical capacities q_u^{TC} for every entry and exit node $u \in V_+ \cup V_-$ of the network. These variables limit the booking quantities q_i^{book} of the players i, which are decided on the second level; see Constraint (1e). The second set of first-level decision variables consists of the price floors $\underline{\pi}_u^{\mathrm{book}}$ for booking quantities at every entry and exit node $u \in V_+ \cup V_-$. Constraint (1c) models that the booking price floors are chosen such that the transport costs $\varphi^4(q^{\mathrm{nom}})$ arising in level four plus additional exogenously given network costs C, e.g., investment costs, are recovered. The actual nominations $q_{i,t}^{\text{nom}}$ are decided on the third level; see (1f). Note that exactly these nominations are entering the objective function of the first level.

Finally, we consider the Constraint (1d). This constraint claims that in dependence of the technical capacities q^{TC} all balanced nominations that potentially may occur, i.e.,

$$\mathcal{N}(q^{\text{TC}}) := \left\{ q \in \mathbb{R}^{V_+ \cup V_-} : 0 \le q \le q^{\text{TC}}, \sum_{u \in V_+} q_u = \sum_{u \in V_-} q_u \right\},$$

need to be feasible with respect to the network. In other words: Every balanced and node-wise aggregated nomination that satisfies given technical capacities needs to be transportable. This is formalized by the condition $\mathcal{F}(\hat{q}^{\text{nom}}) \neq \emptyset$, where $\mathcal{F}(\hat{q}^{\text{nom}})$ is the set of feasible points of the cost-minimization transport problem in the fourth level. We discuss this feasible set in more detail in Section 2.4. Constraint (1d) corresponds to an $\exists$-$\forall$-$\exists$ quantifier structure, i.e., there must exist a vector of technical capacities q^{TC} such that for all possible nominations $\hat{q}^{\text{nom}} \in \mathcal{N}(q^{\text{TC}})$ there exists a feasible point in $\mathcal{F}(\hat{q}^{\text{nom}})$. Thus, from the viewpoint of robust optimization (cf., e.g., [4, 6] and the many references therein), the constraint can be interpreted as an adjustable-robust and thus semi-infinite constraint [57] with here-and-now decisions q^{TC} and wait-and-see decisions in the feasible set of the fourth level. As such, Constraint (1d) adds significant difficulty to the overall model. We will discuss the handling of this constraint in more detail in Section 5 and its computational implications in Section 6.2. As previously mentioned, Constraint (1d) is also problematic from an economic point of view. Since every balanced nomination that is feasible w.r.t. the technical capacities must be guaranteed to be transportable by the TSO, this constraint is suspect to leave network capacities unused. Thus, it is very likely that the robustness constraint causes inefficiencies. The analysis of this effect is the main motivation for this paper and is taken care of in Section 6. Let us mention at this point that we are aware of that, in practice, the TSO might not respect the Constraint (1d) in its strictest form. Instead, she might draw on experience and set technical capacities that guarantee feasibility for likely market outcomes but not for all possible nominations. We account for this by analyzing a respective scenario in the case study in Section 6. Moreover, there are further capacity products such as so-called conditional, restrictively allocable, or interruptible capacities that are used to give the TSO more flexibility to ensure feasibility of the actually realized loads. The modeling of these capacity products is out of scope of this paper and we refer to Chapter 3 in the book [38] for additional details.

2.2. **Level 2: Booking.** At this level, each player $i \in \mathcal{P}_u$ with $u \in V_- \cup V_+$ books capacity rights q_i^{book} to maximize the anticipated revenue $\varphi_{i,t}^3$ (realized in the third level; see Section 2.3) minus booking costs. In line with [24], we assume perfect competition. In particular, we assume that bookings are not made strategically. This disallows to exclude competitors from the market by preemptive bookings and to drive up spot-market prices. The model of each player then reads as follows:

$$\begin{aligned}
\max_{q_i^{\text{book}}} \quad & \sum_{t \in T} \varphi_{i,t}^3(q_i^{\text{book}}) - (\underline{\pi}_u^{\text{book}} + \pi_u^{\text{book}}) q_i^{\text{book}} \\
\text{s.t.} \quad & q_i^{\text{book}} \ge 0, \\
& \sum_{i \in \mathcal{P}_u} q_i^{\text{book}} \le q_u^{\text{TC}}.
\end{aligned} \tag{2}$$

Players potentially compete for scarce technical capacities q^{TC} that are outcome of the first level. The booking price floor $\underline{\pi}_u^{\text{book}}$, that is also outcome of the first level, always applies for every booking. The additional markup π_u^{book} only occurs in case of scarce technical capacities, i.e., as the result of a competitive bidding process for the bookings. It is shown in [24] that the Problems (2) can be aggregated node-wise to obtain a mixed nonlinear complementarity problem (MNCP) per node under some additional assumptions. In the case study in Section 6, we only consider a single player i per node u. We thus refrain from a discussion of these assumptions here. The reasoning behind the modeling decision of one player per node is given in Section 6. In Theorem 1 in [24] it is further shown that the above mentioned MNCP is equivalent to the optimization problem

$$\varphi_u^2(q_u^{\mathrm{TC}}, \underline{\pi}_u) := \max_{q_i^{\mathrm{book}}} \quad \sum_{i \in \mathcal{P}_u} \sum_{t \in T} \varphi_{i,t}^3(q_i^{\mathrm{book}}) - \underline{\pi}_u^{\mathrm{book}} q_i^{\mathrm{book}} \tag{3a}$$

$$\text{s.t.} \quad q_i^{\mathrm{book}} \geq 0 \quad \text{for all } i \in \mathcal{P}_u, \tag{3b}$$

$$\sum_{i \in \mathcal{P}_u} q_i^{\mathrm{book}} \leq q_u^{\mathrm{TC}}, \tag{3c}$$

in which the markup price π_u^{book} is exactly the dual variable of Constraint (3c). This problem is solved at every entry and exit node of the network and all these problems are independent of each other.

2.3. **Level 3: Nomination.** At the third level, all players choose nominations restricted by their bookings of the second level in order to maximize their individual surplus at the equilibrium market price π_t^{nom}, which results endogenously at the third level. Under perfect competition, all players act as price takers. Thus, every gas seller $i \in \mathcal{P}_u$, $u \in V_+$, maximizes its surplus in every time period $t \in T$:

$$\varphi_{i,t}^3(q_i^{\mathrm{book}}) := \max_{q_{i,t}^{\mathrm{nom}}} \quad (\pi_t^{\mathrm{nom}} - c_i^{\mathrm{var}}) q_{i,t}^{\mathrm{nom}} \quad \text{s.t.} \quad 0 \leq q_{i,t}^{\mathrm{nom}} \leq q_i^{\mathrm{book}}.$$

Similarly, every gas buyer $i \in \mathcal{P}_u$, $u \in V_-$, maximizes its benefit in every time period $t \in T$:

$$\varphi_{i,t}^3(q_i^{\mathrm{book}}) := \max_{q_{i,t}^{\mathrm{nom}}} \quad \int_0^{q_i^{\mathrm{nom}}} P_{i,t}(s) \,\mathrm{d}s - \pi_t^{\mathrm{nom}} q_{i,t}^{\mathrm{nom}} \quad \text{s.t.} \quad 0 \leq q_{i,t}^{\mathrm{nom}} \leq q_i^{\mathrm{book}}.$$

Using first-order optimality conditions for every maximization problem, the nomination level can be modeled as an MNCP; see [24]. In Theorem 3 in [24] it is further shown that the resulting MNCP can be recast as an equivalent welfare maximization problem:

$$\varphi^3(q^{\mathrm{book}}) := \max_{q^{\mathrm{nom}}} \quad \sum_{t \in T} \left(\sum_{u \in V_-} \sum_{i \in \mathcal{P}_u} \int_0^{q_{i,t}^{\mathrm{nom}}} P_{i,t}(s) \,\mathrm{d}s - \sum_{u \in V_+} \sum_{i \in \mathcal{P}_u} c_i^{\mathrm{var}} q_{i,t}^{\mathrm{nom}} \right) \tag{4a}$$

$$\text{s.t.} \quad 0 \leq q_{i,t}^{\mathrm{nom}} \leq q_i^{\mathrm{book}} \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_+ \cup V_-,\ t \in T, \tag{4b}$$

$$\sum_{u \in V_-} \sum_{i \in \mathcal{P}_u} q_{i,t}^{\mathrm{nom}} - \sum_{u \in V_+} \sum_{i \in \mathcal{P}_u} q_{i,t}^{\mathrm{nom}} = 0 \quad \text{for all } t \in T. \tag{4c}$$

Constraint (4c) is the market clearing condition. The dual variable of this constraint is the market price π_t^{nom} in time period $t \in T$.

2.4. **Level 4: Cost-Optimal Transport of Nominations.** The fourth level is concerned with the cost-optimal transport of the nominations that are the outcome of the third level. The TSO solves the minimization problem

$$\min_{p,q} \quad \sum_{t \in T} c_t(q^{\mathrm{nom}}) \quad \text{s.t.} \quad (p, q) \in \mathcal{F}(q^{\mathrm{nom}}),$$

where $c_t(q^{\mathrm{nom}})$ is the transportation cost for the given nomination q^{nom}.

For the specification of the feasible set $\mathcal{F}(q^{\mathrm{nom}})$ that restricts gas pressures p and gas mass flows q, we follow [25] and [39]. For every node $u \in V$ of the gas transport network and time period $t \in T$ we denote the gas pressure by $p_{u,t}$ with bounds

$$0 < p_u^- \leq p_{u,t} \leq p_u^+ \leq \infty.$$

Further, gas mass flow on arc $a \in A$ in time period t is denoted by $q_{a,t}$. For an arc $a = (u, v)$, $q_{a,t} > 0$ is interpreted as flow in the direction of the arc, i.e., from u to v, and $q_{a,t} < 0$

as flow in the opposite direction. Additionally, the gas mass flow has to satisfy given capacities

$$-\infty \le q_a^- \le q_{a,t} \le q_a^+ \le \infty \quad \text{for all } a \in A,\ t \in T, \tag{5}$$

and mass balance at every node of the network is modeled using the constraints

$$\begin{aligned}
\sum_{a \in \delta_u^{\text{out}}} q_{a,t} - \sum_{a \in \delta_u^{\text{in}}} q_{a,t} &= \sum_{i \in \mathcal{P}_u} q_{i,t}^{\text{nom}} \quad \text{for all } u \in V_+,\ t \in T,\\
\sum_{a \in \delta_u^{\text{out}}} q_{a,t} - \sum_{a \in \delta_u^{\text{in}}} q_{a,t} &= -\sum_{i \in \mathcal{P}_u} q_{i,t}^{\text{nom}} \quad \text{for all } u \in V_-,\ t \in T,\\
\sum_{a \in \delta_u^{\text{out}}} q_{a,t} - \sum_{a \in \delta_u^{\text{in}}} q_{a,t} &= 0 \quad \text{for all } u \in V_0,\ t \in T.
\end{aligned} \tag{6}$$

In this formulation, δ_u^{out} represents outgoing edges and δ_u^{in} represents incoming edges at node u.

Further, gas pressure needs to be coupled at the incident nodes of an arc with the arc's gas mass flow. In a rather general form, this is achieved by the pressure loss law

$$p_{u,t}^2 - p_{v,t}^2 = \Phi_a(q_{a,t}) \quad \text{for all} \quad a = (u,v) \in A,\ t \in T,$$

where Φ_a denotes the pressure loss function for arc $a \in A$. We can substitute the squared pressure variables using $\pi_{u,t} = p_{u,t}^2$ for all $u \in V$, $t \in T$, and obtain the constraints

$$\pi_{u,t} - \pi_{v,t} = \Phi_a(q_{a,t}) \quad \text{for all } a = (u,v) \in A,\ t \in T, \tag{7}$$

together with the bounds

$$0 < \pi_u^- \le \pi_{u,t} \le \pi_u^+ \le \infty \quad \text{for all } u \in V,\ t \in T. \tag{8}$$

We make the following assumption.

Assumption 1. *The pressure loss function Φ_a is linear for all $a \in A$.*

This assumption renders the transport model of this section linear. Technically, this assumption is not needed for the bilevel reformulation in Section 2.5 and the single-level reformulation in Section 3. However, without this assumption, the transport model occurring in the upper level of the bilevel problem will be nonlinear, which further complicates an already challenging problem. Moreover, the linear reformulation of the robustness constraint (1d) stated in Section 5 is not possible without Assumption 1 unless compact global optimality certificates are available for nonconvex problems. This, however, is strongly related to the P vs. NP problem, which is why Assumption 1 is needed for reasons of computational tractability.

Lastly, we need to specify the transportation costs. As mentioned, we only consider passive gas transport networks in this paper. These are networks that do not contain controllable network elements like compressors or (control) valves. However, transportation costs mainly arise from these controllable elements such as compressors. Usually, transportation costs are driven by pressure losses across the network. In order to mimic cost-optimal transport in our setting, the objective function in the fourth level minimizes costs of squared pressure losses in the entire network that are given by the nonsmooth expression

$$\sum_{t \in T} \sum_{a=(u,v) \in A} c_t^{\text{trans}} |\pi_{u,t} - \pi_{v,t}|,$$

where $c_t^{\text{trans}} > 0$ is a parameter. In total, the problem at level four reads

$$\varphi^4(q^{\text{nom}}) := \min_{\pi, q} \quad \sum_{t \in T} \sum_{a=(u,v) \in A} c_t^{\text{trans}} |\pi_{u,t} - \pi_{v,t}| \quad \text{s.t.} \quad (5)\text{–}(8). \tag{9}$$

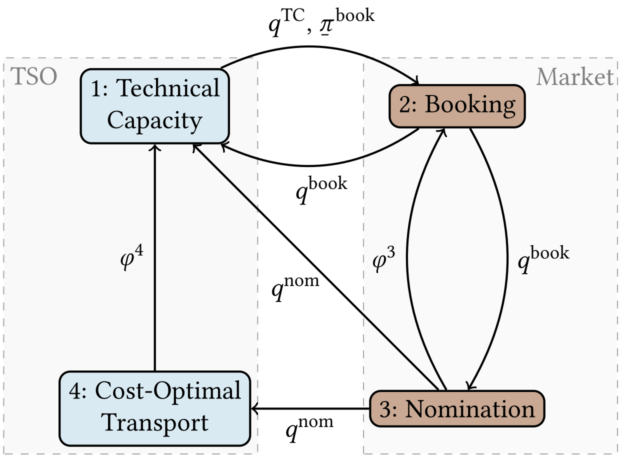

FIGURE 1. Dependencies between the four levels; taken from [24].

2.5. **Reduction to a Bilevel Problem.** The structure of the four-level model presented in the previous sections is rather complicated. Figure 1 sheds more light on the dependencies of the levels and the structure of the full model. It can be seen that, e.g., the variables of the fourth level do not appear in the constraints of any other level. In Section 3 of [24] it is shown that this structure can be exploited to equivalently re-state the four-level model as a bilevel model. In particular, in Theorem 5 in [24] it is shown that the original second- and third-level problem can be merged into an equivalent single-level problem. Thus, the four-level model can be reduced to an equivalent trilevel one. In addition, in Theorem 7 in [24] it is shown that this trilevel model can be further reduced to the following bilevel model by merging the original fourth-level problem into the original first-level problem.

$$
\begin{aligned}
\max_{q^{\mathrm{TC}}, \underline{\pi}^{\mathrm{book}}, \pi, q} \quad & \varphi^{\mathrm{u}}(q^{\mathrm{nom}}, \pi) = \sum_{t \in T} \left(\sum_{u \in V_-} \sum_{i \in \mathcal{P}_u} \int_0^{q_{i,t}^{\mathrm{nom}}} P_{i,t}(s) \, \mathrm{d}s - \sum_{u \in V_+} \sum_{i \in \mathcal{P}_u} c_i^{\mathrm{var}} q_{i,t}^{\mathrm{nom}} \right) && \text{(10a)} \\
& \qquad - \sum_{t \in T} \sum_{a=(u,v) \in A} c_t^{\mathrm{trans}} (|\pi_{u,t} - \pi_{v,t}|) - C \\
\text{s.t.} \quad & 0 \le q_u^{\mathrm{TC}}, \ 0 \le \underline{\pi}_u^{\mathrm{book}} \quad \text{for all } u \in V_+ \cup V_-, && \text{(10b)} \\
& \sum_{u \in V_+ \cup V_-} \sum_{i \in \mathcal{P}_u} \underline{\pi}_u^{\mathrm{book}} q_i^{\mathrm{book}} = \sum_{t \in T} \sum_{a=(u,v) \in A} c_t^{\mathrm{trans}} |\pi_{u,t} - \pi_{v,t}| + C, && \text{(10c)} \\
& \forall \hat{q}^{\mathrm{nom}} \in \mathcal{N}(q^{\mathrm{TC}}) : \mathcal{F}(\hat{q}^{\mathrm{nom}}) \neq \emptyset, && \text{(10d)} \\
& (\pi, q) \in \mathcal{F}(q^{\mathrm{nom}}), \text{ i.e., } \pi, q \text{ fulfill (5)–(8)} && \text{(10e)} \\
& (q^{\mathrm{book}}, q^{\mathrm{nom}}) \in \arg\max \text{(11)}, && \text{(10f)}
\end{aligned}
$$

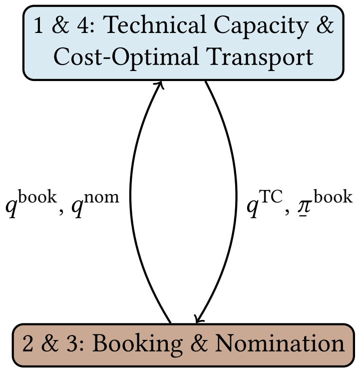

FIGURE 2. Dependencies between the two levels of the reduced bilevel problem; taken from [24].

where the lower level is given by

$$\max_{q^{\text{book}}, q^{\text{nom}}} \quad \sum_{t \in T} \left(\sum_{u \in V_-} \sum_{i \in \mathcal{P}_u} \int_0^{q_{i,t}^{\text{nom}}} P_{i,t}(s)\, \mathrm{d}s - \sum_{u \in V_+} \sum_{i \in \mathcal{P}_u} c_i^{\text{var}} q_{i,t}^{\text{nom}} \right) \tag{11a}$$
$$\qquad - \sum_{u \in V_+ \cup V_-} \sum_{i \in \mathcal{P}_u} \underline{\pi}_u^{\text{book}} q_i^{\text{book}}$$
$$\text{s.t.} \quad \sum_{i \in \mathcal{P}_u} q_i^{\text{book}} \leq q_u^{\text{TC}} \quad \text{for all } u \in V_+ \cup V_-, \qquad [\pi_u^{\text{book}}] \tag{11b}$$
$$0 \leq q_{i,t}^{\text{nom}} \leq q_i^{\text{book}} \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_+ \cup V_-,\ t \in T, \qquad [\gamma_{i,t}^{\pm}] \tag{11c}$$
$$\sum_{u \in V_-} \sum_{i \in \mathcal{P}_u} q_{i,t}^{\text{nom}} - \sum_{u \in V_+} \sum_{i \in \mathcal{P}_u} q_{i,t}^{\text{nom}} = 0 \quad \text{for all } t \in T. \qquad [\pi_t^{\text{nom}}] \tag{11d}$$

The upper level (10) represents the combined first and fourth level of the original four-level model, i.e., the TSO's actions, whereas the lower level (11) consists of the original second and third level and models the market interaction. The lower-level problem has a unique solution under reasonable assumption, see Theorem 6 in [24], so that we do not have to consider aspects related to optimistic or pessimistic bilevel solutions. Note that we omitted the redundant second-level constraint $0 \leq q_i^{\text{book}}$ since it is trivially fulfilled due to Constraint (11c). The dependencies between the two levels are illustrated in Figure 2. The bilevel structure is one of the major challenges of the problem. Bilevel problems are inherently nonconvex and even linear bilevel problems are NP hard [27, 34]. Apart from that the problem is challenging because the upper level itself is challenging due to at least three reasons. First, the upper level has nonsmooth terms in the objective function and in Constraint (10c). Second, Constraint (10c) contains nonconvex bilinear terms $\underline{\pi}_u^{\text{book}} q_i^{\text{book}}$. Third, and most importantly, the upper level contains the semi-infinite robustness constraint (10d), which is, in general, not tractable from a computational point of view. We tackle the bilevel structure in Section 3, the nonsmooth and bilinear terms of the upper level in Section 4, and the robustness constraint in Section 5. Finally, we point out that without Assumption 1, the modeling of the gas transport in Constraint (10e) would be nonlinear and thus introduce additional difficulties to the upper level of the bilevel reformulation.

2.6. **Benchmark Models.** Due to the overall hardness of Problem (10), we introduce two relaxations that serve as a benchmark in the following.

A first-best benchmark can be obtained by optimizing the total welfare considering both the network and market interaction constraints. In bilevel programming, this corresponds to the so-called high-point relaxation; see [45]. From an economic perspective this first-best benchmark assumes an integrated gas company that is fully regulated and that takes all decisions in a welfare-optimal manner. Consequently, we can remove the need for a prior determination of technical capacities and bookings. This renders also the robustness constraint (10d) redundant. We thus obtain a model that determines the decisions of the original third and fourth level, accounting for welfare optimal production and demand as well as cost-optimal transport simultaneously:

$$\max_{q^{\mathrm{nom}} \geq 0, \pi, q} \quad \varphi^{\mathrm{u}}(q^{\mathrm{nom}}, \pi) \quad \text{s.t.} \quad (q^{\mathrm{nom}}, \pi, q) \text{ satisfies (5)–(8).} \tag{12}$$

Another benchmark model can be obtained by omitting only the robustness constraint (10d), which is one of the major difficulties of the bilevel problem (10):

$$\max_{q^{\mathrm{TC}}, \underline{\pi}^{\mathrm{book}}, \pi, q} \quad \varphi^{\mathrm{u}}(q^{\mathrm{nom}}, \pi) \tag{13a}$$

$$\text{s.t.} \quad \text{(10b), (10c), and (10e),} \tag{13b}$$

$$(q^{\mathrm{book}}, q^{\mathrm{nom}}) \in \arg\max \text{(11)}. \tag{13c}$$

From an economic point of view this model assumes that the TSO needs to make sure that only realized nominations need to be transportable. Thus, this model serves as a direct benchmark to measure the inefficiencies caused by the robustness constraint (10d). In this sense, Problem (13) can be interpreted as a second-best model that captures, for instance, a situation in which the network operator is able to set the technical capacities based on experience.

It is easy to see that out of the three models, Problem (12) yields the best welfare outcome and bounds the welfare outcome of Problem (13). The latter, in turn, obviously bounds the welfare outcome of the bilevel reformulation (10) of the four level entry-exit model. In Section 6, we use this model hierarchy to measure inefficiencies in entry-exit-like systems.

3. Reduction to a Single-Level Problem

In the previous section we discussed a four-level model of an entry-exit-like system and reduced it to an equivalent bilevel problem with a convex lower level as described in [24]. In recent years, several algorithms evolved for bilevel problems that exploit certain structural properties. However, most—if not all—bilevel-tailored algorithms whose performance could be demonstrated in computational studies, e.g., [18, 35, 40, 55], are either only applicable to (mixed-integer) linear bilevel problems or rely on the property of integer linking variables (upper-level variables that are present in the lower-level constraints). All these algorithms are not appropriate for our setting of nonlinear upper- and lower-level problems and continuous linking variables q^{TC} and $\underline{\pi}^{\mathrm{book}}$. Thus, we resort to the well-known standard technique of replacing the convex lower level by its necessary and sufficient first-order optimality conditions to obtain a single-level problem. Note that this approach requires Slater's constraint qualification to hold. In the following, we first perform a single-level reformulation based on the Karush–Kuhn–Tucker (KKT) conditions and, in a second step, we provide an exact mixed-integer linear reformulation of the KKT complementarity conditions.

3.1. A KKT-Based Single-Level Reformulation.

A single-level reformulation can be obtained by replacing (10f) either by the strong duality conditions or the KKT conditions of Problem (11). In any way, we use the following assumption to obtain a computationally tractable reformulation.

Assumption 2. *All inverse market demand functions are linear and strictly decreasing, i.e.,* $P_{i,t}(q^{\mathrm{nom}}_{i,t}) = a_{i,t} + b_{i,t} q^{\mathrm{nom}}_{i,t}$ *with* $a_{i,t} > 0$ *and* $b_{i,t} < 0$.

This yields concave-quadratic upper- and lower-level objective functions w.r.t. the upper- or lower-level variables, respectively. Thus, with Assumption 2, the lower-level problem (11) is a concave-quadratic maximization problem over linear constraints and the strong-duality theorem of convex optimization is applicable. Furthermore, its KKT conditions are both necessary and sufficient; see, e.g., [10]. Following the strong-duality approach, one would replace the lower level by primal and dual feasibility as well as the strong-duality equation. Even for linear objective functions, the strong-duality equation of the lower level yields nonconvex bilinear terms of primal upper-level variables and dual lower-level variables; see, e.g., [58]. Thus, we refrain from using the strong-duality reformulation and instead replace the lower-level problem by its KKT conditions, i.e., by the stationarity conditions

$$-\underline{\pi}_u^{\text{book}} - \pi_u^{\text{book}} + \sum_{t \in T} \gamma_{i,t}^+ = 0 \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_+ \cup V_-, \tag{14a}$$

$$a_{i,t} + b_{i,t} q_{i,t}^{\text{nom}} + \gamma_{i,t}^- - \gamma_{i,t}^+ - \pi_t^{\text{nom}} = 0 \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_-,\ t \in T, \tag{14b}$$

$$-c_i^{\text{var}} + \gamma_{i,t}^- - \gamma_{i,t}^+ + \pi_t^{\text{nom}} = 0 \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_+,\ t \in T, \tag{14c}$$

primal feasibility (11b)–(11d), nonnegativity

$$\pi_u^{\text{book}} \geq 0 \quad \text{for all } u \in V_+ \cup V_-, \tag{15a}$$

$$\gamma_{i,t}^-, \gamma_{i,t}^+ \geq 0 \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_+ \cup V_-,\ t \in T, \tag{15b}$$

and complementarity constraints

$$\pi_u^{\text{book}} \left(q_u^{\text{TC}} - \sum_{i \in \mathcal{P}_u} q_i^{\text{book}} \right) = 0 \quad \text{for all } u \in V_+ \cup V_-, \tag{16a}$$

$$\gamma_{i,t}^- q_{i,t}^{\text{nom}} = 0 \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_+ \cup V_-,\ t \in T, \tag{16b}$$

$$\gamma_{i,t}^+ \left(q_i^{\text{book}} - q_{i,t}^{\text{nom}} \right) = 0 \quad \text{for all } i \in \mathcal{P}_u,\ u \in V_+ \cup V_-\ t \in T. \tag{16c}$$

Then, the bilevel problem (10) can be equivalently rephrased as the single-level problem

$$\begin{aligned} \max_z \quad & \varphi^{\text{u}}(q^{\text{nom}}, \pi) \\ \text{s.t.} \quad & \text{upper-level feasibility: (10b)–(10e)}, \\ & \text{lower-level KKT conditions: (11b)–(11d), (14)–(16)}, \end{aligned} \tag{17}$$

where $z = (q^{\text{TC}}, \underline{\pi}^{\text{book}}, q^{\text{book}}, q^{\text{nom}}, \pi, q, \pi^{\text{nom}}, \pi^{\text{book}}, \gamma^{\pm})$ is the vector of primal upper-level as well as primal and dual lower-level variables.

3.2. **Linearization of KKT Complementarity Conditions.** The complementarity conditions (16) are nonconvex but can be linearized using additional binary variables and big-M constants; see [19]. For a linear constraint $a^\top x \leq b$ and its dual variable $\lambda \geq 0$, the complementarity condition reads

$$\lambda(b - a^\top x) = 0. \tag{18}$$

For suitable primal and dual big-M constants M_p, M_d and a binary variable $u \in \{0, 1\}$, (18) can be linearized as follows:

$$b - a^\top x \leq M_p u, \quad \lambda \leq M_d(1 - u). \tag{19}$$

Finding suitable values for M_p and M_d is crucial for a correct linearization. Often, these values are obtained heuristically. In a bilevel context, this may result in suboptimal or infeasible solutions [47]. In fact, finding correct big-Ms may in general be as hard as solving the original bilevel problem [36]. Sometimes, however, problem-specific knowledge can be used to obtain correct big-M values. In our application, finding primal big-Ms M_p is easy because q^{TC}, q^{book}, and q^{nom} are bounded by, e.g., the maximum demand of all gas buyers. Finding dual big-Ms M_d is more complicated. In the following, we derive suitable values similar to [37]. Therefore, we use the notation $\mathcal{P}_- := \cup_{u \in V_-} \mathcal{P}_u$ and $\mathcal{P}_+ := \cup_{u \in V_+} \mathcal{P}_u$ as well as $a_t^{\min} := \min\{a_{i,t} : i \in \mathcal{P}_-\}$, $a_t^{\max} := \max\{a_{i,t} : i \in \mathcal{P}_-\}$, $c_{\min}^{\text{var}} := \min\{c_i^{\text{var}} : i \in \mathcal{P}_+\}$, and $c_{\max}^{\text{var}} := \max\{c_i^{\text{var}} : i \in \mathcal{P}_+\}$ and make the following assumption.

Assumption 3. *For every $t \in T$ it holds $a_t^{\min} \geq c_{\min}^{\text{var}}$ and $a_t^{\max} \geq c_{\max}^{\text{var}}$.*

The economic rationale is that every player is eager to participate in the market in every time period t. The gas buyer with the smallest initial willingness to pay $a_t^{\min}$ could potentially buy gas from the seller with the lowest variable costs $c_{\min}^{\text{var}}$ and the seller with the highest variable costs $c_{\max}^{\text{var}}$ could potentially sell gas to the buyer with the highest initial willingness to pay $a_t^{\max}$. This assumption can be easily checked a priori. Furthermore, we assume that the network is designed in a way such that in every time period we have trade.

Assumption 4. *For every time period $t \in T$ there exists a player $i \in V_+ \cup V_-$ with $q_{i,t}^{\text{nom}} > 0$.*

We first derive bounds on the free dual variable π_t^{nom}. As described in Section 2.3, this variable corresponds to the market price for gas in time period $t \in T$. We do not need bounds on this variable for a correct linearization of the KKT complementarity conditions (16), but bounds on π_t^{nom} will help to bound the other dual variables.

Lemma 1. *For every $t \in T$, it holds $c_{\min}^{\text{var}} \leq \pi_t^{\text{nom}} \leq a_t^{\max}$.*

Proof. Due to Assumption 4 and the market clearing condition (11d), there must exist players $i_- \in \mathcal{P}_-$ and $i_+ \in \mathcal{P}_+$ with $q_{i_-,t}^{\text{nom}} > 0$ and $q_{i_+,t}^{\text{nom}} > 0$. Applying KKT complementarity (16b) and dual feasibility (14b) to i_- and i_+, one obtains

$$\pi_t^{\text{nom}} = a_{i_-,t} + b_{i_-,t} q_{i_-,t}^{\text{nom}} - \gamma_{i_-,t}^+ < a_{i_-,t} \leq a_t^{\max} \quad \text{and} \quad \pi_t^{\text{nom}} = c_{i_+}^{\text{var}} + \gamma_{i_+,t}^+ \geq c_{i_+}^{\text{var}} \geq c_{\min}^{\text{var}},$$

respectively. □

We now turn to the dual variables $\gamma_{i,t}^{\pm}$. The variable $\gamma_{i,t}^-$ denotes the price range that the market price π_t^{nom} would have to decrease (for gas buyers) or increase (for gas sellers) for player i to enter the market in time period t. Contrary, $\gamma_{i,t}^+$ indicates the benefit of player i when trading a further unit.

Lemma 2. *It exists an optimal solution of the single-level reformulation (17) with*

$$\gamma_{i,t}^- \leq 2\left(a_t^{\max} - c_{\min}^{\text{var}}\right) \quad \text{and} \quad \gamma_{i,t}^+ \leq a_t^{\max} - c_{\min}^{\text{var}}$$

for all time periods $t \in T$ and $i \in \mathcal{P}_+ \cup \mathcal{P}_-$.

Proof. We only consider exit nodes $u \in V_-$. The case of $u \in V_+$ can be shown in exactly the same way. Let $t \in T$ and let z be an optimal solution for (17) with q_u^{book}, $q_{u,t}^{\text{nom}}$, $\underline{\pi}_u^{\text{book}}$, π_t^{nom}, $\bar{\pi}_u^{\text{book}}$, and $\gamma_{u,t}^{\pm}$ being part of this optimal solution. We distinguish two cases:

(i) $\sum_{i \in \mathcal{P}_u} q_i^{\text{book}} > 0$ and
(ii) $\sum_{i \in \mathcal{P}_u} q_i^{\text{book}} = 0$.

In case of (i), there exists a player $i_- \in \mathcal{P}_u$ with $q_{i_-}^{\text{book}} > 0$. If $q_{i_-,t}^{\text{nom}} > 0$, KKT complementarity (16b) yields $\gamma_{i_-,t}^- = 0$ and the stationarity condition (14b) simplifies to

$$\gamma_{i_-,t}^+ = a_{i_-,t} - \pi_t^{\text{nom}} + b_{i_-} q_{i_-,t}^{\text{nom}} < a_{i_-,t} - \pi_t^{\text{nom}} \leq a_t^{\max} - c_{\min}^{\text{var}}$$

by applying $b_{i_-} < 0$ and Lemma 1. If on the other hand $q_{i_-,t}^{\text{nom}} = 0$, KKT complementarity (16c) yields $\gamma_{i_-,t}^+ = 0$ and from KKT stationarity (14b) we obtain

$$\gamma_{i_-,t}^- = \pi_t^{\text{nom}} - a_{i_-,t} \leq a_t^{\max} - c_{\min}^{\text{var}}$$

by using Lemma 1 and Assumption 3. Thus, the claim is fulfilled in Case (i).

We now turn to Case (ii). In this case, $q_i^{\text{book}} = q_i^{\text{nom}} = 0$ holds. Let $i \in \mathcal{P}_u$. We set

$$\tilde{\underline{\pi}}_u^{\text{book}} := \sum_{t \in T} \left(a_t^{\max} - c_{\min}^{\text{var}}\right) \quad \text{and} \quad \tilde{\bar{\pi}}_u^{\text{book}} := 0$$

as well as

$$\tilde{\gamma}_{i,t}^- := a_t^{\max} - a_{i,t} + \pi_t^{\text{nom}} - c_{\min}^{\text{var}} \quad \text{and} \quad \tilde{\gamma}_{i,t}^+ := a_t^{\max} - c_{\min}^{\text{var}}. \tag{20}$$

Now, let $\tilde{z}$ be the vector that we obtain by replacing the corresponding entries of z by $\tilde{\underline{\pi}}_u^{\text{book}}$, $\tilde{\pi}_u^{\text{book}}$, and $\tilde{\gamma}_{u,t}^{\pm}$. It is easy to see that stationarity (14a) and (14b) as well as complementarity (16) is fulfilled. Further, nonnegativity (10b) of $\tilde{\underline{\pi}}_u^{\text{book}}$ and nonnegativity (15b) of $\tilde{\gamma}_{i,t}^{\pm}$ hold due to Lemma 1 and Assumption 3. Primal feasibility (11b)–(11d) of the lower level remains unchanged and the upper-level constraint (10c) still holds because $\sum_{i\in\mathcal{P}_u} \underline{\pi}_u^{\text{book}} q_i^{\text{book}} = \sum_{i\in\mathcal{P}_u} \tilde{\underline{\pi}}_u^{\text{book}} q_i^{\text{book}} = 0$ according to $q_i^{\text{book}} = 0$. All other upper-level constraints are trivially fulfilled by the modified vector $\tilde{z}$, and $\tilde{z}$ is thus feasible for (17). Since the objective function value is not affected by the modifications, $\tilde{z}$ is also optimal. Finally, from (20) we obtain

$$\tilde{\gamma}_{i,t}^- \le a_t^{\max} - c_{\min}^{\text{var}} + a_t^{\max} - c_{\min}^{\text{var}} = 2\left(a_t^{\max} - c_{\min}^{\text{var}}\right)$$

due to Lemma 1 and Assumption 3 and

$$\tilde{\gamma}_{i,t}^+ = a_t^{\max} - c_{\min}^{\text{var}}.$$

Thus, also in Case (ii) the claim is fulfilled. □

The dual variable π_u^{book} denotes the markup price for bookings; see the discussion in Section 2.2. It can be bounded as follows.

Lemma 3. *It exists an optimal solution of the single-level reformulation (17) with*

$$\pi_u^{\text{book}} \le \sum_{t\in T} \left(a_t^{\max} - c_{\min}^{\text{var}}\right)$$

for all $u \in V_+ \cup V_-$.

Proof. From KKT stationarity (14a), the nonnegativity of $\underline{\pi}_u^{\text{book}}$ (see Constraint (10b)), and Lemma 2, we obtain

$$\pi_u^{\text{book}} = \sum_{t\in T} \gamma_{i,t}^+ - \underline{\pi}_u^{\text{book}} \le \sum_{t\in T} \left(a_t^{\max} - c_{\min}^{\text{var}}\right). \qquad \square$$

In total, we can now replace Constraint (16) in the single-level reformulation (17) by its linearized variant according to (19). Finally, we end up with a single-level problem that is significantly larger than the bilevel problem (10) due to additional continuous and binary variables and constraints. As previously mentioned, this single-level problem still contains nonsmooth and nonconvex terms in the objective function and in Constraint (10c) and the semi-infinite robustness constraint (10d). In the following Sections 4 and 5 we further reformulate the single-level problem to obtain a computationally tractable problem.

4. Further Reformulations of the Upper Level

In this section, we linearize the nonsmooth absolute values $|\pi_{u,t} - \pi_{v,t}|$ and deal with the nonconvex bilinear terms $\underline{\pi}_u^{\text{book}} q_i^{\text{book}}$ that both appear in the single-level reformulation (17). First, we take care of the former. Without Constraint (10c), the absolute values would only appear in the objective function (10a). In this case, they could be linearized with standard techniques by introducing additional continuous variables and constraints. However, Constraint (10c) couples the absolute values $|\pi_{u,t} - \pi_{v,t}|$ to the bilinear terms $\underline{\pi}_u^{\text{book}} q_i^{\text{book}}$. Thus, besides additional continuous variables $\Delta\pi_{u,v,t} \ge 0$ for all $a = (u, v) \in A$, $t \in T$, we also need additional binary variables $x_{a,t}$ for all $a \in A$, $t \in T$ and the constraints

$$\Delta\pi_{u,v,t} \ge \pi_{u,t} - \pi_{v,t}, \qquad \Delta\pi_{u,v,t} \le 2\pi_v^+ x_{a,t} + \pi_{u,t} - \pi_{v,t}, \tag{21a}$$

$$\Delta\pi_{u,v,t} \ge \pi_{v,t} - \pi_{u,t}, \qquad \Delta\pi_{u,v,t} \le 2\pi_u^+ (1 - x_{a,t}) + \pi_{v,t} - \pi_{u,t}, \tag{21b}$$

for all $a = (u, v) \in A$, $t \in T$. It is easy to see that for $x_{a,t} = 0$, Constraints (21a) are active and enforce $\Delta\pi_{u,v,t} = \pi_{u,t} - \pi_{v,t}$, whereas Constraints (21b) are inactive due to the squared pressure bounds (8). The same holds with flipped roles for $x_{a,t} = 1$. We can thus replace all terms $|\pi_{u,t} - \pi_{v,t}|$ by the continuous variables $\Delta\pi_{u,v,t}$ and Constraints (21).

Second, we turn to the bilinear terms in Constraint (10c). Given upper bounds $\bar{\underline{\pi}}_u^{\text{book}}$ for $\underline{\pi}_u^{\text{book}}$ and $\bar{q}_i^{\text{book}}$ for q_i^{book} for every $u \in V_+ \cup V_-$ and $i \in \mathcal{P}_u$, these bilinear terms can be approximated using a piecewise linearization such as, e.g., the incremental method. See [53] for a detailed overview and a numerical comparison of existing methods. On the other hand, modern solvers like, e.g., Gurobi, can handle bilinear terms by using McCormick envelopes [41] in combination with spatial branching. A preliminary numerical study revealed that in our case the latter approach using Gurobi 9 works significantly better in terms of running times. Since the spatial-branching-based approach also requires upper bounds for both continuous variables $\underline{\pi}_u^{\text{book}}$ and q_i^{book}, we derive them in the following. We first state bounds for $\underline{\pi}_u^{\text{book}}$.

Lemma 4. *It exists an optimal solution of the single-level reformulation (17) with*

$$\underline{\pi}_u^{\text{book}} \le \sum_{t \in T} \left(a_t^{\max} - c_{\min}^{\text{var}}\right)$$

for all $u \in V_+ \cup V_-$.

Proof. From the lower-level KKT stationarity condition (14a), nonnegativity (15a), and Lemma 2, we obtain

$$\underline{\pi}_u^{\text{book}} = \sum_{t \in T} \gamma_{i,t}^+ - \pi_u^{\text{book}} \le \sum_{t \in T} \left(a_t^{\max} - c_{\min}^{\text{var}}\right). \qquad \square$$

In order to bound q_i^{book}, we first discuss that it is feasible to express bookings in terms of maximal nominations, i.e., to set $\bar{q}_i^{\text{book}} = \max_{t \in T}\{q_i^{\text{nom}}\}$.

Lemma 5. *Suppose a player $i \in \mathcal{P}_u$ at a node $u \in V_+ \cup V_-$ with*

$$\max_{t \in T}\{q_{i,t}^{\text{nom}}\} < q_i^{\text{book}}.$$

Then, it holds $\underline{\pi}_u^{\text{book}} = \pi_u^{\text{book}} = 0$.

Proof. Let $i \in \mathcal{P}_u$ be a player at node $u \in V_+ \cup V_-$ with $\max_{t \in T}\{q_{i,t}^{\text{nom}}\} < q_i^{\text{book}}$. Then, we know from KKT complementarity (16c) that the dual variables $\gamma_{i,t}^+$ must be zero for all $t \in T$. Thus, KKT stationarity (14a) and nonnegativity (15a) and (10b) give

$$0 \le \pi_u^{\text{book}} = -\underline{\pi}_u^{\text{book}} \le 0,$$

i.e., $\underline{\pi}_u^{\text{book}} = \pi_u^{\text{book}} = 0$. $\square$

This means that player $i \in \mathcal{P}_u$, $u \in V_+ \cup V_-$, only books above the maximum nomination, if the booking price floor $\underline{\pi}_u^{\text{book}}$ and the markup π_u^{book} are zero. In other words, the remaining booking $\Delta_i^{\text{book}} = q_i^{\text{book}} - \max_{t \in T}\{q_{i,t}^{\text{nom}}\} > 0$ does not affect the total welfare φ^1, which yields the following corollary.

Corollary 1. *Among the optimal solutions of the bilevel problem (10), there exists a solution for which*

$$\max_{t \in T}\{q_{i,t}^{\text{nom}}\} = q_i^{\text{book}}$$

holds for all $i \in \mathcal{P}$, $u \in V_+ \cup V_-$.

Thus, we can bound bookings by bounds for nominations. Since nominations must be feasible for the physical network, we can bound nominations by the capacity of adjacent arcs a to the node u.

Proposition 1. *With*

$$\bar{q}_i^{\text{nom}} = \sum_{a=(u,v)\in A} \max\{0, q_a^+\} + \sum_{a=(v,u)\in A} \max\{0, -q_a^-\} \quad \textit{for all } i \in \mathcal{P}_u, u \in V_+,$$
$$\bar{q}_i^{\text{nom}} = \sum_{a=(u,v)\in A} \max\{0, -q_a^-\} + \sum_{a=(v,u)\in A} \max\{0, q_a^+\} \quad \textit{for all } i \in \mathcal{P}_u, u \in V_-,$$

it holds

$$0 \leq q_{i,t}^{\text{nom}} \leq \bar{q}_i^{\text{nom}}$$

for all $i \in \mathcal{P}_u$, $u \in V_+ \cup V_-$ *and* $t \in T$.

Since the bounds $\bar{q}_i^{\text{nom}}$ are independent of $t \in T$, we directly obtain the following corollary from Corollary 1 and Lemma 1.

Corollary 2. *It holds* $0 \leq q_i^{\text{book}} \leq \bar{q}_i^{\text{nom}} = \bar{q}_i^{\text{book}}$ *for all* $i \in \mathcal{P}_u$, $u \in V_+ \cup V_-$.

The derived bounds can now be used in a spatial branching approach to tackle the bilinear terms $\underline{\pi}_u^{\text{book}} q_i^{\text{book}}$.

5. Handling of the Upper Level's Robustness Constraint

In this section, we deal with the robustness constraint (10d) that is part of the upper-level problem of the bilevel reformulation (10), respectively part of the single-level reformulation (17). We briefly recap that this constraint reads

$$\forall \hat{q}^{\text{nom}} \in \mathcal{N}(q^{\text{TC}}) : \mathcal{F}(\hat{q}^{\text{nom}}) \neq \emptyset,$$

where

$$\mathcal{N}(q^{\text{TC}}) = \left\{ q \in \mathbb{R}^{V_+ \cup V_-} : 0 \leq q \leq q^{\text{TC}}, \sum_{u \in V_+} q_u = \sum_{u \in V_-} q_u \right\},$$

and $\mathcal{F}(\hat{q}^{\text{nom}})$ is the set of feasible points $(q_a)_{a \in A}$ and $(\pi_u)_{u \in V}$, i.e., the points that satisfy (5)–(8) for a given nomination $\hat{q}^{\text{nom}}$. As already mentioned in Section 2.1, this can be seen as an adjustable-robust constraint. As such, it is semi-infinite and one of the major challenges of the entire problem. One opportunity to deal with this constraint is to leave it out entirely, which yields a relaxation of the problem; see Section 2.6. On the other hand, Theorem 10 in [39] gives a characterization for feasible bookings that we can use to reformulate this constraint. From this characterization it also follows directly that—in contrast to nonlinear pressure loss functions—feasible bookings can be validated in polynomial time in case of linear pressure loss functions $\Phi_a(\cdot)$. The reason is that the characterization incorporates the solution of several optimization problems. In case of linear pressure loss functions, all these optimization problems are linear. Thus, one can use a mixed-integer linear reformulation of the optimality conditions of these optimization problems to obtain a mixed-integer linear reformulation of the robustness constraint (10d). This is not possible anymore for nonlinear pressure loss functions. Thus, Assumption 1 is crucial for obtaining a more tractable reformulation. Note that we specify the exact choice of $\Phi_a(\cdot)$ later in Section 6.

For the sake of self-containment we state Theorem 10 of [39] adapted to our setting in the following. To this end, we first need some more notation that is taken from [39] as well. By $M \in \mathbb{R}^{V \times A}$ we denote the node-arc incidence matrix of the network G, i.e., for any node $u \in V$ and arc $a \in A$, the entry m_{ua} is defined by

$$m_{ua} = \begin{cases} +1, & \text{if } a = (u, v), \\ -1, & \text{if } a = (v, u), \\ 0, & \text{otherwise.} \end{cases}$$

By choosing a reference node $0 \in V$ and a spanning tree T of G, the arcs can be decomposed into basis arcs $B := A(T)$ and nonbasis arcs $N := A \setminus A(T)$. As a result, by reordering the arcs of the graph, one obtains the representation

$$M = \begin{bmatrix} m_{0B} & m_{0N} \\ M_B & M_N \end{bmatrix}$$

of the incidence matrix M, where $m_0 = (m_{0B}\ m_{0N}) \in \mathbb{R}^{\{0\}\times B}$ denotes the row vector corresponding to the reference node 0. In [39], it is shown that the submatrix M_B is invertible.

We also introduce a vector of nominations $q^n \in \mathbb{R}^V_{\geq 0}$. We thereby comply with the notation used in [39] and denote nominations at exit nodes $u \in V_-$ with a negative sign. This means, we have

$$q^n_u = q^{\text{nom}}_u \quad \text{for all } u \in V_+, \quad q^n_u = -q^{\text{nom}}_u \quad \text{for all } u \in V_-, \quad \text{and} \quad q^n_u = 0 \quad \text{for all } u \in V_0.$$

Further, $\hat{q}^n_0 \in \mathbb{R}^{V-1}$ denotes the vector of nominations without the component corresponding to the reference node 0. In addition, the function $g : \mathbb{R}^V \times \mathbb{R}^N \to \mathbb{R}^V$ with

$$g(q^n, q_N) := \begin{pmatrix} 0 \\ M_B^{-\top}\Phi_B(M_B^{-1}(\hat{q}^n_0 - M_N q_N)) \end{pmatrix} \tag{22}$$

represents the potential, i.e., squared pressure, loss caused by gas flow from the reference node 0 to any other node in the network. We point out that this potential change depends on the nomination q^n as well as on the nonbasis flows q_N. We further use the notation Φ_B and Φ_N that denote vector-valued variants of Φ_a.

Theorem 10 of [39] gives a characterization of feasible bookings. A booking is considered feasible if every balanced and node-wise aggregated nomination that fulfills the booking bounds is transportable. The robustness constraint (10d) in turn requires every balanced and node-wise aggregated nomination that satisfies given technical capacities to be transportable. In this light, we can interpret the technical capacities q^{TC} as a "booking" and use Theorem 10 of [39] in our setting.

Theorem 1 (Labbé et al. [39]). *Let $G = (V, A)$ be a network with given potential bounds $0 < \pi^-_u \leq \pi^+_u \leq \infty$ for every node $u \in V$ and arc capacities $-\infty \leq q^-_a \leq q^+_a \leq \infty$ for every arc $a \in A$. Then, technical capacities q^{TC} are feasible with respect to the robustness constraint (10d) if and only if*

$$\Delta g^*_{w_1,w_2} \leq \pi^+_{w_1} - \pi^-_{w_2} \qquad \text{for all } w_1, w_2 \in V, \tag{23a}$$

$$q^-_a \leq \underline{q}_{B_a} \leq \overline{q}_{B_a} \leq q^+_a \qquad \text{for all } a \in B, \tag{23b}$$

$$q^-_a \leq \underline{q}_{N_a} \leq \overline{q}_{N_a} \leq q^+_a \qquad \text{for all } a \in N, \tag{23c}$$

with

$$\Delta g^*_{w_1,w_2} := \max_{q^n, q_N} \quad g_{w_1}(q^n, q_N) - g_{w_2}(q^n, q_N) \tag{24a}$$

$$\text{s.t.} \quad \Phi_N(q_N) = M_N^\top M_B^{-\top}\Phi_B(M_B^{-1}(\hat{q}^n_0 - M_N q_N)), \qquad [\alpha] \tag{24b}$$

$$q^n_u \in [0, q^{\text{TC}}_u] \quad \text{for all } u \in V_+, \qquad [\beta^\pm_u] \tag{24c}$$

$$q^n_u \in [-q^{\text{TC}}_u, 0] \quad \text{for all } u \in V_-, \qquad [\beta^\pm_u] \tag{24d}$$

$$q^n_u = 0 \quad \text{for all } u \in V_0, \qquad [\beta_u] \tag{24e}$$

$$\sum_{u\in V} q^n_u = 0, \qquad [\gamma] \tag{24f}$$

and

$$\overline{q}_{B_a} := \max_{q^n, q_N} \quad (M_B^{-1}(\hat{q}_0^n - M_N q_N))_a \quad \text{s.t.} \quad \text{(24b)–(24f)}, \tag{25}$$

$$\underline{q}_{B_a} := \min_{q^n, q_N} \quad (M_B^{-1}(\hat{q}_0^n - M_N q_N))_a \quad \text{s.t.} \quad \text{(24b)–(24f)}, \tag{26}$$

$$\overline{q}_{N_a} := \max_{q^n, q_N} \quad q_a \quad \text{s.t.} \quad \text{(24b)–(24f)}, \tag{27}$$

$$\underline{q}_{N_a} := \min_{q^n, q_N} \quad q_a \quad \text{s.t.} \quad \text{(24b)–(24f)} \tag{28}$$

holds.

All optimization problems (24)–(28) are linear under Assumption 1. In total, we have $|V|^2$-many LPs (24), $|B|$-many LPs (25) and (26) and $|N|$-many LPs (27) and (28). Thus, we can replace Constraint (10d) with a total of $|V|^2 + 2|A|$ LPs. These optimization problems can be reformulated using their KKT conditions, which are necessary and sufficient for linear problems. In Problem (24), we denoted the dual variables in brackets. Of course, since we need the KKT conditions of each of the $|V|^2 + 2|A|$ LPs, we need a distinguished set of dual variables per LP.

First, we denote the KKT conditions of Problem (24). We therefore assume without loss of generality that the reference node 0 is an entry, i.e., $0 \in V_+$. Here and in what follows, we also substitute the dual variables $\beta_u = \beta_u^+ - \beta_u^-$ for $u \in V_0$ with $\beta_u^\pm \geq 0$ for the ease of presentation. For given $(w_1, w_2) \in V \times V$, the KKT stationarity conditions read

$$\begin{aligned} \beta_0^- - \beta_0^+ &= \gamma, \\ (A)_{w_1,u} - (A)_{w_2,u} + (AM_N\alpha)_u + \beta_u^- - \beta_u^+ &= \gamma \quad \text{for all } u \in V \setminus \{0\}, \\ -(AM_N)_{w_1}^\top + (AM_N)_{w_2}^\top - (D_N^\Phi + M_N^\top A M_N)\alpha &= 0, \end{aligned} \tag{29}$$

where the matrix A is given by

$$A := M_B^{-\top} D_B^\Phi M_B^{-1}$$

and D_B^Φ and D_N^Φ denote the diagonal Jacobian matrices of the vector-valued functions Φ_B and Φ_N. Further, we have KKT primal feasibility (24b)–(24f), nonnegativity

$$\beta_u^\pm \geq 0 \quad \text{for all } u \in V, \tag{30}$$

and KKT complementarity conditions

$$\begin{aligned} \beta_u^- q_u^n = 0, \quad \beta_u^+(q_u^{\text{TC}} - q_u^n) = 0 \quad \text{for all } u \in V_+, \\ \beta_u^-(-q_u^n) = 0, \quad \beta_u^+(q_u^n + q_u^{\text{TC}}) = 0 \quad \text{for all } u \in V_-. \end{aligned} \tag{31}$$

Thus, for every pair of nodes (w_1, w_2), we can replace Problem (24) by its KKT conditions (24b)–(24f) and (29)–(31).

All other LPs (25)–(28) only differ from Problem (24) in their objective function. Thus, only the KKT stationarity conditions differ from (29), whereas primal feasibility, nonnegativity, and complementarity are given by (24b)–(24f), (30), and (31) (with a separate set of primal and dual variables per LP). Hence, for the remaining LPs, we only specify the KKT stationarity conditions.

For an arc $a \in B$, the corresponding KKT stationarity conditions of Problem (25) are given by

$$\begin{aligned} \beta_0^- - \beta_0^+ - \gamma &= 0, \\ M_{B_{a,u}}^{-1} + (AM_N\alpha)_u + \beta_u^- - \beta_u^+ - \gamma &= 0 \quad \text{for all } u \in V \setminus \{0\}, \\ -(M_B^{-1}M_N)_a^\top - (D_N^\Phi + M_N^\top A M_N)\alpha &= 0, \end{aligned} \tag{32}$$

and the KKT stationarity conditions of Problem (26) are given by

$$\begin{aligned} \beta_0^- - \beta_0^+ - \gamma &= 0, \\ -M_{B_{a,u}}^{-1} + (AM_N\alpha)_u + \beta_u^- - \beta_u^+ - \gamma &= 0 \quad \text{for all } u \in V \setminus \{0\}, \\ (M_B^{-1}M_N)_a^\top - (D_N^\Phi + M_N^\top A M_N)\alpha &= 0. \end{aligned} \tag{33}$$

Finally, for every $a \in N$, the KKT stationarity conditions of Problem (27) are given by

$$\begin{aligned} \beta_0^- - \beta_0^+ - \gamma &= 0, \\ (AM_N\alpha)_u + \beta_u^- - \beta_u^+ - \gamma &= 0 \quad \text{for all } u \in V \setminus \{0\}, \\ e_{N_a} - (D_N^\Phi + M_N^\top A M_N)\alpha &= 0, \end{aligned} \tag{34}$$

and for Problem (28) by

$$\begin{aligned} \beta_0^- - \beta_0^+ - \gamma &= 0, \\ (AM_N\alpha)_u + \beta_u^- - \beta_u^+ - \gamma &= 0 \quad \text{for all } u \in V \setminus \{0\}, \\ -e_{N_a} - (D_N^\Phi + M_N^\top A M_N)\alpha &= 0, \end{aligned} \tag{35}$$

with $e_{N_a} \in \mathbb{R}^N$ being the N_ath unit vector in $\mathbb{R}^N$. In summary, Constraint (10d) is fulfilled if and only if the following holds:

$$\begin{aligned} &\Delta g^*_{w_1,w_2} \le \pi^+_{w_1} - \pi^-_{w_2} \quad \text{for all } w_1, w_2 \in V, \\ &q_a^- \le \underline{q}_{B_a} \le \overline{q}_{B_a} \le q_a^+ \quad \text{for all } a \in B, \\ &q_a^- \le \underline{q}_{N_a} \le \overline{q}_{N_a} \le q_a^+ \quad \text{for all } a \in N, \\ &\text{Constraints (24b)–(24f) and (29), (30), (31)} \quad \text{for all } w_1, w_2 \in V, \\ &\text{Constraints (24b)–(24f) and (32), (30), (31)} \quad \text{for all } a \in B, \\ &\text{Constraints (24b)–(24f) and (33), (30), (31)} \quad \text{for all } a \in B, \\ &\text{Constraints (24b)–(24f) and (34), (30), (31)} \quad \text{for all } a \in N, \\ &\text{Constraints (24b)–(24f) and (35), (30), (31)} \quad \text{for all } a \in N. \end{aligned} \tag{36}$$

We emphasize again that the system (36) contains distinguished variables q^n, q_N, α, $\beta^\pm$, and γ for every involved LP. We further point out that all KKT complementarity conditions (31) occurring in (36) can be linearized again with a big-M linearization as specified in (19). For a provably correct reformulation, we need again primal and dual bounds M_p and M_d. However, in the LPs (24)–(28) neither technical bounds, like flow bounds or squared pressure bounds, nor economic quantities are involved. Furthermore, apart from $\beta_u^\pm$, all other dual variables are free variables. This renders finding provable bounds for $\beta_u^\pm$ way more difficult; the specific numerical values are given in the next section.

Finally, under Assumptions 1–4 and after all reformulations discussed in this and the previous sections, we obtained a mixed-integer single-level problem with a convex-quadratic objective function over linear and bilinear constraints. Although this is still a very challenging problem class, it can be tackled by modern solvers such as, e.g., Gurobi. If the big-M constants used to reformulate the robustness constraint are chosen large enough, then all reformulations are exact. We are now ready to use the derived single-level problem to measure inefficiencies that occur in entry-exit-like gas markets.

6. The Cost of Decoupling Trade and Transport

As mentioned throughout the paper it is very likely that the European entry-exit gas market system leads to inefficiencies due to the decoupling of trade and transport. In particular, it is reasonable to expect that the robustness constraint (10d) is "expensive" in this sense. In this section, we measure this cost of decoupling trade and transport for an exemplary network. For a better understanding of the effects, we analyze inefficiencies

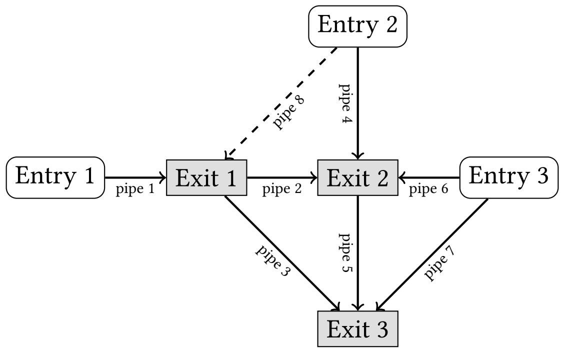

FIGURE 3. The 6-node network.

arising (i) for different demand levels, (ii) in dependence of network expansion, and (iii) for different pricing regimes for the booking price floors $\underline{\pi}_u^{\text{book}}$. To this end, we compare the welfare outcomes of the entry-exit system (10) with the two benchmark models stated in Section 2.6. In the following, we label the entry-exit model by *Entry-Exit*, the single-level benchmark model (12) by *1st-Best* and the bilevel benchmark model (13), which disregards the robustness constraint, by *2nd-Best*. For all our computations, we use the respective single-level reformulations of Problem (10) and (13) that apply all reformulations derived in Sections 3–5.

6.1. **Data and Economic Setup.** For the economic analysis, we consider the 6-node network depicted in Figure 3 with four time periods, i.e., $|T| = 4$. The network has 3 entry nodes and 3 exit nodes with one player per node, no inner nodes, and 7 arcs (without the dashed arc). We do not consider strategic interaction in our analysis. The reasons are manifold. Most importantly, a rigorous analysis of strategic interaction within this already highly complex setup would impose challenges that are far beyond tractability. Moreover, the analysis of non-strategic agents at the nodes allows us to clearly pin down the effects of the market design on efficiency that are present already without the presence of market power. To better isolate the effects that are caused by the decoupling of trade and transport, we simplify the physical data as much as possible w.r.t. maintaining enough granularity to illustrate welfare effects. We assume identical pipes $a \in A$ that are characterized by their length $L_a = 350\,\text{km}$, diameter $D_a = 0.5\,\text{m}$, roughness $k_a = 0.1\,\text{mm}$, and capacities $q_a^{\pm} = \pm 435\,\text{kg}\,\text{s}^{-1}$. For deriving a linear transport model, see Assumption 1, we consider stationary gas flow, i.e., we abstract from temporal dependencies. Further, we only consider horizontal pipes. In such a setting, one can use the following well-known pressure loss function as given in, e.g., [20].

$$\Phi(q_a) = \frac{\lambda_a c^2 L_a}{(0.25\pi)^2 D_a^5} |q_a| q_a. \tag{37}$$

The constant $c = 340\,\text{m}\,\text{s}^{-1}$ denotes the speed of sound in natural gas. The friction coefficient λ_a can be approximated by the formula of Nikuradse; see [20]:

$$\lambda_a = \left(2 \log_{10}\left(\frac{D_a}{k_a}\right) + 1.138\right)^{-2}.$$

Equation (37) is a suitable simplification of gas flow physics but is still nonlinear. Thus, to arrive at a linear approximation, we replace $|q_a|$ by a mean flow q_a^{mean}. Preliminary computational studies showed that $q_a^{\text{mean}} = 100\,\text{kg}\,\text{s}^{-1}$ is a reasonable choice. We are aware that this linearization yields only a coarse approximation of (37) but, as already discussed, without Assumption 1 the problem is computationally not tractable. In case the chosen mean flow underestimates the actual flow, the linearization underestimates

TABLE 1. Variable costs c_i^{var} (in EUR/(1000 Nm^3/h)) of entries and intercepts $a_{i,t}$ (in EUR/(1000 Nm^3/h)) and slopes b_i (in EUR/(1000 Nm^3/h)^2) of exits.

	Entries	Exits				
	c_i^{var}	$a_{i,0}$	$a_{i,1}$	$a_{i,2}$	$a_{i,3}$	b_i
	EUR/(1000 Nm^3/h)	EUR/(1000 Nm^3/h)				EUR/(1000 Nm^3/h)^2
1	63	1000	450	850	1700	-4.5
2	57	800	300	700	1500	-5.0
3	71	3900	2500	3700	5500	-20.0

the pressure loss, which in turn may result in pressure bound violations w.r.t. the more accurate nonlinear flow model in (37). The situation may be improved by fitting individual average flows for each arc via a trial-and-error process. On the other hand, the flow varies over the time periods on each arc. Thus, choosing the average flow as the mean flow over the time periods does not prevent the existence of time periods, in which the average flow underestimates the actual flow. Finally, making the average flow time-period dependent is not feasible w.r.t. the characterization of the robustness constraint as discussed in Section 5. Consequently, and as expected, a linearization of (37) will be inexact for certain flow situations. Thus, we use the average flows as given above and check for the resulting violations ex-post to get an idea of the level of (in)accuracy of the chosen linearization; see Appendix A for the detailed results of this ex-post check. Let us be as transparent as possible at this point: Of course, we observe physical errors that are not negligible. However, the used modeling approach is the best possible due to the already discussed theoretical complexity reasons. Nevertheless, we are convinced that our results shed light on the economic implications of the entry-exit gas market system in Europe.

For all entry nodes $u \in V_+$, we have pressure bounds $p_u^- = 40$ bar and $p_u^+ = 65$ bar, and for all exit nodes $u \in V_-$, we have pressure bounds $p_u^- = 40$ bar and $p_u^+ = 50$ bar. The economic data for every gas seller and buyer is specified in Table 1. The four time periods $t \in T$ model the seasons spring ($t = 0$), summer ($t = 1$), autumn ($t = 2$), and winter ($t = 3$). Thus, the willingness to pay $a_{i,t}$ is lowest in time period $t = 1$ and highest in time period $t = 3$. We can also see from Table 1 that Entry 2 is the cheapest gas producer with the lowest variable costs, while Entry 3 is the most expensive one. Exit 2 has the lowest willingness to pay but a larger absolute slope than Exit 1. Exit 3 has the highest willingness to pay and is very inelastic with the largest absolute slope. Finally, we assume no exogenous network costs, i.e., we set $C = 0$.

6.2. **Computational Setup.** We now turn to the computational setup. All optimization problems used for the economic analysis have been implemented in Python 3 and solved with Gurobi 9.0.1. All computations have been carried out on a compute cluster; see [48] for the details about the installed hardware. The big-M constants discussed at the end of Section 5 are chosen to be 1×10^6. We now discuss some statistics of the three different models *1st-Best*, *2nd-Best*, and *Entry-Exit* as displayed in Table 2. It reveals that the *1st-Best* model is not challenging at all, which is expected since it is a very small continuous convex-quadratic problem. The model has 104 continuous variables and 108 constraints. Note that for the linearization of the absolute pressure losses in the objective function of the *1st-Best* model, only the two left inequalities in (21) and no additional binary variables are required. In the *2nd-Best*, the absolute pressure losses are coupled to bilinear terms; see Section 4. Thus, the full set of the constraints (21) and $|T| \cdot |A| = 28$ binary variables are required. More significantly, the linearization of the bilevel structure of the *2nd-Best* model adds $|V| + 2|T| \cdot |V| = 54$ binary variables. Still, the *2nd-Best* model is not challenging. In

TABLE 2. Model statistics of *1st-Best*, *2nd-Best*, and *Entry-Exit* for a 6-node and a 9-node network.

	Variables	Binaries	Constraints	Time in s
6-node Network				
1st-Best	104	0	108	0.12
2nd-Best	262	82	336	0.42
Entry-Exit	1980	682	2983	43.72
9-node Network				
1st-Best	176	0	192	0.05
2nd-Best	421	133	552	1.17
Entry-Exit	6342	2059	9679	5003.07

contrast, introducing the robustness constraint adds noticeable difficulty to the problem both in terms of problem size and running time. The total number of variables, the number of binary variables, and the number of constraints all roughly increase by a factor 8. In fact, the reformulation of the robustness constraint adds $2|V|(|V|^2 + 2|A|) = 600$ binary variables. As a result, the running time increases by a factor around 100. While the resulting *Entry-Exit* model is still neither big nor challenging (around 43 s running time) for state-of-the-art solvers, this analysis demonstrates that the correct modeling of entry-exit-like systems introduces significant complexity even for small networks like the 6-node network of Figure 3. This becomes even more obvious for bigger networks. Although we stick to the 6-node network for the economic analysis, we analyze the computational scalability of the *Entry-Exit* model. Table 2 also displays model sizes and running times for a 9-node network. Again, the *1st-Best* and *2nd-Best* models are not challenging, while the *Entry-Exit* model requires significantly more variables and constraints. The additional binary variables (around 15 times more compared to *2nd-Best*) are critical, which results in a running time of over 5000 s. This is almost 5000 times the running time needed for the *2nd-Best* model and underlines the hardness that is added by the robustness constraint. We also carried out some numerical experiments on a 12-node network. For this network, the first-best model was solved in 0.04 s and the second-best model was solved in 2.01 s. The entry-exit model, however, was not solved within the time limit of 24 h. In fact, the gap did not improve anymore after six hours and stayed at 120 %. This again underlines the tremendous hardness that is introduced by the robustness condition.

6.3. **The Effect of Scaling Demand.** In this section, we analyze the cost of decoupling trade and transport in dependence of different demand levels. We compare the *standard* demand as specified in Table 1 to a setting with *low* demand, i.e., the intercepts $a_{i,t}$ of the standard setting are decreased by 30 %, and to a setting with *high* demand, i.e., the intercepts of the standard setting are increased by 40 %. As already mentioned, the robustness constraint is suspect to shrink the used capacity in the network because the TSO might be forced to set the technical capacities too conservatively. In Table 3 we display the relative reduction of total technical capacities, bookings, and nominations of the *2nd-Best* model and the *Entry-Exit* model w.r.t. the *1st-Best* model. To that aim, the total "technical capacities" and "bookings" for the *1st-Best* model are computed by $|T| \cdot \sum_{u \in V_+ \cup V_-} \max_{t \in T} \left\{q^{\text{nom}}_{u,t}\right\}$. The table also denotes the relative cost of decoupling trade and transport, which is given by the relative reduction of total welfare W. For the standard demand level, the market system associated with the *2nd-Best* decreases available technical capacities and bookings by 10.24 % and nominations by 5.64 %. This yields a moderate welfare reduction of 1.93 % compared to *1st-Best*. The effect is way more pronounced for the *Entry-Exit* system, in which technical capacities and bookings are decreased by 27.34 % and nominations by 21.55 %. As a result, the welfare decreases by 8.34 % compared to the

TABLE 3. Relative reduction (in %) of total technical capacities (q^{TC}), bookings (q^{book}), nominations (q^{nom}), and welfare (W) of the *2nd-Best* model and the *Entry-Exit* model w.r.t. the *1st-Best* model under low, normal, and high demand. Note that for the *1st-Best* model, we consider $q^{\mathrm{TC}} = q^{\mathrm{book}} = |T| \sum_{u \in V_+ \cup V_-} \max_{t \in T} \left\{q^{\mathrm{nom}}_{u,t}\right\}$.

Reduction in %	2nd-Best				Entry-Exit			
	q^{TC}	q^{book}	q^{nom}	W	q^{TC}	q^{book}	q^{nom}	W
Low	7.16	7.16	1.47	0.85	32.00	32.00	16.33	6.72
Normal	10.24	10.24	5.64	1.93	27.34	27.34	21.55	8.34
High	9.03	9.03	4.00	0.84	17.62	17.62	15.05	8.72

TABLE 4. Relative reduction (in %) of total technical capacities (q^{TC}), bookings (q^{book}), nominations (q^{nom}), and welfare (W) of the *2nd-Best* model and the *Entry-Exit* model w.r.t. the *1st-Best* model for the standard and the expanded network. Note that for the *1st-Best* model, we consider $q^{\mathrm{TC}} = q^{\mathrm{book}} = |T| \sum_{u \in V_+ \cup V_-} \max_{t \in T} \left\{q^{\mathrm{nom}}_{u,t}\right\}$.

Reduction in %	2nd-Best				Entry-Exit			
	q^{TC}	q^{book}	q^{nom}	W	q^{TC}	q^{book}	q^{nom}	W
Standard	10.24	10.24	5.64	1.93	27.34	27.34	21.55	8.34
Expanded	11.39	11.39	3.33	1.50	13.41	13.41	3.85	1.82

1st-Best. When compared to the *2nd-Best*, we find that the robustness constraint alone accounts for more than 6.4 % of welfare loss. In a setting with low demand, these effects are similar although not as pronounced. The welfare loss caused by the *Entry-Exit* model is 6.72 % compared to the *1st-Best* and around 5.9 % compared to the *2nd-Best*. In a setting with high demand, the welfare loss caused by the *2nd-Best* model is only 0.84 %, which is lower compared to the normal demand level. This underlines that, even when physics are modeled in a linear fashion, the welfare effects of different market design choices cannot be expected to be monotonic. In line with the latter statement, the welfare loss caused by the *Entry-Exit* model (8.72 %) is slightly higher compared to the setting with intermediate demand. For high demand, the robustness constraint alone accounts for a welfare loss of almost 7.9 %.

Overall, the effects are quite similar throughout the various demand levels: The *2nd-Best* model causes a noticeable drop in available capacities, but results only in a moderate welfare loss compared to the *1st-Best* model. In contrast, the *Entry-Exit* model causes a significantly higher drop in available capacities due to the robustness constraint. This results in severe inefficiencies, i.e., in a high cost of decoupling trade and transport.

6.4. **The Effect of Different Network Configurations.** We now analyze the cost of decoupling trade and transport in different network configurations. In particular, we compare the *standard* network of Figure 3 without the dashed pipe 8 with an *expanded* network that includes pipe 8. Note that this improves the connection of the seller at Entry 2, who is the seller with the overall lowest cost, with all other agents. We compare again the relative reduction of total technical capacities, bookings, and nominations of the *2nd-Best* model and the *Entry-Exit* model w.r.t. the *1st-Best* model as well as the relative reduction of total welfare W; see Table 4. For the standard network, the welfare loss caused by the entry-exit system has already been discussed in Section 6.3. The observations differ systematically for the expanded network. While the relative reduction in capacities and

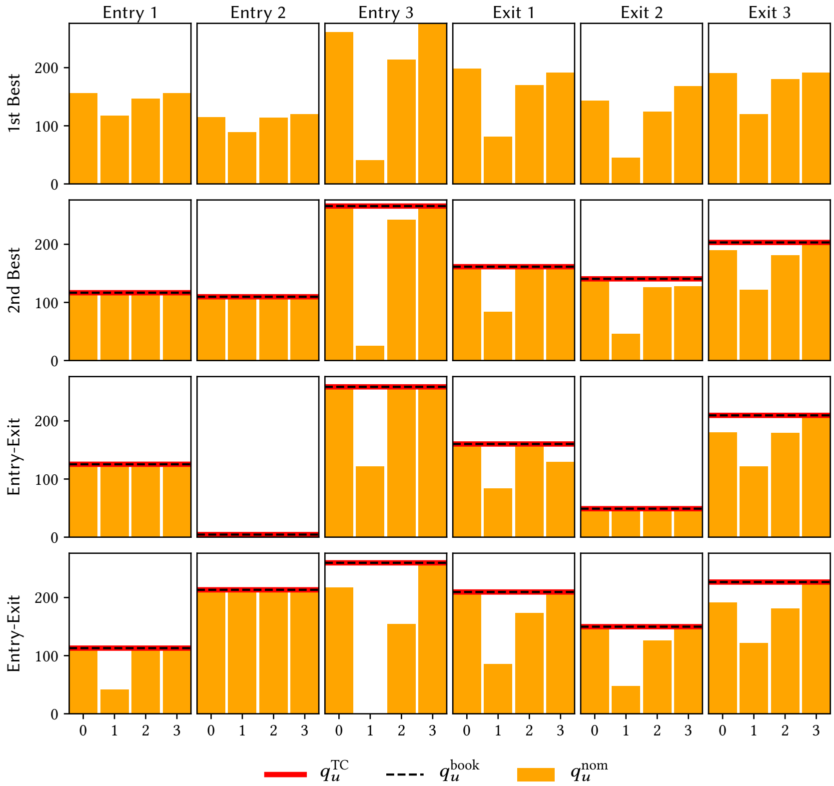

FIGURE 4. Technical capacities (q_u^{TC}), bookings (q_i^{book}), and nominations ($q_{i,t}^{\mathrm{nom}}$) for the *1st-Best*, *2nd-Best*, and *Entry-Exit* models of the standard network and for the *Entry-Exit* model of the expanded network.

the relative welfare loss caused by the *2nd-Best* model are quite comparable to the standard network, the *Entry-Exit* model hardly adds further inefficiencies for the expanded network. The relative decrease in technical capacities, bookings, and nominations is only slightly more pronounced compared to *2nd-Best*, which results in a moderate relative reduction of total welfare W by 1.82 % compared to the *1st-Best*. This is only 0.32 % more than what is observed for the *2nd-Best*. Thus, the additional loss of relative welfare caused by the robustness constraint alone is almost negligible.

To shed some more light on this observation, we look at Figure 4, which displays technical capacities, bookings, and nominations per time period for *1st-Best*, *2nd-Best*, and *Entry-Exit* of the standard network (upper three rows) and for *Entry-Exit* of the expanded network (bottom row). Obviously, in the *1st-Best* model, it is welfare-optimal to include the cheapest gas seller Entry 2 into the market. The nominations in the *2nd-Best* are a bit more leveled but quite comparable to the outcomes of the *1st-Best*. However, in the *Entry-Exit* model, the TSO sets very low technical capacities for Entry 2, which almost entirely excludes the seller from the market. One explanation for this is that Entry 2 is connected only poorly to the network, especially to Exit 3. We recap that Exit 3 is the buyer with the most inelastic demand and the highest willingness to pay. It thus turns out to be more beneficial with regard to the robustness constraint and the involved physics to allocate high technical capacities to the more expensive Entry 3, that has a direct link to Exit 3. The situation changes immediately when pipe 8 is introduced. Entry 2 is now connected better to the network and can supply Exits 1 and 2 directly. It is now beneficial to include Entry 2 into the trade by increasing its technical capacities; see Figure 4 (bottom row).

The discussion in this section indicates that the network configuration strongly impacts the market outcomes of the individual players. In this light, trade and transport are not decoupled after all, which is most obvious in the setting of the standard network, where it is welfare-optimal to exclude the cheapest producer from the market. The results of this section also indicate that in more expanded networks, the cost of decoupling trade and transport is rather small. This low cost however comes at the expense of possibly high investment costs for building and maintaining an (over-)expanded network.

Finally note that, in this section, we just compared the efficiency losses from an *Entry-Exit* model as compared to the *1st-Best* and *2nd-Best* for different network configurations. The network design problem that would have to be solved in order to set up the network optimally cannot be addressed in this paper. Obviously, one would have to account for the costs and benefits of adding pipes, accounting for all endogenous feedback effects of network expansion on the market outcome.

6.5. **The Effect of Different Pricing Regimes.** In the last two sections, we have seen that entry-exit-like systems can cause severe reductions in total welfare W compared to the *1st-Best* and *2nd-Best* models. This cost of decoupling trade and transport is "payed" by the gas buying and selling firms by decreasing rents. However, in general, this cost will not be shared equally but some players may be discriminated by the market system, while others may even benefit from the decoupling of trade and transport.

In this section, we analyze how the individual rents

$$R_i = \sum_{t \in T} (\pi_t^{\text{nom}} - c_i^{\text{var}}) q_{i,t}^{\text{nom}} - \underline{\pi}_u^{\text{book}} q_i^{\text{book}}, \quad i \in \mathcal{P}_u, \ u \in V_+,$$
$$R_i = \sum_{t \in T} (P_{i,t}(q_{i,t}^{\text{nom}}) - \pi_t^{\text{nom}}) q_{i,t}^{\text{nom}} - \underline{\pi}_u^{\text{book}} q_i^{\text{book}}, \quad i \in \mathcal{P}_u, \ u \in V_-$$

are affected by entry-exit-like systems. The two equations above represent the buyers' and the sellers' rents, respectively. Note that the price floor $\underline{\pi}_u^{\text{book}}$, that is set by the TSO to recover transportation and exogenous network costs (cf. Constraint (1c)), directly impacts the individual rents of the players. The price floors can be chosen by the TSO in various ways. In this section, we analyze four different such pricing regimes: (i) a regime in which price floors are chosen *efficiently*, (ii) a regime with a single *uniform* booking price floor at all entry and exit nodes, (iii) a regime in which *only exit* nodes pay a uniform booking price floor, and (iv) a regime in which *only entry* nodes pay a uniform booking price floor. Note that (i) is assumed implicitly for our modeling in Section 2 and that the other three regimes can be modeled by additional linear constraints. To better illustrate the effects of the various pricing schemes we use exogenously given network costs of $C = 300\,000$ EUR throughout this section. This high value of the investment costs, which must be earned via the price floors (cf. Constraint (1c)), allows to illustrate the effect of the price floors on efficiency very clearly. Since the comparison between the different market designs in the previous sections is not enriched by this assumption about investment costs, we have not specified the costs there. Note further that we use the standard network without the dashed pipe 8, see Figure 3, in this subsection.

Table 5 displays the reduction in total welfare W as compared to the *1st-Best* model for each of the four different pricing regimes and shows how the rents R_i of all gas sellers and buyers are affected. All values represent absolute reductions in comparison to the welfare and rents achieved in the *1st-Best* model, which is the same for all four pricing regimes. In addition, Figure 5 shows the welfare-maximizing price floors $\underline{\pi}_u^{\text{book}}$ and payments $\underline{\pi}_u^{\text{book}} q_i^{\text{book}}$ for the four pricing schemes. At a first glance, the different pricing regimes yield diverse price floor patterns, see Figure 5, which is not surprising. Also, the total welfare is affected by the choice of the pricing scheme; see Table 5. Of course, the *efficient* pricing scheme yields the lowest reduction of welfare compared to the *1st-Best* because it shifts payments to agents where they do not induce welfare losses. The same welfare result

TABLE 5. Absolute reduction (in 1000 EUR) of rents R_i and total welfare W compared to the *1st-Best* model for different pricing regimes.

Reduction in 1000 EUR	Entry 1	Entry 2	Entry 3	Exit 1	Exit 2	Exit 3	W
Efficient	41.87	41.87	−181.11	122.48	148.41	7.35	180.86
Uniform	−36.28	38.94	−113.44	143.75	153.33	0.66	186.96
Only Exit	−35.43	39.02	−109.22	142.17	154.53	−1.74	189.33
Only Entry	−19.62	39.58	−76.63	122.48	148.41	−33.35	180.86

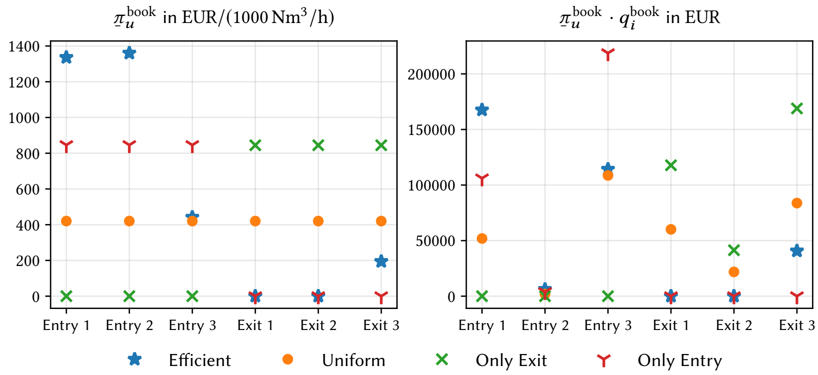

FIGURE 5. Comparison of price floors (left) and products of price floors times bookings (right) for the pricing regimes *efficient*, *uniform*, *only exit*, and *only entry*.

can be obtained by charging price floors for entries only, since obviously those payments do not induce any effect on the quantity supplied at the market. As long as revenues of gas sellers are high enough, the TSO can collect his expenses for gas transportation and other exogenous network costs from the sellers. The fact that Exit 3 receives a price floor of around 200 EUR/(1000Nm^3/h) in the *efficient* pricing scheme compared to a price floor of 0 EUR/(1000Nm^3/h) in the *only entry* pricing scheme is an artifact of non-unique welfare-maximizing price floors. In particular, the *only entry* pricing scheme, which charges the same price floor at every entry node, is also a feasible efficient pricing scheme. It can also be seen in Table 5 that imposing price floors for exit players like it is done in the pricing regimes *uniform* and *only exit* results in additional welfare losses. This is due to the fact that the players' marginal units are driven out of the market at Exit 1 and 2 if a booking fee is charged. We now take a closer look at the implications for the individual players.

From Table 5 it can be seen that Entry 2 is not affected much by changes in the pricing scheme. The rent of Entry 2 is reduced by around 42 000 EUR compared to the *1st-Best* under the *efficient* pricing scheme. This reduction decreases slightly, when other pricing schemes are used. As seen in the previous sections, the reason is the following. Since Entry 2 is the cheapest seller, it is assigned as much nomination q_i^{nom} as possible with respect to physics in the *1st-Best* model; cf. Figure 4. However, due to its poor network access in combination with the robustness constraint of the entry-exit system, Entry 2 is excluded from the market by having its technical capacities set to zero by the TSO. Thus, the high price floor that Entry 2 receives is irrelevant and the contribution $\pi_u^{\text{book}} q_i^{\text{book}}$ of Entry 2 to the reimbursement of the TSO for transportation and network expenses is marginal; see Figure 5 (right). A direct consequence is that the rent of Entry 3 increases compared to the *1st-Best* throughout all pricing schemes. The amount that Entry 1 sells in *1st-Best* but not in *Entry-Exit* is compensated by Entry 3. In the *efficient* pricing scheme, this results in an increase of the rent of Entry 3 of 181 110 EUR compared to the *1st-Best*, which is significant, taking into account that the reduction of total welfare amounts to 180 860 EUR. In the *efficient* pricing scheme, the increasing rent of Entry 3 is in parts at the expense of Entry 1, whose rent is reduced by around 42 000 EUR compared to the first best. This reduction is also caused by a very high price floor that is set by the TSO. It is interesting to see that in the *efficient* pricing scheme, the highest contribution to cover the expenses of the TSO comes from Entry 1; see Figure 5 (right). The reason is that at Entry 1 the expenses of the TSO can be recovered with the least distorting effects. Note also that Entry 2 seems to be charged a high fee but in effect pays nothing since the entry receives no bookable capacity. Using any other pricing scheme than *efficient* immediately yields an increasing rent of Entry 1, since the burden is shifted also to other agents. However, this happens at a cost, in case trade is suppressed.

Similar observations can be drawn when looking at the exit players. Exits 1 and 2 do not receive a price floor in the *efficient* and (obviously) in the *only entry* pricing regime. Thus, these two regimes yield the lowest rent reduction for the two exits. Still, the reduction is, compared to the *1st-Best* and the reduction of total welfare W, very significant. This is not the case for Exit 3 in the *efficient* pricing regime. Since the demand of Exit 3 is very inelastic, the reduction of the rent compared to the *1st-Best* is rather low. The rent even increases compared to the *1st-Best*, when only the entries are charged with a price floor. It is also noteworthy that the *uniform* and *only exit* pricing regimes increase the rent of Exit 3 compared to the *efficient* regime, although the former two regimes yield significantly higher booking price floors and booking payments for Exit 3; see Figure 5. The reason is that Exit 3 can realize higher nominations, which overcompensates the increasing payments.

From the analysis in this section, we can draw several conclusions. The actual choice of the pricing regime for booking price floors can affect both, the total welfare as well as the individual rents, and can add to the cost of decoupling trade and transport. The *efficient* booking price floor regime yields the best overall welfare outcome but is highly discriminating in the sense that the rent of Entry 3 increases drastically at the expense of all other players. The regime *only entry* yields the same welfare outcome but distributes the reductions of the rents way more evenly. This makes also clear that in all those considerations one has to account for the fact that the *Entry-Exit* model might have multiple solutions.

7. Conclusion

In this paper, we analyzed inefficiencies that arise in entry-exit gas markets due to the decoupling of trade and transport. To this end, we used an existing multilevel model of the European entry-exit system and an equivalent bilevel reformulation thereof from the literature [24]. We reformulated this bilevel problem as an equivalent nonconvex and nonsmooth mixed-integer single-level problem. The major difficulty of the single-level problem is an entry-exit-market specific robustness constraint that directly results from a proper modeling of the underlying gas market system: The TSO needs to ensure that every balanced nomination that is feasible w.r.t. technical capacities is transportable. This constraint alone renders the problem intractable in general. However, under the assumption of passive networks (i.e., without active elements such as compressor stations), and a potential-based but linear flow model, we derived a "more tractable" single-level reformulation by applying a characterization of the feasible points of the robustness constraint from [39].

In a case study for a stylized network, we analyzed the inefficiencies caused by entry-exit systems and identified four major effects. First, the robustness constraint accounts for significant welfare losses compared to first-best and second-best models. This means that—under the assumptions made—the cost of decoupling trade and transport can indeed be measured by the model and techniques that we propose. Second, the design of the network has a significant impact on which gas traders receive technical capacities from the TSO. Thus, our results show that, in fact, trade and transport are not decoupled in a long run. Third, the booking price floors established by the TSO to collect payments from the traders to reimburse network costs may be ambiguous and, depending on the pricing regime used, can be quite discriminatory. Fourth and finally, our computational analysis revealed that the computational effort required to solve the problem mostly stems from the robustness constraint.

In practice, the robustness constraint is likely not imposed in its strictest form. To capture a situation in which the network operator sets the technical capacities based on experience, we also analyzed a second-best entry-exit model, where only feasibility of the actually resulting market outcomes is required. As one would expect, we find that for this model, the welfare losses are lower than under the strict robustness constraint. Less restrictive feasibility constraints might also be possible in practice if, e.g., interruptible contracts are used. Although this case can possibly be approximated by our second-best entry-exit scenario, it is too complex to be modeled and solved in the context of our formal analysis and case study. Thus, while our contribution for the first time provides a formal model to asses the efficiency losses that may result from the decoupling of network and market, it naturally cannot address various other aspects of gas markets that are of practical relevance. In addition to the interruptible contracts mentioned above, these include, e.g., aspects such as market liquidity, security of supply, as well as the implementation cost of an alternative market mechanism itself.

In order to extend the findings in this paper to instances of real-world size, several challenges arise. Most importantly, it is not clear how to handle the robustness constraint for nonlinear flow models on general (cyclic) networks. We discussed theoretical complexity reasons why the consideration of nonlinear flow models on general networks is out of reach right now. In our opinion, this legitimates the usage of linear flow models. Note that this is in analogy to very many electricity market models with DC power flow constraints. These models are well-accepted in the literature as well although it is known that there are instances, in which no DC-feasible points exist that are feasible for the nonlinear AC model [2]. Nevertheless, the development of techniques that allow to better capture the nonlinear aspects of gas flow is a very reasonable field for future research. Furthermore, adding active elements would add another layer of complexity to an already very challenging model. The same holds true for transient gas flow models that take time-dependent aspects of gas flow such as linepack into a account. Although this increases the complexity of the modeling, results with these additional aspects will provide even better insights into the European gas market system. In addition, a more effective solution approach tailored to bilevel problems with a structure as specified in Section 3 is required. However, this is out of scope of this paper and will be part of our future research.

Acknowledgments

This research has been performed as part of the Energie Campus Nürnberg and is supported by funding of the Bavarian State Government. We also thank the Deutsche Forschungsgemeinschaft for their support within project A05 and B08 in the "Sonderforschungsbereich/Transregio 154 Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks". Finally, the third author thanks Johannes Thürauf for several fruitful discussions on the topic of this paper.

Appendix A. Appendix: Feasibility Check for a Nonlinear Flow Model

For the sake of completeness, we discuss possible infeasibilities resulting from the linear flow model as specified in Section 6.1. For our ex-post analysis, we minimize the

TABLE 6. Results of the nonlinear flow feasibility check.

Instance	Model	# Viol.	Total Viol.	Max. Viol.	At Node
6 Nodes	1st-Best	10	210.5	34.1	Entry 1
6 Nodes	2nd-Best	6	62.7	24.3	Entry 3
6 Nodes	Entry-Exit	11	104.2	17.4	Entry 3
6 Nodes, Low Demand	1st-Best	8	149.7	34.8	Entry 1
6 Nodes, Low Demand	2nd-Best	7	101.2	19.6	Entry 3
6 Nodes, Low Demand	Entry-Exit	2	10.0	6.6	Entry 3
6 Nodes, High Demand	1st-Best	14	249.5	24.7	Entry 1
6 Nodes, High Demand	2nd-Best	11	153.4	20.5	Exit 3
6 Nodes, High Demand	Entry-Exit	6	68.9	15.9	Exit 3
6 Nodes, Extended	1st-Best	6	94.4	20.2	Exit 3
6 Nodes, Extended	2nd-Best	2	36.8	20.7	Entry 3
6 Nodes, Extended	Entry-Exit	2	28.9	16.9	Exit 3
6 Nodes, Network Investment	1st-Best	10	210.5	34.1	Entry 1
6 Nodes, Network Investment	2nd-Best	5	61.2	24.3	Entry 3
6 Nodes, Network Investment	Entry-Exit	11	104.2	17.4	Entry 3
6 Nodes, Uniform Pricefloor	2nd-Best	6	57.5	24.6	Entry 3
6 Nodes, Uniform Pricefloor	Entry-Exit	7	72.8	16.7	Entry 3
6 Nodes, Only Entry Pricefloor	2nd-Best	6	62.9	24.3	Entry 3
6 Nodes, Only Entry Pricefloor	Entry-Exit	11	104.2	17.4	Entry 3
6 Nodes, Only Exit Pricefloor	2nd-Best	6	57.5	24.6	Entry 3
6 Nodes, Only Exit Pricefloor	Entry-Exit	10	82.9	15.5	Entry 3
9 Nodes	1st-Best	15	293.3	34.2	Entry 4
9 Nodes	2nd-Best	4	53.6	23.0	Entry 3
9 Nodes	Entry-Exit	6	62.9	18.2	Entry 3

sum of squared pressure bound violations, i.e., we solve the nonlinear problem

$$\begin{aligned}
\min_{\pi,q,s^-,s^+} \quad & \sum_{t\in T}\sum_{u\in V}\left(s^-_{u,t}+s^+_{u,t}\right)\\
\text{s.t.} \quad & (5)\text{–}(7),\\
& \pi^-_u - s^-_{u,t} \le \pi_{u,t} \le \pi^+_u \le +s^+_{u,t} \quad \text{for all } u\in V,\ t\in T,\\
& s^-_{u,t}, s^+_{u,t} \ge 0 \quad \text{for all } u\in V,\ t\in T,
\end{aligned}$$

in which the nominations $q^{\mathrm{nom}}_{i,t}$ are fixed to the solution provided by our solution approach. As a solver, we used the local solver CONOPT4. We also checked for some exemplary instances with the global solver BARON, if the solution provided by CONOPT4 is a global optimum. For all our tests, this was the case. In Table 6, we summarize the results for all instances that we used throughout this paper. The first two columns specify the instance. The third column lists the number of pressure bound violations (over all nodes in the network and over all time periods), whereas the fourth column lists the aggregated bound violation and the fifth one lists the maximum violation. The last column denotes the node in the network at which the largest violation occurred.

REFERENCES

[1] Alonso, A., Olmos, L., and Serrano, M. "Application of an entry–exit tariff model to the gas transport system in Spain." In: *Energy Policy* 38.5 (2010), pp. 5133–5140. DOI: 10.1016/j.enpol.2010.04.043.

[2] Baker, K. *Solutions of DC OPF are Never AC Feasible.* Tech. rep. 2020. URL: https://arxiv.org/pdf/1912.00319.pdf.

[3] Baltensperger, T., Füchslin, R. M., Krütli, P., and Lygeros, J. "Multiplicity of equilibria in conjectural variations models of natural gas markets." In: *European Journal of Operational Research* 252.2 (2016), pp. 646–656. DOI: 10.1016/j.ejor.2016.01.032.

[4] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. *Robust Optimization*. Princeton University Press, 2009.

[5] Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. "Adjustable robust solutions of uncertain linear programs." In: *Mathematical Programming* 99.2 (2004), pp. 351–376. DOI: 10.1007/s10107-003-0454-y.

[6] Bertsimas, D., Brown, D. B., and Caramanis, C. "Theory and Applications of Robust Optimization." In: *SIAM Review* 53.3 (2011), pp. 464–501. DOI: 10.1137/080734510.

[7] Boots, M. G., Rijkers, F. A., and Hobbs, B. F. "Trading in the Downstream European Gas Market: A Successive Oligopoly Approach." In: *The Energy Journal* 25.3 (2004), pp. 73–102. URL: http://www.jstor.org/stable/41323043.

[8] Boucher, J. and Smeers, Y. "Economic forces in the European gas market — a 1985 prospective." In: *Energy Economics* 9.1 (1987), pp. 2–16. DOI: 10.1016/0140-9883(87)90002-8.

[9] Boucher, J. and Smeers, Y. "Gas trade in the European community during the 1970s." In: *Energy Economics* 7.2 (1985), pp. 102–116. DOI: 10.1016/0140-9883(85)90025-8.

[10] Boyd, S. and Vandenberghe, L. *Convex Optimization*. Cambridge University Press, 2004. DOI: 10.1017/cbo9780511804441.

[11] Chyong, C. K. and Hobbs, B. F. "Strategic Eurasian natural gas market model for energy security and policy analysis: Formulation and application to South Stream." In: *Energy Economics* 44 (2014), pp. 198–211. DOI: 10.1016/j.eneco.2014.04.006.

[12] Cremer, H. and Laffont, J.-J. "Competition in gas markets." In: *European Economic Review* 46.4–5 (2002), pp. 928–935. DOI: 10.1016/S0014-2921(01)00226-4.

[13] Egging, R., Gabriel, S. A., Holz, F., and Zhuang, J. "A complementarity model for the European natural gas market." In: *Energy Policy* 36.7 (2008), pp. 2385–2414. DOI: 10.1016/j.enpol.2008.01.044.

[14] Egging, R., Holz, F., and Gabriel, S. A. "The World Gas Model: A multi-period mixed complementarity model for the global natural gas market." In: *Energy* 35.10 (2010), pp. 4016–4029. DOI: 10.1016/j.energy.2010.03.053.

[15] European Parliament and Council of the European Union. *Directive 2003/55/EC of the European Parliament and of the Council of 26 June 2003 concerning common rules for the internal market in natural gas and repealing Directive 98/30/EC*. 2003.

[16] European Parliament and Council of the European Union. *Directive 2009/73/EC of the European Parliament and of the Council concerning common rules for the internal market in natural gas and repealing Directive 2003/55/EC*. 2009.

[17] European Parliament and Council of the European Union. *Directive 98/30/EC of the European Parliament and of the Council of 22 June 1998 concerning common rules for the internal market in natural gas*. 1998.

[18] Fischetti, M., Ljubić, I., Monaci, M., and Sinnl, M. "A New General-Purpose Algorithm for Mixed-Integer Bilevel Linear Programs." In: *Operations Research* 65.6 (2017), pp. 1615–1637. DOI: 10.1287/opre.2017.1650.

[19] Fortuny-Amat, J. and McCarl, B. "A Representation and Economic Interpretation of a Two-Level Programming Problem." In: *The Journal of the Operational Research Society* 32.9 (1981), pp. 783–792. JSTOR: 2581394.

[20] Fügenschuh, A., Geißler, B., Gollmer, R., Morsi, A., Pfetsch, M. E., Rövekamp, J., Schmidt, M., Spreckelsen, K., and Steinbach, M. C. "Physical and technical fundamentals of gas networks." In: *Evaluating Gas Network Capacities*. Ed. by T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe. SIAM-MOS series on Optimization. SIAM, 2015. Chap. 2, pp. 17–44. DOI: 10.1137/1.9781611973693.ch2.

[21] Gabriel, S. A., Kiet, S., and Zhuang, J. "A Mixed Complementarity-Based Equilibrium Model of Natural Gas Markets." In: *Operations Research* 53.5 (2005), pp. 799–818. DOI: 10.1287/opre.1040.0199.

[22] Glachant, J.-M., Hallack, M., and Vazquez, M. *Building Competitive Gas Markets in the EU*. Cheltenham: Edward Elgar Publishing, 2013.

[23] Grimm, V., Grübel, J., Schewe, L., Schmidt, M., and Zöttl, G. "Nonconvex equilibrium models for gas market analysis: Failure of standard techniques and alternative modeling approaches." In: *European Journal of Operational Research* 273.3 (2019), pp. 1097–1108. DOI: 10.1016/j.ejor.2018.09.016.

[24] Grimm, V., Schewe, L., Schmidt, M., and Zöttl, G. "A Multilevel Model of the European Entry-Exit Gas Market." In: *Mathematical Methods of Operations Research* 89.2 (2019), pp. 223–255. DOI: 10.1007/s00186-018-0647-z.

[25] Groß, M., Pfetsch, M. E., Schewe, L., Schmidt, M., and Skutella, M. "Algorithmic results for potential-based flows: Easy and hard cases." In: *Networks* (2019). DOI: 10.1002/net.21865.

[26] Hallack, M. and Vazquez, M. "European Union Regulation of Gas Transmission Services: Challenges in the Allocation of Network Resources through Entry/exit Schemes." In: *Utilities Policy 25* 25.5 (2013), pp. 23–32. DOI: 10.1016/j.jup.2013.01.003.

[27] Hansen, P., Jaumard, B., and Savard, G. "New branch-and-bound rules for linear bilevel programming." In: *SIAM Journal on Scientific and Statistical Computing* 13.5 (1992), pp. 1194–1217. DOI: 10.1137/0913069.

[28] Hirschhausen, C. "Reform der Erdgaswirtschaft in der EU und in Deutschland: Wieviel Regulierung braucht der Wettbewerb?" In: *Perspektiven der Wirtschaftspolitik* 7.1 (2006), pp. 89–103. DOI: 10.1111/j.1465-6493.2006.00200.x.

[29] Holz, F., Hirschhausen, C. von, and Kemfert, C. "A strategic model of European gas supply (GASMOD)." In: *Energy Economics* 30.3 (2008), pp. 766–788. DOI: 10.1016/j.eneco.2007.01.018.

[30] Hunt, P. *Entry-Exit transmission pricing with national hubs. Can it deliver a Pan-European wholesale market in gas?* Tech. rep. Oxford Institute of Energy Studies, 2008.

[31] Huppmann, D. "Endogenous production capacity investment in natural gas market equilibrium models." In: *European Journal of Operational Research* 231.2 (2013), pp. 503–506. DOI: 10.1016/j.ejor.2013.05.048.

[32] Ikonnikova, S. and Zwart, G. T. "Trade quotas and buyer power, with an application to the E.U. natural gas market." In: *Journal of the European Economic Association* 12.1 (2014), pp. 177–199. DOI: 10.1111/jeea.12064.

[33] Jansen, T., Lier, A. van, Witteloostuijn, A. van, and Ochssée, T. B. von. "A modified Cournot model of the natural gas market in the European Union: Mixed-motives delegation in a politicized environment." In: *Energy Policy* 41 (2012). Modeling Transport (Energy) Demand and Policies, pp. 280–285. DOI: 10.1016/j.enpol.2011.10.047.

[34] Jeroslow, R. G. "The polynomial hierarchy and a simple model for competitive analysis." In: *Mathematical Programming* 32.2 (1985), pp. 146–164. DOI: 10.1007/BF01586088.

[35] Kleinert, T., Grimm, V., and Schmidt, M. *Outer Approximation for Global Optimization of Mixed-Integer Quadratic Bilevel Problems*. Tech. rep. 2020. URL: http://www.optimization-online.org/DB_FILE/2019/12/7534.pdf.

[36] Kleinert, T., Labbé, M., Plein, F., and Schmidt, M. "There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization." In: *Operations Research* (2019). URL: http://www.optimization-online.org/DB_HTML/2019/04/7172.html. Forthcoming.

[37] Kleinert, T. and Schmidt, M. "Global Optimization of Multilevel Electricity Market Models Including Network Design and Graph Partitioning." In: *Discrete Optimization* 33 (2019), pp. 43–69. DOI: 10.1016/j.disopt.2019.02.002.

[38] Koch, T., Hiller, B., Pfetsch, M. E., and Schewe, L. *Evaluating Gas Network Capacities*. SIAM-MOS series on Optimization. SIAM, 2015. DOI: 10.1137/1.9781611973693.

[39] Labbé, M., Plein, F., and Schmidt, M. "Bookings in the European Gas Market: Characterisation of Feasibility and Computational Complexity Results." In: *Optimization and Engineering* 21.1 (2020), pp. 305–334. DOI: 10.1007/s11081-019-09447-0.

[40] Lozano, L. and Smith, J. C. "A value-function-based exact approach for the bilevel mixed-integer programming problem." In: *Operations Research* 65.3 (2017), pp. 768–786. DOI: 10.1287/opre.2017.1589.

[41] McCormick, G. P. "Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems." In: *Mathematical Programming* 10.1 (1976), pp. 147–175. DOI: 10.1007/BF01580665.

[42] Meran, G., Hirschhausen, C. von, and Neumann, A. "Access Pricing and Network Expansion in Natural Gas Markets." In: *Zeitschrift für Energiewirtschaft* 34.3 (2010), pp. 179–183. DOI: 10.1007/s12398-010-0028-7.

[43] Midthun, K. T., Bjorndal, M., and Tomasgard, A. "Modeling Optimal Economic Dispatch and System Effects in Natural Gas Networks." In: *The Energy Journal* 30.4 (2009), pp. 155–180. URL: https://ideas.repec.org/a/aen/journl/2009v30-04-a06.html.

[44] Midthun, K. T., Fodstad, M., and Hellemo, L. "Optimization Model to Analyse Optimal Development of Natural Gas Fields and Infrastructure." In: *Energy Procedia* 64 (2015), pp. 111–119. DOI: 10.1016/j.egypro.2015.01.014.

[45] Moore, J. T. and Bard, J. F. "The mixed integer linear bilevel programming problem." In: *Operations Research* 38.5 (1990), pp. 911–921. DOI: 10.1287/opre.38.5.911.

[46] Oliver, M. E., Mason, C. F., and Finnoff, D. "Pipeline congestion and basis differentials." In: *Journal of Regulatory Economics* 46.3 (2014), pp. 261–291. DOI: 10.1007/s11149-014-9256-9.

[47] Pineda, S. and Morales, J. M. "Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold." In: *IEEE Transactions on Power Systems* (2019). DOI: 10.1109/TPWRS.2019.2892607.

[48] Regionales Rechenzentrum Erlangen. *Woodcrest Cluster*. URL: https://www.anleitungen.rrze.fau.de/hpc/woody-cluster/ (visited on 06/18/2020).

[49] Rømo, F., Tomasgard, A., Hellemo, L., Fodstad, M., Eidesen, B. H., and Pedersen, B. "Optimizing the Norwegian Natural Gas Production and Transport." In: *Interfaces* 39.1 (2009), pp. 46–56. DOI: 10.1287/inte.1080.0414.

[50] Siddiqui, S. and Gabriel, S. A. "Modeling market power in the U.S. shale gas market." In: *Optimization and Engineering* 18.1 (2017), pp. 203–213. DOI: 10.1007/s11081-016-9310-9.

[51] Smeers, Y. *Gas models and three difficult objectives*. Tech. rep. Core Discussion Paper, 2008. URL: http://www.uclouvain.be/cps/ucl/doc/core/documents/coreDP2008_9.pdf.

[52] Vazquez, M., Hallack, M., and Glachant, J.-M. "Designing the European Gas Market: More Liquid and Less Natural?" In: *Economics of Energy and Environmental Policy* 1.3 (2012). DOI: 10.5547/2160-5890.1.3.3.

[53] Vielma, J. P., Ahmed, S., and Nemhauser, G. "Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions." In: *Operations Research* 58.2 (2010), pp. 303–315. DOI: 10.1287/opre.1090.0721.

[54] Weymouth, T. R. "Problems in Natural Gas Engineering." In: *Trans. Amer. Soc. of Mech. Eng.* 34.1349 (1912), pp. 185–231.

[55] Xu, P. and Wang, L. "An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions." In: *Computers & Operations Research* 41 (2014), pp. 309–318. DOI: 10.1016/j.cor.2013.07.016.

[56] Yang, Z., Zhang, R., and Zhang, Z. "An exploration of a strategic competition model for the European Union natural gas market." In: *Energy Economics* 57 (2016), pp. 236–242. DOI: 10.1016/j.eneco.2016.05.008.

[57] Yanıkoğlu, İ., Gorissen, B., and Hertog, D. den. "A Survey of Adjustable Robust Optimization." In: *European Journal of Operational Research* 277.3 (2019), pp. 799–813. DOI: 10.1016/j.ejor.2018.08.031.

[58] Zare, M. H., Borrero, J. S., Zeng, B., and Prokopyev, O. A. "A note on linearized reformulations for a class of bilevel linear integer problems." In: *Annals of Operations Research* 272.1-2 (2019), pp. 99–117. DOI: 10.1007/s10479-017-2694-x.

[59] Zwart, G. and Mulder, M. *NATGAS: a model of the European natural gas market*. Tech. rep. 144. CPB Netherlands Bureau for Economic Policy Analysis, 2006. URL: https://ideas.repec.org/p/cpb/memodm/144.html.

(T. Böttger) Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Discrete Optimization, Cauerstr. 11, 91058 Erlangen, Germany

Email address: tom.tb.boettger@fau.de

(V. Grimm) (A) Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Economic Theory, Lange Gasse 20, 90403 Nürnberg, Germany; (B) Energie Campus Nürnberg, Fürther Str. 250, 90429 Nürnberg, Germany

Email address: veronika.grimm@fau.de

(T. Kleinert) (A) Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Discrete Optimization, Cauerstr. 11, 91058 Erlangen, Germany; (B) Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Economic Theory, Lange Gasse 20, 90403 Nürnberg, Germany; (C) Energie Campus Nürnberg, Fürther Str. 250, 90429 Nürnberg, Germany

Email address: thomas.kleinert@fau.de

(M. Schmidt) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany

Email address: martin.schmidt@uni-trier.de