Efficient Formulations and Decomposition Approaches for Power Peak Reduction in Railway Traffic via Timetabling

Andreas Bärmann[1], Alexander Martin[1] and Oskar Schneider[2]

[1] Andreas.Baermann@fau.de
Alexander.Martin@fau.de
Lehrstuhl für Wirtschaftsmathematik,
Department Mathematik,
Friedrich-Alexander-Universität Erlangen-Nürnberg,
Cauerstraße 11, 91058 Erlangen, Germany

[2] Oskar.Schneider@fau.de
Gruppe Optimization
Fraunhofer Arbeitsgruppe für Supply Chain Services SCS,
Fraunhofer-Institut für Integrierte Schaltungen IIS,
Nordostpark 93, 90411 Nürnberg, Germany

Abstract

Over the last few years, optimization models for the energy-efficient operation of railway traffic have received more and more attention, particularly in connection with timetable design. In this work, we study the effect of load management via timetabling. The idea is to consider trains as time-flexible consumers in the railway power supply network and to use slight shifts in the departure times from the stations to avoid too many simultaneous departures. This limits peak consumption and can help to improve the stability of the power supply. To this end, we derive efficient formulations for the problem of an optimal timetable adjustment based on a given timetable draft, two of which even allow for totally unimodular polyhedral descriptions. The proper choice of the objective function allows to incorporate either priorities of the train operating companies or the infrastructure manager. These include the avoidance of large peaks in average or instantaneous consumption and the improved use of recuperated braking energy. To solve the arising optimization models efficiently, we develop specially-tailored exact Benders decomposition schemes which allow for the computation of high-quality solutions within very short time. In an extensive case study for German railway passenger traffic, we show that our methods are capable of solving the problem on a nationwide scale. We will see that the optimal adjustment of timetables entails a tremendous potential for reducing energy consumption.

Keywords: Railway Timetabling - Energy-Efficiency - Total Unimodularity - Decomposition - Mixed-Integer Programming

Mathematics Subject Classification: 90C90 - 90C57 - 49M27 - 90C11

1 Introduction

Traction energy is a major cost factor for any train-operating company. Deutsche Bahn AG, Germany's largest railway company, for example, pays an annual electricity bill in the order of 1 billion Euros for traction electricity alone. The largest part of this cost, about 75%, are due the high overall electricity consumption of railway traffic – in the case of Deutsche Bahn it is around

11 billion kWh per year. In fact, with about 2% of the total consumption, railway traffic is the biggest electricity consumer in Germany. However, up to 25% of the annual electricity costs of Deutsche Bahn AG do not depend on the amount of electricity consumed, but on the temporal distribution of the consumption. Having a power consumption profile that shows high peaks in some parts of the day and low valleys in others parts of the day is rather costly, while having a mostly balanced consumption profile without high peaks over the course of the day is cheapest. Indeed, pricing schemes which charge both overall consumption and peak demand are very typical for large electricity consumers, also in the manufacturing industry, for example.

Measures to reschedule variable loads in order to reduce peak power consumption are commonly called *load management* or *peak shaving*. In the case of railway traffic, this effect may be achieved by adapting the timetable and thus influencing the departure times of the trains. A train draws most power while accelerating; therefore too many simultaneous departures should be avoided in order to keep the overall consumption balanced. On the other hand, synchronizing the departure of one train with the arrival of another train is desirable, as the braking train can feed back its recuperated kinetic energy into the power supply network for the accelerating train to use. If no other train is nearby that could make use of it, this energy is lost. Altogether, considering electricity consumption when planning the timetable is a worthwhile consideration for a train-operating company in order to save energy costs.

Although the above numbers suggest a significant benefit by load management in railway traffic, there have been relatively few publications on the topic so far. The first major work in this field is that of Sansó and Girard (1997) who study the reduction of instantaneous power peaks in metro systems by delaying train departures, an approach known as *dwell-time control*. To this end, they develop a mixed-integer optimization model incorporating both operational timetabling constraints as well as the power consumption of the trains. The model is solved via a decomposition heuristic. Similar timetabling approaches are taken in Kim et al. (2010) and Kim et al. (2011) who use a rather simplistic model without operational constraints but allow shifting train departures in both directions and integrate the efficient use of recuperated energy. In the former paper, they use a local-search algorithm while in the latter the mixed-integer program (MIP) is solved directly. Albrecht (2010) focuses on *running-time control*, i.e. slowing down the trains, instead of dwell-time control to reduce energy consumption as well as power consumption and finds solutions via a genetic algorithm. Dwell-time control based on a genetic algorithm is pursued by Chen et al. (2005).

On the general field of energy-saving train operations, there is much more literature available. This includes energy-efficient driving strategies (see the extensive surveys by Feng et al. (2013), Scheepmaker et al. (2017) and Yin et al. (2017a)), operational measures (for instance Ragunathan et al. (2014), Hasegawa et al. (2014) and Kimura and Miyatake (2014)) or timetabling focussed on reducing overall energy consumption (see Su et al. (2013), Peña-Alcaraz et al. (2011), Li and Lo (2014), Fournier et al. (2012) and Gong et al. (2014)). Bärmann et al. (2020) undertake polyhedral studies for a clique-based combinatorial optimization problem which is then used to model and solve optimal timetable adaptation problems including travel time choice for the trains. Zhou et al. (2017) explicitly combine trajectory optimization and timetabling via a modelling based on shortest paths within a time-expanded network, discretizing space, time and velocity. The combination of both timetabling and trajectory optimization is also considered by Wang and Goverde (2019), who do trajectory optimization to redistribute travel times between different line segments of multiple train routes in order to fine-tune a given timetable. Further works in this direction are that of Mo et al. (2019), who also integrate passenger waiting times into their model, and that of Canca and Zarzo (2017), who consider line frequency and fleet size as well. Yin et al. (2016) investigate energy-efficient rescheduling of timetables under uncertain passenger demand, while the case of dynamic passenger demand is studied in Yin et al. (2017b). Optimized train routing, scheduling and control as part of an integrated real-time traffic management framework have been studied by Luan et al. (2018a) and Luan

et al. (2018b). For general approaches in railway timetabling, we refer to Cacchiani and Toth (2012). Optimal load management has also been investigated in other problem contexts, such as industrial production (see Lorenz et al. (2012) and Nelson et al. (2013), for example).

To the best of our knowledge, there is no prior published work on load management including the optimal use of braking energy on the scale of a large national railway network. In a joint project with Deutsche Bahn AG, Germany's largest railway company, our motivation was to explore the cost-saving potential of adapting a given timetable draft in order to achieve a more balanced expected power consumption profile over the day. The principal measure to achieve this aim is shifting the departure times of the trains within a certain prescribed range of plus-minus a few minutes around their departure times stated in the draft. The adapted timetable shall maintain the key structure of the timetable draft; in particular it shall keep up the same level of service to the customers and meet all security requirements. Using recuperated energy from braking trains to the fullest possible extent is a key tool to reduce the overall energy consumption of the train operations in a railway network. Recuperated energy can only be used if there is another train in the network which accelerates in the same moment. Thus, we also strive to synchronize arrivals with departures. To this end, we devise mixed-integer programming models for optimized timetabling capturing the above aspects as well as an efficient decomposition approach to solve country-scale instances.

Our present work continues that of Bärmann et al. (2017), who evaluated optimized timetabling with respect to different performance metrics for the power consumption induced by the timetable. These included reducing 15-minute average peak consumption both with and without considering recuperated energy as well as reducing fluctuation of consumption measured in different norms. However, the models used there were not yet able to solve instances of relevant size. In Bärmann et al. (2018), an analysis of the underlying polytope was carried out which allowed the reduction of 15-minute average peak consumption for much larger instances. However, these approaches still did not consider the optimized use of recuperation energy – a direction that is relevant especially for reducing the costs of the train-operating company.

In this article, we derive compact and tight timetabling models together with a specially-tailored Benders decomposition scheme to solve the problem including the efficient use of recuperated energy. In addition, we control instantaneous peak consumption in the optimized solution in order to ensure stability of the power supply system. We also give a detailed case study for an optimal adaptation of the Germany-wide Deutsche Bahn timetable for passenger traffic in which we demonstrate the benefits of our approach.

Our exposition begins with a description of the problem background in Section 2. In Section 3, we describe our timetable optimization model as well as two enhanced versions with integral underlying timetabling polytopes (the non-linearity of the objective functions still makes the problem NP-hard, however). Furthermore, we develop our Benders decomposition approach which greatly improves solution quality within the first few minutes of computation and is able to solve large-scale instances. The detailed case study for German passenger traffic can be found in Section 4, and our conclusions are given in Section 5.

2 Problem Description

In the following, we explain the technical background governing the operation of the railway power supply system and the load-dependent costs railway companies usually incur. We also introduce our assumptions for feasible timetable adaptations to maintain the structure of a given planning draft with respect to customer service and safety provisions.

2.1 Characteristics of Traction Energy Supply in Germany

The German railway network is almost completely owned and operated by DB Netz AG, which is a subsidiary of Deutsche Bahn AG (DB) and Germany's most important infrastructure manager (IM). It leases the tracks to the train operating companies (TOCs), of which there are around 300 in Germany. The major TOCs in passenger traffic are the DB-subsidiaries DB Regio AG for regional transport and DB Fernverkehr AG for long-distance transport. Private competition exists mostly in regional traffic on local lines and in freight transport, where the DB-subsidiary DB Cargo AG had to suffer from a significant decline in market share in recent years.

The traction energy for the trains is provided by DB Energie GmbH, which is itself a subsidiary of DB Netz. The energy is either directly produced from dedicated power plants or transferred from the public electricity network. For its distribution, DB Energie operates an almost nationwide traction power supply network throughout Germany, separate from the public power grid, with a total length of more than 7700 km. To minimize energy losses over long distances, the electricity is transported via the high-voltage transmission layer operated at 110 kV and a frequency of 16.7 Hz. So-called *substations* nearby the tracks transform the electricity to a voltage of 15 kV at which the trains are operating. There are about 180 such substations in Germany, or one at about every 60 km of electrified track, many being coupling points to the public power grid. The energy is then fed from the substation to the catenaries, which in turn power the trains. The newer trains can feed energy back to the catenaries when braking. This is called *recuperation*. Recuperated energy can be used by nearby trains for their acceleration, which reduces the overall energy consumption. The closer the trains are to each other, the less energy is lost in the transmission.

The workload of the network can change very rapidly – drops and increases of 300 MW within few minutes are very frequent as trains accelerate and brake. DB Energie has to satisfy the total demand for traction energy and has to keep the current frequency constant. When the demand increases, the rotation frequency of the generators in the power plants decreases. To compensate this effect, the generators need to be supplied with more primary energy. The opposite effect occurs when the demand drops and primary energy needs to be taken away from the generators to prevent a frequency increase.

In the following, we describe how optimized timetabling can help to reduce peak power consumption in the railway system and improve the use of recuperated braking energy.

2.2 Energy-Related Costs in the Railway System

DB Energie charges the TOCs both according to the total energy consumption of the trains they operate *and* for their maximum power consumption. For their total consumption, they pay a certain price per kilowatt-hour to which federal fees and apportionments are added. The second component is determined by the maximum peak power consumption of all trains over any 15-minute interval in the billing period. It is calculated by summing up the measured power consumptions of all trains together and then averaging this value over the four 15-minute intervals of the clock. To this end, each locomotive has a built-in counter which measures the power consumption in every second. The TOC then pays a certain rate per kilowatt which is multiplied with the highest such average value. For the invoicing, the railway company can choose between a monthly or a yearly billing period, where the load-dependent charge is typically somewhat higher in the monthly pricing scheme. The reason for including a load-dependent charge – which is usual for contracts with big consumers of electricity – is that the load-dependent charge covers the costs for the layout of the infrastructure. This accounts for the fact that it is basically the biggest consumer who determines the necessary dimensioning of the power supply system, as additional generation and transmission capacities cause extra costs. For feeding back energy to the catenaries, TOCs are rewarded with a fixed price per kilowatt-hour of recuperated braking energy, which is somewhat lower than what they pay to

buy the energy. This recuperated energy is, however, not directly subtracted from the amount of energy drawn, but is refunded separately.

As a result of this pricing scheme, the TOC has an active interest in keeping its peak consumption as low as possible to save money – the ideal demand pattern in this respect would be a consumption that is constant over time. Naturally, there exist periods in which there is a higher demand than at other times. For example, in the morning and afternoon rush hours, there is much commuter traffic, which, as a result, induces a higher load in the power supply system than in off-peak hours, when fewer trains are running. As recuperated energy is refunded separately and not subtracted when computing its peak consumption, we neglect it when taking the point of view of the TOC.

A way to reduce the power-dependent energy costs of a TOC can now be to adjust the timetable such that too many simultaneous departures, and the resulting energy-intensive acceleration phases, are desynchronized. This can already be achieved by slight shifts in the departure times of the trains by few minutes in both directions – making use of remaining degrees of freedom in the timetable draft before finally publishing it. Figure 1 shows the effect this can have on the power consumption profile of the TOC at the example of regional traffic in the German state of Bavaria (cf. instance *Bayern* in Section 4).

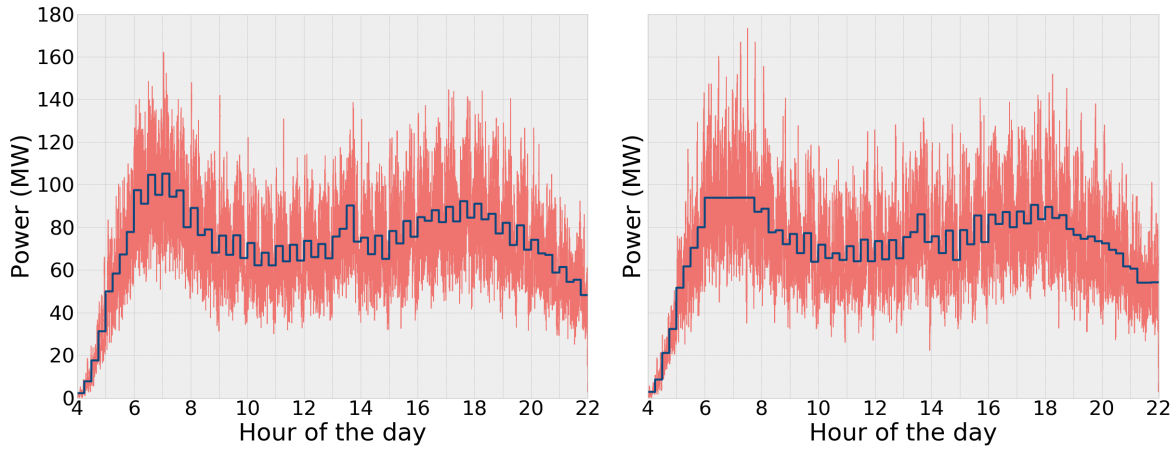

Figure 1: Power consumption profile for regional traffic in Bavaria (instance *Bayern*, cf. Section 4) with 3554 trains between 4 a.m. and 10 p.m., before (left) and after (right) optimization – instantaneous consumption in *red*, 15-minute averages in *blue*

It shows the power consumption profile of all regional trains running in the state between 4 a.m. and 10 p.m. – both before (left) and after (right) an optimal adjustment of the timetable. In *red*, we see the combined power consumption of the trains in each second, in *blue* we see the resulting 15-minute averages. For this instance, our approach can balance power consumption especially in the morning hours between 6 a.m. and 8 a.m., where we can reduce peak consumption by 11 MW or more than 10%. According to the official price sheet by DB Energie for 2020 (found on their homepage `https://www.dbenergie.de`), TOCs are charged 110.02 € per kilowatt in peak demand over a given year. The 105 MW in peak consumption in the unoptimized timetable would thus amount to a power-dependent cost of 11.6 million € for the current year of planning, compared to 10.3 million € for the 94 MW in the optimized timetable. Thus, by slightly adjusting the considered timetable, DB Regio could save well over 1 million € annually for this subnetwork alone. This result was achieved by allowing to shift train departures by not more than $\pm$ 3 minutes around their currently-planned departure times.

In a very similar fashion, the IM can use timetable adaptation to reduce peaks in power consumption from the public power supply network (or dedicated power plants feeding the railway

network). The difference is that it needs to take into account recuperated energy from braking trains which can be reused by other trains in order to lower the overall amount of energy drawn from the power supply. However, recuperated energy which cannot immediately be used by other trains is lost, such that a synchronization of braking trains with accelerating trains is necessary. Consequently, when braking energy does not find a receiving train it cannot be used to lower the average power consumption over the current 15-minute interval. A further concern to the IM is to ensure stability in the railway power supply system. To this end, it is also necessary to keep instantaneous peaks in consumption, lasting up to several seconds, under control.

2.3 Modelling Assumptions Concerning the Timetable Adaptation

For the optimal adaptation of a given timetable draft with respect to energy consumption, we take several modelling assumptions as described in the following. As our planning horizon, we consider a representative day of the week for which we want to adapt the departure times of the trains in the stations slightly. In doing so, we want to preserve the structure of the timetable draft, making use of the remaining degrees of freedom in the current state of planning to come to favourable energy consumption patterns. Firstly, we only allow that the departures of the trains at each station can be varied within a short time interval around the currently planned departure time. The typical value we consider here is a shift of $\pm$ 3 minutes, although we also experiment with smaller and larger values. This already ensures that the frequency at which stations are serviced remains vastly intact, while enabling the desynchronization of too many simultaneous departures and the better synchronization of accelerating trains with braking trains. Secondly, we demand that the dwell time of each train in each station is at least as long as in the current timetable draft. This ensures that passengers have enough time to get on and off the train. Thirdly, we require that the order of the departures in a given station in a given direction be retained as established in the timetable draft. In other words, our model considers the order in which trains pass the tracks in the network as fixed. In most of the available timetabling literature, the order of the trains on the tracks is subject to optimization. Usually this leads to timetabling models which either use big-M-formulations or act on time-expanded networks to take into account conflicts between train departures with respect to safety distances (see Cacchiani and Toth (2012) for a discussion of these two approaches). In our case, it is not too conservative to assume fixed orders as the existing timetable draft is to be adapted only slightly. Furthermore, it will allow for a much more compact model formulation which neither requires big-M-formulations nor time-expanded networks to model minimum headway time constraints. The minimum headway time between two trains depends on the track characteristics and on the types of trains involved. For a fast train following a slow train, it is higher than the other way round, for example. A given order of the trains on a track implies that we already know the required minimum headway time by which any two consecutive trains on that track have to be separated at least. Thus, the usual binary ordering variables in big-M-models can be considered as fixed in our case. For the incorporation of the safety constraints, we then assume that it is enough to enforce the minimum headway time at the two terminal stations of the track to ensure a sufficient separation of the trains. That means we require that the departure times of two consecutive trains are separated by at least the minimum headway time later than the preceding train, and similarly for the arrival times in the subsequent station. Finally, we require that any interchange relations between two trains in the network are maintained as established in the timetable draft. That means all connecting trains that are reachable from a given train in a given station have to be reachable after timetable optimization as well. To enforce this, we use the following rule: whenever one train arrives at a certain station, and a second trains departs within a certain time interval after that arrival (e.g. between 5 and 15 minutes later) in the timetable draft, we require that this time window be kept up in the optimized timetable as well. In addition, we require that the changeover time between the two trains does not change by too much, e.g. by at most $\pm$ 3 minutes. This is to ensure the same degree of service to the

passenger as in the currently planned timetable.

Using the feasible adaptations of the timetable draft outlined above, we look for a new timetable that leads to a lower peak power consumption. As a basis for these considerations, we assume that the power profiles of the trains are known in advance, i.e. the power consumption over their journey between two stations is given as an input and is independent from the chosen departure time. Power consumptions are assumed to be given in a resolution of 1 second, which corresponds to the resolution of the consumption counters in the locomotives.

3 Mathematical Models

In the following, we will introduce two optimization models that try to achieve different goals with respect to energy efficiency. The first model will take the point of view of a TOC that wants to reduce its load-dependent charges by the IM. The second model intends to reduce the overall power that has to be provided by the IM, thus decreasing the power production costs.

Both models optimize over the same feasible set, namely the set of all possible timetable adaptations, and differ only in their objective functions, which are non-linear in all cases, but can be linearized by adding further variables and constraints. Nevertheless, we will see that the first model will be significantly easier to solve, as the linearization of its objective function is much more lightweight. For the feasible set of the models, we will be able to give totally unimodular polyhedral descriptions of polynomial size – making them as tight as possible and compact at the same time.

3.1 Modelling Feasible Timetable Adaptations

We start by describing the feasible set underlying both models, namely the set of possible timetable adaptations. For this set, we will give three equivalent formulations of which the last two possess totally unimodular constraint matrices.

3.1.1 Naive Formulation

Let $T := \{0, 1, \dots, \Delta\}$ with $\Delta \in \mathbb{N}$ be the planning horizon of the problem, i.e. a part of the considered sample day. We assume that its resolutions is in seconds and that Δ is divisible by 900 for ease of exposition when computing power consumption averages over 15-minute-intervals. Let R be the set of trains running during the planning horizon. Then we denote by V^r the set of stations from which train $r \in R$ departs and by A^r the set of ordered pairs of subsequently serviced stations. In other words, if $(v, w) \in A^r$, then train r visits station w directly after station v. Let $J^{rv} \subseteq T$ be the set of feasible departure times for train r in station $v \in V^r$. The time train r needs to travel line $a = (v, w) \in A^r$ is denoted by Γ^{ra}, while c^{rw} denotes the minimum dwell time in station w. For each line $a = (v, w)$ between to stations v and w, we introduce the set L^a, which contains all ordered pairs (r_1, r_2) of consecutive trains r_1 and r_2 on that line (r_2 needs to travel the line a after r_1). Any two such trains have to respect a corresponding minimum headway time $s^{r_1 r_2 a}$. Finally, for each station v, we introduce the set U^v of all ordered pairs (r_1, r_2) of trains r_1 and r_2 for which an interchange at station v must be possible. In this case, the time that may pass between the arrival of train r_1 at station v and the departure of train r_2 from that station must be at least $\rho^{r_1 r_2 v}$ and at most $\sigma^{r_1 r_2 v}$ time steps after the arrival of train r_1.

For the decision to have train $r \in R$ depart from station $v \in V^r$ at instant $j \in J^{rv}$ we introduce the binary variable x_j^{rv}, which takes a value of 1 if this decision is taken, and 0 otherwise. Furthermore, let $V := \bigcup_{r \in R} V^r$ denote the set of all stations from which trains depart, and let $A := \bigcup_{r \in R} A^r$ denote the set of all lines between to stations travelled in the timetable. A feasible

timetable adaptation can then be described by the following set of constraints:

$$\sum_{j \in J^{rv}} x_j^{rv} = 1 \qquad (\forall r \in R)(\forall v \in V^r) \tag{1a}$$

$$x_j^{rv} \leq \sum_{\substack{h \in J^{rw}: \\ h \geq j + \Gamma^{ra} + c^{rw}}} x_h^{rw} \qquad (\forall r \in R)(\forall a = (v,w) \in A^r)(\forall j \in J^{rv}) \tag{1b}$$

$$x_j^{r_1 v} \leq \sum_{\substack{h \in J^{r_2 v}: \\ h \geq j + \max\{0, \Gamma^{r_1 a} - \Gamma^{r_2 a}\} + s^{r_1 r_2 a}}} x_h^{r_2 v} \qquad (\forall a = (v,w) \in A)(\forall (r_1, r_2) \in L^a)(\forall j \in J^{r_1 v}) \tag{1c}$$

$$x_j^{r_1 u} \leq \sum_{\substack{h \in J^{r_2 v}: \\ h \geq j + \Gamma^{r_1 a} + \rho^{r_1 r_2 v} \\ h \leq j + \Gamma^{r_1 a} + \sigma^{r_1 r_2 v}}} x_h^{r_2 w} \qquad \begin{array}{l}(\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1}) \\ (\forall j \in J^{r_1 u})\end{array} \tag{1d}$$

$$x_j^{rv} \in \{0,1\} \qquad (\forall r \in R)(\forall v \in V^r)(\forall j \in J^{rv}). \tag{1e}$$

Constraint (1a) ensures that each train departs from each of its stations at exactly one of the predefined possible departure times. With Constraint (1b) we make sure that the minimum dwell times in each station are respected. Having train r depart from station v at instant j along line (v,w) implies that the earliest possible point in time to depart from the next station w is after the travel time Γ^{ra} and the dwell time c^{rw} have passed. In a similar fashion, Constraint (1c) enforces the minimum headway time between two successive trains on a line. If train r_1 on line a is followed by a train r_2 that is at most as fast as r_1, this second train can depart $s^{r_1 r_2 a}$ time steps after r_1 at the earliest. If train r_2 is faster than r_1, then the time that has to pass after the departure of r_1 is at least $\Gamma^{r_1 a} - \Gamma^{r_2 a}$ to ensure that r_2 arrives at least $s^{r_1 r_2 a}$ time steps after r_1 at the next station. We assume that this suffices to guarantee in both cases that the two trains are separated by at least $s^{r_1 r_2 a}$ time steps while passing line a. Next, Constraint (1d) models the requirement to keep all desired possibilities for interchanges between the trains in the stations intact. The departure of train r_2 from station v has to happen within a certain time window around the arrival of train r_1 in station v if an interchange shall be possible. Finally, Constraint (1e) requires all x-variables to take binary values only. Note that the sets J^{rv} are assumed to contain the currently planned departure times of the corresponding trains from the corresponding stations, and we assume that these departure times fulfil all the requirements laid out above. In other words, the timetable draft to be optimized is always a feasible solution to the above system.

Constraints (1a)–(1e) can be interpreted as the feasible set of a project scheduling problem, where Constraint (1a) is the requirement to schedule all the tasks, and where the latter three constraints represent precedences between these jobs. It is known from the literature that this feasible set can be reformulated such that the constraint matrix of the corresponding constraints is totally unimodular (see e.g. Schwindt and Zimmermann (2015)). In a more general sense, Constraints (1a)–(1e) can be interpreted as a clique problem with multiple-choice constraints (CPMC). In the recent work Bärmann et al. (2018), the authors study so-called *staircase compatibility*, a generalization of the precedence structure of a project scheduling problem. It leads to compatibility graphs for which this clique problem can be modelled by a linear program (LP) with a totally unimodular constraint matrix. In the following, we will use the reformulations from Schwindt and Zimmermann (2015); Bärmann et al. (2018) to obtain compact unimodular formulations for System (1).

3.1.2 Totally Unimodular Formulations

Our second formulation for the set of feasible timetable adaptations results from a lifting of Constraints (1b)– (1d). In Model (1), their left-hand sides consist of a single variable each. It models the departure of a certain train from a certain station at a certain point in time. By

adding all variables on top which correspond to possible later departures of the same train at the same station, we can strengthen these constraints. This leads to the new system

$$\sum_{j \in J^{rv}} x_j^{rv} = 1 \qquad (\forall r \in R)(\forall v \in V^r) \tag{2a}$$

$$\sum_{\substack{k \in J^{rv}: \\ k \geq j}} x_k^{rv} \leq \sum_{\substack{h \in J^{rw}: \\ h \geq j + \Gamma^{ra} + c^{rw}}} x_h^{rw} \qquad (\forall r \in R)(\forall a = (v,w) \in A^r)(\forall j \in J^{rv}) \tag{2b}$$

$$\sum_{\substack{k \in J^{r_1 v}: \\ k \geq j}} x_k^{r_1 v} \leq \sum_{\substack{h \in J^{r_2 v}: \\ h \geq j + \max\{0, \Gamma^{r_1 a} - \Gamma^{r_2 a}\} + s^{r_1 r_2 a}}} x_h^{r_2 v} \qquad \begin{array}{l}(\forall a = (v,w) \in A)(\forall (r_1, r_2) \in L^a) \\ (\forall j \in J^{r_1 v})\end{array} \tag{2c}$$

$$\sum_{\substack{k \in J^{r_1 u}: \\ k \geq j}} x_k^{r_1 u} \leq \sum_{\substack{h \in J^{r_2 v}: \\ h \geq j + \Gamma^{r_1 a} + \rho^{r_1 r_2 v}}} x_h^{r_2 v} \qquad \begin{array}{l}(\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1}) \\ (\forall j \in J^{r_1 u})\end{array} \tag{2d}$$

$$\sum_{\substack{i \in J^{r_2 v}: \\ i \geq h}} x_i^{r_2 v} \leq \sum_{\substack{j \in J^{r_1 u}: \\ j \geq h - \Gamma^{r_1 a} - \sigma^{r_1 r_2 v}}} x_j^{r_1 u} \qquad \begin{array}{l}(\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1}) \\ (\forall h \in J^{r_2 v})\end{array} \tag{2e}$$

$$x_j^{rv} \in \{0,1\} \qquad (\forall r \in R)(\forall v \in V^r)(\forall j \in J^{rv}). \tag{2f}$$

The validity of these new constraints can easily be verified. Take for example Constraint (2b) for some $r \in R$, $a = (v,w) \in A^r$ and $j \in J^{rv}$. It arises by considering Constraints (1b) for the same r and a as well as all $k \in J^{rv}$ with $k \geq j$ and summing them up, which yields

$$\sum_{\substack{k \in J^{rv}: \\ k \geq j}} x_k^{rv} \leq \sum_{\substack{k \in J^{rv}: \\ k \geq j}} \sum_{\substack{h \in J^{rw}: \\ h \geq k + \Gamma^{ra} + c^{rw}}} x_h^{rw} = \sum_{\substack{h \in J^{rw}: \\ h \geq j + \Gamma^{ra} + c^{rw}}} (h - j - \Gamma^{ra} - c^{rw} + 1) x_h^{rw}.$$

Now we observe that the left-hand side can never take a value higher than 1 due to Constraint (1a). Moreover, we see above that the right-hand side can be reformulated such that all x-variables have integer coefficients of value 1 or higher. Changing all these coefficients to 1 does not change the set of binary solutions fulfilling the constraint, as a value of 1 on the right-hand side is already enough to dominate the left-hand side, which can be either 0 or 1. This reasoning leads to Constraint (2b). Another way to obtain this strengthened constraint is as a 0-$\frac{1}{2}$-cut by suitably combining Constraints (1a) and (1b). In a similar fashion, we can obtain the remaining constraints of System (2), where for reformulating Constraint (1d) we split it up into

$$x_j^{r_1 u} \leq \sum_{\substack{h \in J^{r_2 v}: \\ h \geq j + \Gamma^{r_1 a} + \rho^{r_1 r_2 v}}} x_h^{r_2 v} \qquad (\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1})(\forall j \in J^{r_1 u})$$

and

$$x_j^{r_1 u} \leq \sum_{\substack{h \in J^{r_2 v}: \\ h \leq j + \Gamma^{r_1 a} + \sigma^{r_1 r_2 v}}} x_h^{r_2 v} \qquad (\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1})(\forall j \in J^{r_1 u}).$$

With the former constraint, we can proceed as before to obtain Constraint (2d). The latter can be reformulated by fixing some v, (r_1, r_2) and the according a and then summing over all $k \leq j$ for all possible choices of j. A similar consideration for the variable coefficients as above and exploiting Constraint (1a) on both sides then yields Constraint (2e). Altogether, we have derived a tighter formulation for the set of feasible timetable adjustments. As shown in Bärmann et al. (2018), this lifting of the constraints leads to a totally unimodular constraint matrix. Therefore, we can replace Constraint (2f) by

$$x_j^{rv} \geq 0 \quad (\forall r \in R)(\forall v \in V^r)(\forall j \in J^{rv}), \tag{3}$$

where we use that this constraint together with Constraint (2a) already imply the upper bounds of 1. We will refer to the formulation above as the totally unimodular formulation (TU) for the set of feasible timetable adaptations.

From TU, it is possible to obtain a second reformulation by performing a variable substitution. We introduce new variables y_j^{rv}, $r \in R$, $v \in V^r$, $j \in J^{rv}$, which are defined via

$$y_j^{rv} := \sum_{\substack{k \in J^{rv}: \\ k \geq j}} x_k^{rv}.$$

Their interpretation is as follows: variable y_j^{rv} takes a value of 1 if train r departs from station v at minute j or a later feasible point in time. Otherwise, y_j^{rv} takes a value of 0. Replacing the x-variables in System (2) by these new variables leads to

$$\begin{aligned}
y^{rv}_{\min(J^{rv})} &= 1 && (\forall r \in R)(\forall v \in V^r) && \text{(4a)}\\
y_j^{rv} &\leq y_h^{rw} && (\forall r \in R)(\forall a = (v,w) \in A^r) \\ & && (\forall j \in J^{rv} : h := j + \Gamma^{ra} + c^{rw} \leq \max(J^{rw})) && \text{(4b)}\\
y_j^{rv} &= 0 && (\forall r \in R)(\forall a = (v,w) \in A^r) \\ & && (\forall j \in J^{rv} : j + \Gamma^{ra} + c^{rw} > \max(J^{rw})) && \text{(4c)}\\
y_j^{r_1 v} &\leq y_h^{r_2 v} && (\forall a = (v,w) \in A)(\forall (r_1, r_2) \in L^a) \\ & && (\forall j \in J^{r_1 v} : h := j + \max\{0, \Gamma^{r_1 a} - \Gamma^{r_2 a}\} + s^{r_1 r_2 a} \leq \max(J^{r_2 v})) && \text{(4d)}\\
y_j^{r_1 v} &= 0 && (\forall a = (v,w) \in A)(\forall (r_1, r_2) \in L^a) \\ & && (\forall j \in J^{r_1 v} : j + \max\{0, \Gamma^{r_1 a} - \Gamma^{r_2 a}\} + s^{r_1 r_2 a} > \max(J^{r_2 v})) && \text{(4e)}\\
y_j^{r_1 u} &\leq y_h^{r_2 v} && (\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1}) \\ & && (\forall j \in J^{r_1 u} : h := j + \Gamma^{r_1 a} + \rho^{r_1 r_2 v} \leq \max(J^{r_2 v})) && \text{(4f)}\\
y_j^{r_1 u} &= 0 && (\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1}) \\ & && (\forall j \in J^{r_1 u} : j + \Gamma^{r_1 a} + \rho^{r_1 r_2 v} > \max(J^{r_2 v})) && \text{(4g)}\\
y_h^{r_2 v} &\leq y_j^{r_1 u} && (\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1}) \\ & && (\forall h \in J^{r_2 v} : j := h - \Gamma^{r_1 a} - \sigma^{r_1 r_2 v} \leq \max(J^{r_1 v})) && \text{(4h)}\\
y_h^{r_2 v} &= 0 && (\forall v \in V)(\forall (r_1, r_2) \in U^v, a = (u,v) \in A^{r_1}) \\ & && (\forall h \in J^{r_2 v} : h - \Gamma^{r_1 a} - \sigma^{r_1 r_2 v} > \max(J^{r_1 v})) && \text{(4i)}\\
y_{j+1}^{rv} &\leq y_j^{rv} && (\forall r \in R)(\forall v \in V^r)(\forall j \in J^{rv} : j < \max(J^{rv})) && \text{(4j)}\\
y^{rv}_{\max(J^{rv})} &\geq 0 && (\forall r \in R)(\forall v \in V^r), && \text{(4k)}
\end{aligned}$$

where $\min(J^{rv})$ denotes the earliest possible departure time in that set, and $\max(J^{rv})$ the latest possible departure time. Constraints (4j) and (4k) originate from the lower bounds (3) of the x-variables. System (4) is the feasible set of a dual flow problem, recognizable by the fact that each row has at most two non-zero entries and each line with two non-zeros has exactly one coefficient $+1$ and one coefficient -1. Accordingly, we refer to System (4) as the dual-flow formulation (DF). By its nature, the constraint matrix corresponding to this formulation is totally unimodular, too. The simplex method in an MIP solver can usually benefit from the sparsity of the constraints in formulation (DF). Indeed, our results in Section 4 will strongly confirm this.

Although we have now found totally unimodular formulations for the set of feasible timetable adaptations, we will see in the following that finding optimal adaptations with respect to reducing peak consumption is still a hard problem.

3.2 Modelling the Reduction of Peak Power Consumption

In this section, we present two optimization models for the reduction of peak power consumption – one focussed on reducing the energy costs of the TOCs, the other focussing on the cost of

energy provision by the IM as well as stability of supply. For both models, we can use all three formulations for feasible timetable adaptations from above. In doing so, we will use the set X for the integral solutions of either Formulation (1) or (2) (where both are used interchangeably in this section, as they live in the same space of variables and contain the same integer points). Accordingly, we will use the set Y for the integral solutions of Formulation (4).

3.2.1 Reducing the Peak Power Charge of TOCs

We start with a model for minimizing the component of the energy bill of a TOC that depends on peak power consumption. Its main purpose is the desynchronization of too many simultaneous train departures (involving trains of that company). To this end, let $p^{rat} \in \mathbb{R}$ be the consumption of train $r \in R$ on line $a = (v, w) \in A^r$ at instant $0 \leq t \leq \Gamma^{ra}$ after departure. Shifting the departure time of the train amounts to shifting the occurrence of the resulting power consumption in time. Thus, if train r departs from station v at time j, its consumption at instant $t \in T$ is given by

$$\bar{p}_j^{rat} := \begin{cases} p^{rat}, & 0 \leq t - j \leq \Gamma^{ra} \\ 0, & \text{otherwise.} \end{cases}$$

By summing over all the trains of a TOC running at a given point in time $t \in T$ according to a feasible timetable $x \in X$, we obtain its total gross power consumption

$$P^+(x, t) := \sum_{r \in R} \sum_{a=(v,w) \in A^r} \sum_{j \in J^{rv}} \max\{\bar{p}_j^{rat}, 0\} x_j^{rv}, \tag{5}$$

where we cut off negative consumption values as recuperated energy is refunded separately. Now, the peak consumption charge is not computed from the instantaneous consumption values but from the average consumption values over 15-minute intervals. Therefore, we introduce the set $T_{15} := \{[900 \cdot i, 900 \cdot [i+1] \mid i \in \{0, \ldots, \frac{\Delta}{900} - 1\}\}$ of all consecutive 15-minute (= 900-second) intervals in the planning horizon T. The gross energy consumption over one such interval $I \in T_{15}$ under a chosen timetable $x \in X$ is then given by

$$E^+(x, I) = \frac{1}{2}(P^+(x, t_0(I)) + P^+(x, t_{900}(I))) + \sum_{\substack{t \in I: \\ t \neq t_0(I) \wedge \\ t \neq t_{900}(I)}} P^+(x, t), \tag{6}$$

where $t_0(I)$ and $t_{900}(I)$ denote the first and the last second of that interval respectively. The above formula assumes that the actual power consumption profile, for which we know the values at the points $t \in T$, is a continuous function, such that it is reasonable to approximate the area between its graph and the horizontal axis via the trapezoidal rule. Altogether, the optimization problem of finding a feasible timetable adaptation which minimizes the maximum average gross power consumption over any 15-minute interval in the planning horizon can be stated as

$$\frac{1}{900} \cdot \min_{x \in X} \max_{I \in T_{15}} E^+(x, I).$$

We denote this problem by (TTMAPBR), as it tries to find a TimeTable with a minimal Maximum Average Power Consumption Before Recuperation. By introducing variables for the power consumption in each second, this problem can easily linearized. An important observation is now that in Equation (6) the summation over t interchanges with the summations over r, v and j in Equation (5). Thus, the computation of $E^+(x, I)$ can be dramatically simplified to

$$E^+(x, I) = \sum_{r \in R} \sum_{a=(v,w) \in A^r} \sum_{j \in J^{rv}} E^{+,\text{trip}}(r, v, j, I) x_j^{rv},$$

where the portion of the gross energy consumption $E^{+,\text{trip}}(r,v,j,I)$ of train r on line a that falls into interval I when departing at instant j can be precomputed as

$$E^{+,\text{trip}}(r,v,j,I) = \frac{1}{2}\left(\max\{\bar{p}_j^{rat_0(I)},0\} + \max\{\bar{p}_j^{rat_{900}(I)},0\}\right) + \sum_{\substack{t\in I:\\ t\neq t_0(I)\wedge\\ t\neq t_{900}(I)}} \max\{\bar{p}_j^{rat},0\}.$$

This leads to a much lower number of variables and constraints necessary for the linearization of Problem (TTMAPBR). Indeed, we only need one additional variable z for the linearization of the objective function, with which the problem can be stated as the following MIP:

$$\min \quad z \tag{7a}$$

$$\text{s.t.} \sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}} E^{+,\text{trip}}(r,v,j,I)x_j^{rv} \leq 900z \quad (\forall I\in T_{15}) \tag{7b}$$

$$x \in X. \tag{7c}$$

Alternatively, we can also use the representation of the set of feasible timetables in the y-variables and write the problem as

$$\min \quad z \tag{8a}$$

$$\text{s.t.} \quad \frac{1}{900}\cdot\sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}} E^{+,\text{trip}}(r,v,j,I)(y_j^{rv}-y_{j+1}^{rv}) \leq z \quad (\forall I\in T_{15}) \tag{8b}$$

$$y \in Y, \tag{8c}$$

where we define $y^{rv}_{\max(J^{rv})+1} \coloneqq 0$ for all $r\in R$, $v\in V^r$.

Remark 3.1. *Note that the gross energy consumption of a train r on line a that falls into interval I, depending on the the chosen departure time j, can also be written as*

$$E^{+,trip}(r,v,\min(J^{rv}),I) + \sum_{\substack{j\in J^{rv}:\\ j>min(J^{rv})}} (E^{+,trip}(r,v,j,I) - E^{+,trip}(r,v,j-1,I))y_j^{rv},$$

where we use that $y^{rv}_{\min(J^{rv})} = 1$ holds according to System (4)*. Thus, we can interpret the change from the x- to the y-variables as the change between the multiple-choice method and the incremental method for modelling piecewise-linear functions. Constraints* (4j) *and* (4k) *together can be recognized as a filling condition in this context. See Vielma et al. (2010); Bärmann et al. (2018); Liers and Merkert (2016) for a broader discussion of this aspect. The incorporation of the piecewise-linear function representing the power consumption of the trains was the reason for choosing time-discretized variables in the first place.*

The fact that the solution of either Model (7) or Model (8) amounts to the optimization of a piecewise-linear function over a totally unimodular system might suggest that this timetabling problem is easy to solve. However, this is not the case, as we show in the following.

Theorem 3.2 (Personal communication with Zhu (2014))**.** *Problem (TTMAPBR) is weakly NP-hard, even if there is only one departure to schedule per train, there are only two possible departure times to choose from, and the power consumption values are positive integers.*

Proof. We show the claim by a reduction from the partition problem (see Garey and Johnson (1979), Problem SP12). Let S be some index set and $a_s \in \mathbb{N}$ for $s\in S$ be positive integers. To decide whether there are two disjoint subsets $S_1, S_2 \subseteq S$ with $S = S_1 \cup S_2$ and $\sum_{s\in S_1} a_s = \sum_{s\in S_2} a_s$, we can solve the following instance of (TTMAPBR): Let $T = \{0,1\}$ be the planning horizon and assume for the sake of simplicity that the intervals to average the consumption over are only one second long each. Interpret S as the set of trains of which each has one

departure to be scheduled, i.e. $|A^s| = 1$ for all $s \in S$, and assume that these departures can be scheduled independently. Let all the trains have the same set of two possible departure times, namely, $J^s = \{0, 1\}$ for all $s \in S$. Furthermore, assume that the travel time of each train s on the trip to be scheduled is only one second, and that the corresponding power consumption value in that second is a_s. Then solving (TTMAPBR) for this instance essentially amounts to solving the original partition instance, as (TTMAPBR) will try to find the schedule that perfectly balances the power consumptions between the two instants of the planning horizon. If a perfect balancing is possible, this indicates the existence of a feasible partition, otherwise there is no such partition. This proves the claim. □

3.2.2 Reducing the Energy Provision Costs of the IM

Minimizing the costs of providing traction energy to the TOCs from the point of view of the IM leads to an optimization problem very similar to (TTMAPBR). The important difference is that the IM is very interested in ensuring effective use of the energy recuperated from braking to reduce peaks in power consumption from the public power supply network. This requires to take negative power consumption values into account, too, as laid out in the following.

We define the net power consumption of all trains running according to a given timetable $x \in X$ at instant $t \in T$ as

$$P(x,t) := \max\left(\sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}} \bar{p}_j^{rat} x_j^{rv}, 0\right). \tag{9}$$

The difference to the gross power consumption $P^+(x,t)$ is that we do not cut off negative contributions, but balance them against positive contributions instead. If at any instant there is more recuperated energy than can be used by trains for their traction, this energy is lost, which explains the maximum in Equation (9). The net energy consumption over an interval $I \in T_{15}$ is then given by

$$E(x,I) = \frac{1}{2}(P(x,t_0(I)) + P(x,t_{900}(I))) + \sum_{\substack{t\in I:\\ t\neq t_0(I)\wedge\\ t\neq t_{900}(I)}} P(x,t).$$

Here we cannot simplify the computation of the energy consumption as before, as the involved summations do not interchange. This is the main reason why the problem of minimizing the average net power consumption, to which we refer as (TTMAP) for TimeTabling with respect to Maximum Average Power Consumption, is much harder to solve. It can be written as

$$\frac{1}{900}\cdot \min_{x\in X}\max_{I\in T_{15}} E(x,I).$$

A similar linearization as before now requires continuous variables $w_t \in \mathbb{R}$ for the net power consumption in each instant $t \in T$, in addition to variable z for linearizing the objective function. Problem (TTMAP) can then be written as the following MIP:

$$\begin{aligned}
\min\quad & z && && && (10a)\\
\text{s.t.}\quad & \frac{1}{2}(w_{t_0(I)} + w_{t_{900}(I)}) + \sum_{t\in I} w_t &\leq\ & 900z && (\forall I \in T_{15}) && (10b)\\
& \sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}} \bar{p}_j^{rat} x_j^{rv} &\leq\ & w_t && (\forall t \in T) && (10c)\\
& w_t &\geq\ & 0 && (\forall t \in T) && (10d)\\
& x &\in\ & X. && && (10e)
\end{aligned}$$

Within Formulation (DF), Constraint (10b) can be stated as

$$\sum_{r\in R}\sum_{a=(v,w)\in A^r}\left(\bar{p}^{rat}_{\min(J^{rv})}y^{rv}_{\min(J^{rv})}+\sum_{\substack{j\in J^{rv}:\\ j<\max(J^{rv})}}(\bar{p}^{rat}_{j+1}-\bar{p}^{rat}_{j})y^{rv}_{j}\right)\leq w_t \quad (\forall t\in T). \tag{11}$$

An important consideration for the IM is not only to reduce average peak consumption but also to limit instantaneous power consumption. As the solutions produced by Model (10) might increase instantaneous power consumption, we also consider a variant of the model providing an upper limit. More precisely, if $U\in\mathbb{R}$ is the maximum power consumption occurring in the initial timetable in any instant during the planning horizon, we add the following constraint ensuring that this value is never exceeded:

$$w_t\leq U \quad (\forall t\in T). \tag{12}$$

We will see that the addition of this constraint empirically does not significantly limit the solution space of feasible timetables, such that we retain most of the potential reduction in maximum average power consumption. In the following, we will refer to Problem (TTMAP) together with Constraint (12) for limiting Instantaneous power consumption (TTMAPI).

3.2.3 Solving Problems (TTMAP) and (TTMAPI) via Benders Decomposition

Note that independent from the choice of variables, our models for (TTMAP) require significantly more auxiliary variables and constraints than those for (TTMAPBR). This is why we could solve (TTMAPBR) within very short time using standard methods when applying our totally unimodular reformulations, even for very large-scale instances. This was not the case for (TTMAP) as needs it to assess the total power consumption in every second, compared to assessing only each 15-minute interval in (TTMAPBR). Moreover, Constraint (10b) and its reformulation (11) are both very dense inequalities, with potentially tens of thousands of non-zero entries for each $t\in T$ in an instance of realistic size. As this slows down the simplex method considerably, we now give a Benders reformulation of Model (10) with which we could significantly improve our results on large-scale instances (see Section 4). The idea is to project out the continuous w-variables to obtain a master problem which is basically formulated in the timetable variables only.

Assume that we are already given a feasible timetable $\bar{x}\in X$. The remaining subproblem is then to compute the actual objective value of $\bar{x}$ in Model (10):

$$\begin{aligned}
\min\quad & z && && && (13a)\\
\text{s.t.}\quad & \frac{1}{2}(w_{t_0(I)}+w_{t_{900}(I)})+\sum_{t\in I}w_t && \leq && 900z \quad (\forall I\in T_{15}) && (13b)\\
& \sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}}\bar{p}^{rat}_{j}\bar{x}^{rv}_{j} && \leq && w_t \quad (\forall t\in T) && (13c)\\
& w_t && \geq && 0 \quad (\forall t\in T). && (13d)
\end{aligned}$$

The corresponding dual subproblem is given by

$$\max \quad \sum_{t\in T}\left(\sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}}\bar{p}_j^{rat}\bar{x}_j^{rv}\right)\beta_t \tag{14a}$$

$$\begin{aligned}
\text{s.t.} \qquad \beta_t - \alpha_{I(t)} &\leq 0 && (\forall t \in T_1) && (14b)\\
\beta_t - \frac{1}{2}\alpha_{I_1(t)} - \frac{1}{2}\alpha_{I_2(t)} &\leq 0 && (\forall t \in T_2) && (14c)\\
\beta_t - \frac{1}{2}\alpha_{I(t)} &\leq 0 && (t = 0 \vee t = \Delta) && (14d)\\
900\sum_{I\in T_{15}}\alpha_I &= 1 && && (14e)\\
\alpha_I &\geq 0 && (\forall I \in T_{15}) && (14f)\\
\beta_t &\geq 0 && (\forall t \in T), && (14g)
\end{aligned}$$

where the newly-introduced symbols have the following meanings: α is the dual variable to Constraint (13b) and β that to Constraint (13c), $I(t)$ denotes the interval in T_{15} to which second t belongs, using $I_1(t)$ and $I_2(t)$ for the earlier and the later such interval, respectively, if t belongs to both, while $T_1 \coloneqq \{t \in T \mid t \bmod 900 \neq 0\}$ and $T_2 \coloneqq \{t \in T \mid t \bmod 900 = 0 \wedge t \neq 0, \Delta\}$. In an optimal solution to Model (14), each β_t will take a value that is as high as possible whenever its objective function coefficient is positive, which means that the corresponding constraint out of (14b)– (14d) is fulfilled with equality in this case. Thus, the model can be simplified to

$$\max \quad \sum_{t\in T_1}P(x,t)\alpha_{I(t)} + \frac{1}{2}\sum_{t\in T_2}P(x,t)(\alpha_{I_1(t)}+\alpha_{I_2(t)}) + \frac{1}{2}(P(x,0)\alpha_{I(0)} + P(x,\Delta)\alpha_{I(\Delta)}) \tag{15a}$$

$$\begin{aligned}
900\sum_{I\in T_{15}}\alpha_I &= 1 && && (15b)\\
\alpha_I &\geq 0 && (\forall I \in T_{15}). && (15c)
\end{aligned}$$

This remaining model is solvable by inspection, as clearly the α_I with the highest objective function coefficient will be set to $\frac{1}{900}$ and the others to 0. Let this index I be denoted by $I_{\max}(\bar{x})$ under timetable $\bar{x}$. Then $I_{\max}(\bar{x})$ is precisely the interval with the highest average power consumption. If $I^+_{\max}(\bar{x})$ denotes the instants $t \in I_{\max}(\bar{x})$ for which the power consumption $P(x,t)$ is positive, we can derive the following Benders cut for $\bar{x}$:

$$\sum_{t\in I^+_{\max}(\bar{x})}\sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}}\bar{p}_j^{rat}x_j^{rv} \leq 900z. \tag{16}$$

Inequality (16) computes the correct objective function value $\bar{z}$ for any given timetable $\bar{x} \in X$. For any timetable close in structure to $\bar{x}$, it can be expected to give a meaningful lower bound for the objective value of that timetable. This can be exploited by eliminating Constraints (10b), (10c) and (10d) from Model (10) to initialize the restricted Benders master problem and to separate Inequality (16) in a Benders fashion until an optimal timetable is found. In order to have reasonably good bounds for the objective value from the start, we add cutting planes similar to (7b) to the initial Benders master problem. To this end, we compute the values

$$E^{\text{trip}}(r,v,j,I) = \frac{1}{2}(\bar{p}_j^{rat_0(I)} + \bar{p}_j^{rat_{900}(I)}) + \sum_{\substack{t\in I:\\ t\neq t_0(I)\wedge\\ t\neq t_{900}(I)}}\bar{p}_j^{rat},$$

which are a lower bound for the net energy consumption that train r on line a contributes in time interval I when departing at instant j. It differs from $E^{+,\text{trip}}(r,v,j,I)$ in that we do not cut

off negative contributions. The inequality

$$\sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}} E^{\text{trip}}(r,v,j,I)x_j^{rv} \le 900z \quad (\forall I \in T_{15})$$

thus yields a valid lower bound for the objective value of any feasible timetable $x \in X$. They showed a significant benefit in speeding up our Benders scheme. We also compute the actual objective function value of any solution $\bar{x}$ cut off by the Benders cut (16) and propose it to the solver again, this time with the correct value for z. This way, we do not have to wait until the solution is found again.

In the case of Problem (TTMAPI), which limits instantaneous power consumption while minimizing average peak consumption, we can reformulate Constraint (12) to fit into this Benders scheme. The dual subproblem taking into account Constraint (12) and its corresponding dual variable γ takes the form

$$\begin{aligned}
\max \quad & \sum_{t\in T}\left(\sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}} \bar{p}_j^{rat}\bar{x}_j^{rv}\right)\beta_t - U\sum_{t\in T}\gamma_t && && (17a)\\
\text{s.t.} \quad & \beta_t - \gamma_t - \alpha_{I(t)} && \le\ 0 \quad (\forall t \in T_1) && (17b)\\
& \beta_t - \gamma_t - \frac{1}{2}\alpha_{I_1(t)} - \frac{1}{2}\alpha_{I_2(t)} && \le\ 0 \quad (\forall t \in T_2) && (17c)\\
& \beta_t - \gamma_t - \frac{1}{2}\alpha_{I(t)} && \le\ 0 \quad (t = 0 \vee t = \Delta) && (17d)\\
& 900\sum_{I\in T_{15}} \alpha_I && =\ 1 && (17e)\\
& \alpha_I && \ge\ 0 \quad (\forall I \in T_{15}) && (17f)\\
& \beta_t && \ge\ 0 \quad (\forall t \in T). && (17g)\\
& \gamma_t && \ge\ 0 \quad (\forall t \in T). && (17h)
\end{aligned}$$

We can directly see that whenever there is a $t \in T$ for which the inequality

$$\left(\sum_{r\in R}\sum_{a=(v,w)\in A^r}\sum_{j\in J^{rv}} \bar{p}_j^{rat}\bar{x}_j^{rv}\right) \le U \tag{18}$$

is violated, the problem is unbounded and its feasible set contains an unbounded ray with $\beta_t = \gamma_t = 1$. Therefore, it is precisely the Benders feasibility cut (18) which cuts of this unbounded ray and removes the corresponding cause of infeasibility in the restricted Benders master problem. Note that all the above considerations can similarly be applied to the timetable formulation in the y-variables to obtain an analogous Benders scheme which can profit from the higher sparsity of (DF). It will indeed be this version of the algorithm which we will use in our computational experiments.

4 A Case Study for German Passenger Traffic

In this section, we present a real-world case study for a significant part of German railway traffic. It demonstrates energy-efficient timetable adaptation based on our optimization models for the complete passenger traffic operated by Deutsche Bahn AG. First, we will describe the problem instances we created based on their data. Then we will compare the computational performance of the models and algorithms we have established to solve the problem. Finally, we will analyse the solutions we have obtained with respect to the reduction in energy costs as well as timetable performance.

All computations have been carried out on a server with Intel Xeon E5-2690v2 3.00 GHz processors and 128 GB RAM, using 5 cores and a specified time limit. We always first state results for a time limit of 10 minutes in order to demonstrate that most instances can be solved satisfactorily within rather short time. For the more difficult instances, we also state the result after 10 hours to show that they can be solved overnight. Our implementation uses the Python-API of Gurobi 8.1.1 (Gurobi Optimization, Inc. (2019)). Note that we always pass the timetable draft to be optimized as an initial solution to the solver as it is feasible for all models under consideration.

4.1 Instance Compilation

For our computational study, we created instances based on real-world data provided by our industry partner Deutsche Bahn AG. They are derived from the German railway passenger timetable for 2015 for the trains operated by DB Regio AG and DB Fernverkehr AG, which represents about 80% of passenger traffic in Germany.

All instances we created contain the traffic between 4 a.m. and 10 p.m. on a typical working day and differ only in the trains selected from the timetable. They mostly describe different regions of Germany and can be grouped into three types according to their scope. The first type are instances where for a given station all trains with at least one trip starting or ending there within the planning interval are considered (both regional and long-distance traffic). We call these the *local* instances. The stations these instances are based on range from smaller stations where no transfer is possible to big hubs in the network. The second type are the *regional* instances, which comprise regional traffic according to the regional subdivisions of DB Regio AG as of 2015. Finally, we consider a set of *national* instances, which are one instance for all regional traffic in Germany (*Regionalverkehr*), one instance for the German long-distance traffic (*Fernverkehr*) and one instance for the complete German passenger traffic with both regional and long-distance traffic combined (*Deutschland*). Altogether, we have 18 local instances, 10 regional instances and 3 national instances, or 31 in total.

For each train and each of its trips in a given instance, we created a representative power profile depending on the type of train, the travel time and the heights of start and end station. To this end, we grouped the trains into four classes, which we name in short after their main representative train type. These classes are the long-distance trains Intercity-Express (ICE), the two regional train types Regional-Express (RE) and Regionalbahn (RB) as well as the urban trains S-Bahn (S). The two classes RB and RE differ in that the former stops at every station, while the latter only stops at somewhat bigger stations. The power profiles we generated take into account typical train characteristics according to this classification and incorporate the necessary acceleration power (based on simplified velocity profiles), the power required to overcome the downhill force (based on the average slope from the start to the end of each trip) as well as the power required to overcome the rolling and the air friction. Note that absent any data on the locomotives used on the individual lines, we have assumed all trains to use electric traction and to be able to recuperate energy. In this sense, the results we present in this study represent a best-case scenario. A typical velocity and power consumption profile for an ICE train as we have used it for our computations is shown in Figure 2.

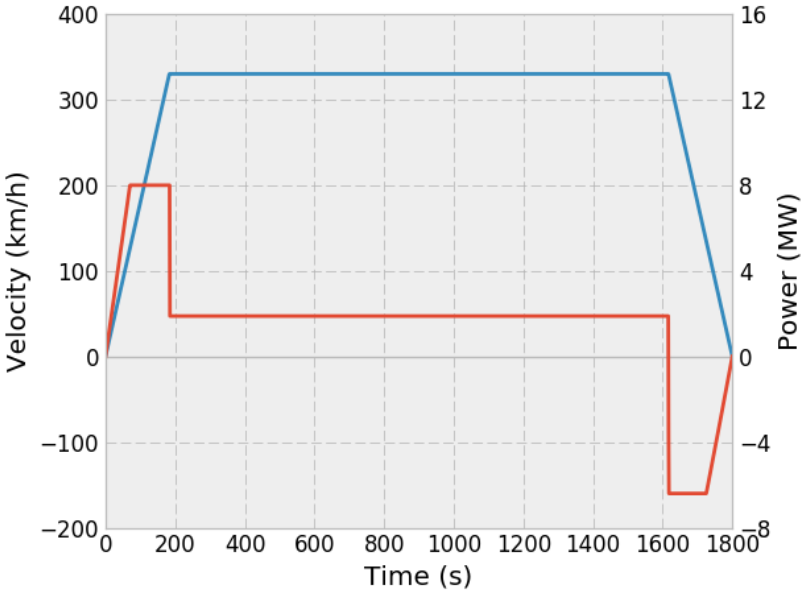

Figure 2: The velocity profile (*blue*) and the power consumption profile (*red*) for an ICE train running on an even track (i.e. start and end station are on the same height) and a journey time of 30 minutes

As we had no data on the physical railway network, we took the assumption that the trains travel on tracks directly connecting the stations they service consecutively. This construction leads to the artefact that regional trains and long-distance trains travel on separate sets of tracks in many cases, as the former have many intermediate stops between large stations which the latter have not. However, we consider our instances as sufficiently accurate to assess the potential of the overall approach. Note that this assumption only pertains to our interpretation of the available data, not to our optimization models. The latter can handle the case of mixed track use if we stipulate that the optimization leave the assignment of trains to tracks untouched; for practical computations this requires, of course, to have an ordered list of all trains passing a certain track during the planning horizon.

In Figure 3, we show the connections between the stations established in the timetable. To find an optimal adjustment of the timetable, we allowed to shift each train departure by up to $\pm$ 3 minutes in steps of full minutes compared to the original departure time. However, we did not allow the departure or arrival of a train to be shifted outside of the considered planning interval. The departure times for trips starting before 4 a.m. and extending into the planning interval as well as those starting in the planning interval and extending past 10 p.m. were considered as fixed. In this case, the part of their power consumption which falls inside the planning interval was added to the consumption of the variable trains as a fixed base load.

The minimum dwell time for each train in each of its station was chosen as the time between arrival and departure in the current timetable. The minimum headway times were taken from (Pachl, 2016, Table 5.4) and rounded up to full minutes. They are displayed in Table 1, subsuming class S under RB.

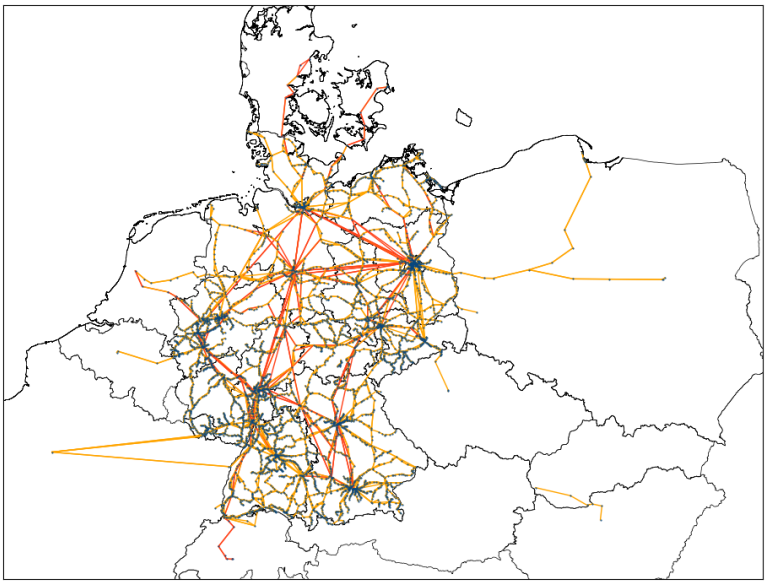

Figure 3: The connections between the railway stations established in the timetable – regional traffic in *light orange* and long-distance traffic in *dark orange*, train stations in *blue*

	ICE	RE	RB
ICE	4	3	3
RE	7	4	4
RB	9	7	4

Table 1: Minimum headway times for each pair of train types in minutes. The columns refer to the train departing earlier, while the columns refer to the following train.

Finally, for a given train arriving at a station, we consider all trains departing between 5 and 15 minutes after this arrival in the current timetable as connecting trains. For such a pairing, we require that the current transfer time be changed by at most $\pm$ 3 minutes and stay between 5 and 15 minutes.

4.2 Computational Results for Problem (TTMAPBR)

As a first step, we compare the performance of the three timetabling formulations introduced in Section 3.1 for solving Problem (TTMAPBR) – i.e. the minimization of maximum average power consumption neglecting regenerative braking, as we consider it to be the point of view of the TOC. Table 2 shows the number of trains and the number of trips to schedule in each instance, the computation time of the instance according to the naive (NA), totally unimodular (TU) and the dual-flow (DF) formulation as well as the achieved savings in energy cost in the optimal solution.

Instance	#Trains	#Trips	Time [s] / Gap [%] NA	TU	DF	Saving [%] DF
Zeil	42	762	11.59 s	2.75 s	**0.76 s**	**14.89%**
Bayreuth Hbf	68	327	0.80 s	0.41 s	**0.21 s**	**22.18%**
Passau	75	1040	366.75 s	124.49 s	**7.91 s**	**14.48%**
Jena Paradies	78	1102	113.79 s	21.94 s	**6.00 s**	**12.46%**
Lichtenfels	113	1650	261.73 s	100.81 s	**8.46 s**	**15.25%**
Erlangen	142	2969	1.02%	272.40 s	**24.22 s**	**15.28%**
Bamberg	209	3644	0.55%	73.49 s	**27.51 s**	**13.08%**
Aschaffenburg	245	3463	77.73 s	3.01 s	**2.74 s**	**12.95%**
Kiel Hbf	297	2130	94.26 s	5.90 s	**2.31 s**	**11.20%**
Leipzig Hbf (tief)	369	6810	3.34 s	3.71 s	**0.57 s**	**6.40%**
Würzburg Hbf	371	4456	0.57%	0.03%	**0.01%**	**8.32%**
Dresden	422	6936	0.74%	442.72 s	**59.89 s**	**9.30%**
Ulm Hbf	468	5729	0.02%	24.19 s	**4.49 s**	**11.23%**
Stuttgart Hbf (tief)	628	11594	1.00%	348.22 s	**11.43 s**	**0.93%**
Berlin Hbf (S-Bahn)	639	16114	19.39 s	23.05 s	**13.15 s**	**2.97%**
Hamburg-Altona(S)	722	12373	0.10%	130.91 s	**10.00 s**	**1.29%**
Frankfurt(Main)Hbf	728	8626	1.06%	71.85 s	**19.13 s**	**9.34%**
Nürnberg	951	12189	22.71%	43.39 s	**5.82 s**	**7.10%**
S-Bahn Hamburg	1208	17533	3.56%	275.16 s	**18.53 s**	**2.36%**
Regio Nord	1476	13379	77.00 s	20.68 s	**7.92 s**	**12.79%**
Regio Nordost	1494	16496	15.56%	28.68 s	**10.05 s**	**15.50%**
Regio Hessen	1547	25092	26.67 s	31.29 s	**10.57 s**	**5.65%**
Regio Südwest	1863	24191	87.99 s	65.58 s	**10.22 s**	**13.00%**
Regio Südost	2357	31917	8.96%	72.39 s	**17.71 s**	**8.96%**
Regio BW	2382	30172	14.30%	287.12 s	**22.19 s**	**13.36%**
S-Bahn Berlin	2578	53353	84.41 s	195.45 s	**83.50 s**	**1.73%**
Regio NRW	2826	47026	26.03%	-	**28.27 s**	**5.13%**
Regio Bayern	3554	49262	519.73 s	94.28 s	**41.86 s**	**10.73%**
Fernverkehr	667	7053	149.87 s	21.56 s	**12.28 s**	**5.38%**
Regionalverkehr	21288	308472	-	-	-	-
Deutschland	21955	315525	5.06%	-	**482.30 s**	**5.05%**

Table 2: Computational results for Problem (TTMAPBR) on our 31 instances, stating the computation times or the gap of the three different formulations for a feasible timetable adjustment and the potential for saving energy costs in a (close to) optimal solution. A dash ('-') indicates hitting the time limit of 10 minutes and no feasible solution and/or bound found. Instance *Regionalverkehr* can be solved to optimality within 1.5 hours via formulation DF, yielding energy savings of 9.53%.

Formulation NA only solves 14 instances to optimality, often leaving large optimality gaps for the unsolved ones. In contrast, TU yields an optimal solution for almost twice as many instances, as the total unimodularity of its underlying timetabling polytope reduces the branching effort. Recall that the overall problem is not totally unimodular as the piecewise-linear objective function to be optimized destroys this property. However, it is visible at first sight that DF outperforms the other two formulations significantly, as it solves all but one instance within the time limit. Using this formulation with its much sparser constraints saves an order of magnitude in computation time on many instances, as the node relaxations in the branch-

and-bound tree are solved much faster.

Instance *Würzburg* resulted to be the most difficult of the smaller instances, which we assume is due to the long-drawn-out phase of high power consumption in the afternoon that makes many of the objective function constraints binding. A striking reduction in solution time can be seen for instance *Nürnberg*, where DF is solved within 6 seconds, while NA still has a gap of 22% after 10 minutes. The reason seemed to be that NA had to cope with numerical instabilities while solving the root relaxation here, unlike the two totally unimodular formulations. Finally, the Germany-wide *Deutschland* instance, arguably the most interesting one, could be solved within about 6 minutes by DF. NA, on the other hand, leaves an optimality gap of 5% after reaching the time limit. With more available computation time, NA is able to solve the instance within 3 hours, which means that DF could reduce the solution time by a factor of 30. To provide some more context, we remark that the *Deutschland* instance could be represented with 380,296 constraints, 173,922 variables and 951,047 non-zero coefficients after preprocessing of formulation DF. Optimization was done almost immediately after solving the corresponding root relaxation. Altogether, this result is very encouraging, as it points to the possibility of giving decision support for nationwide timetabling in close to real time.

When assessing the reduction in peak power consumption, we see that the smaller instances tend to profit more, percentagewise. The most likely explanation is that shifting few departures already has a high impact here. In the larger instances, many individual acceleration phases already average out via the much higher number of trains. This also explains why the solution times of DF only grows moderately when passing to the larger instances: the LP bounds remain very tight (cf. Tables 9 and 10 in the appendix), and it does not require excessive branching on individual departures to close the gap. The reductions which can be obtained for the larger instances are still tremendous, given the high peak consumption charges. For the regional instances, they lie between 5% and 16%, with the exception of those five instances encompassing (parts of) the urban traffic of Stuttgart, Hamburg and Berlin, as the dense schedules there do not provide much room for improvement. The three nationwide instances allow savings between 5% and 10% (when allowing more computation time for instance *Regionalverkehr*). Altogether, Table 2 shows that for all instances peak consumption could be reduced compared to the initial timetable, with improvements ranging from 0.93% to 22.18%.

In the following, we investigate the potential for savings, also financial ones, more closely for some selected instances. Instance *Fernverkehr* contains all ICE and IC trains, which are the most energy demanding ones. The average peak consumption of 335.66 MW for the initial timetable can be reduced to 317.59 MW by our optimization. Applying the official cost factor of 110.02 € per kilowatt peak consumption (see Section 2.2) shows that the responsible TOC (DB Fernverkehr) could save almost 2 million € per year when using the optimized timetables.

Instance *Regionalverkehr* contains all regional trains, which run between closeby cities and villages. Formulation DF was able to reduce the maximum average peak consumption from 601.58 MW to 544.27 MW. This improvement by 57.31 MW would mean cost savings of 6.3 million € for the operating company DB Regio.

Instance *Deutschland*, for which the result is shown in Figure 4, comprises both the trains in *Fernverkehr* and *Regionalverkehr*. As such, it has two about equally high consumption peaks in the morning and in the evening, both of which can be balanced by the optimization. Peak consumption can be reduced by 44.44 MW in this combined instance – less than possible for the individual two instances, due to effects discussed before. Nevertheless, the benefit of a combined optimization of German regional and long-distance traffic could be savings in the order of 5 million € per year.

For the Germany-wide instance, we carried out an additional study where we varied the size of the interval within which the departures can be shifted. The results are given in Table 3.

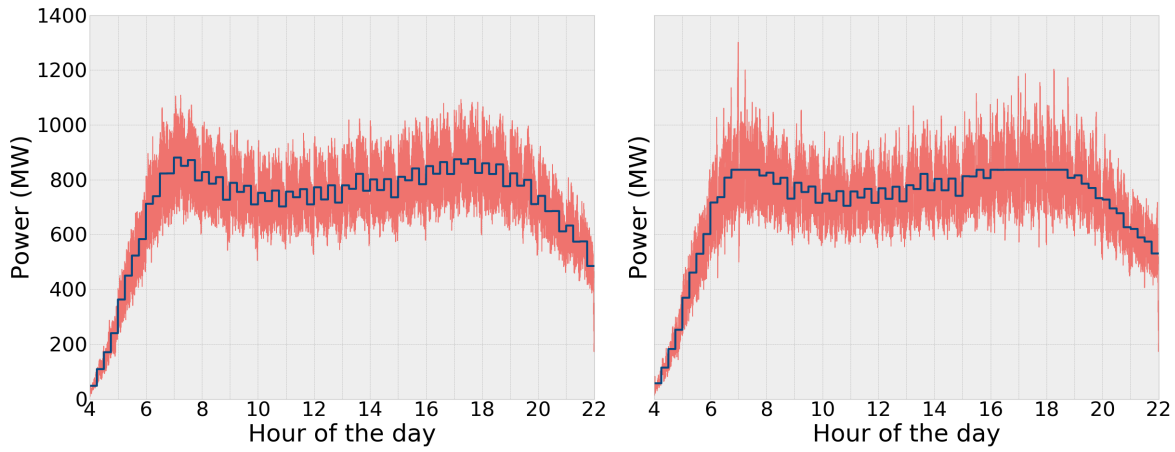

Figure 4: Power consumption profile of instance *Deutschland* before (left) and after (right) optimization according to (TTMAPBR)

Max. shift [min]	Solution time [s]	Optimized [MW]	Abs. red. [MW]	Rel. red. [%]
± 1	425.21	849.20	31.28	3.55
± 2	6,319.82	839.60	40.88	4.64
± 3	482.30	836.04	44.44	5.05
± 5	233.61	833.32	47.16	5.36
± 10	408.07	831.15	49.33	5.60

Table 3: Average peak load reduction according to (TTMAPBR) for instance *Deutschland*. The initial timetable had a maximum average peak consumption of 880.48 MW

Its columns show the maximum allowed shift in each case, the solution time of formulation DF, the peak consumption in the optimized timetable as well as the absolute and relative reduction compared to the initial timetable respectively. We see that shifts of up to only 1 minute already allow for a significant reduction of 3.55 %. With increasing freedom to shift departures, the possible reduction increases to 5.60% for shifts of up to 10 minutes.

For the same instance, we also looked at the effect of the optimization on the timetable performance itself. Table 4 shows for each maximally-allowed shift which percentage of trains are shifted by which absolute amount and how the average departure from and dwell time in a station is affected.

	% of departures shifted by [min]						Average deviation [min]		
Max. shift [min]	0	1	2	3	4–5	6–10	Abs. dep.	Dep.	Dwell
± 1	63	37	-	-	-	-	0.37	0.06	0.08
± 2	60	24	16	-	-	-	0.56	0.08	0.12
± 3	62	20	10	8	-	-	0.63	0.02	0.14
± 5	61	19	9	5	6	-	0.78	0.07	0.16
± 10	61	18	9	5	4	4	0.97	0.00	0.21

Table 4: Statistics for departure time shifts and dwell time increases according to (TTMAPBR) for instance *Deutschland* in an optimal solution

The interesting conclusion is that in all cases, the number of unchanged departures in the timetable stays approximately the same, namely around 60%. The other 40% are drawn-out

over the other the allowed departure intervals. Column *Abs. dep.* shows the absolute average of the number of minutes by which departures are shifted. While a general trend is that less and less trains use the additional freedom for larger intervals, the average shift increases nevertheless. In Column *Dep.* we see the signed averages of the departure time shifts. It becomes visible that about as many trains are shifted to later departure times than are shifted to earlier departure times, independent from the size of the allowed shift interval. Finally, Column *Dwell* contains the average increase in dwell time over all stations (note that it can not decrease, as we use the current dwell time as the minimum dwell time in our computations). We see that average dwell times increase only slightly in all cases. Overall, it becomes visible that shifts of up to 3 minutes are a reasonable choice, as they allow for the realization of most of the potential savings while largely maintaining the initial timetable.

4.3 Computational Results for Problem (TTMAP)

Now we describe our results for Problem (TTMAP). It reduces the highest 15-minute average peak consumption value like (TTMAPBR), but in addition it is able use recuperation energy to increase the effect and even tends to reduce the overall energy consumption in the process. We see the IM as most interested in solving this problem in order to reduce the costs of energy provision.

As DF was vastly superior in solving (TTMAPBR), we only state the results of this formulation here. Table 5 summarizes the solution times and peak reductions for solving (TTMAP) via three different methods after 10 minutes of computation. The first one is a baseline heuristic which solves Model (TTMAPBR) instead of (TTMAP) in order to make use of the short solution times of the former. It simply ignores that the power profiles contain time steps with negative consumption (i.e. recuperation) and uses the simplified reformulation of the objective function in Section 3.2.1 for the derivation of (TTMAPBR). Of course, this leads to an error in the computation of the actual peak power consumption, because interchanging summation and the averaging as done there are not valid when negative consumption values are present. This renders the approach a heuristic only. The second method is the plain solution of (TTMAP) as the MIP presented in Section 3.2.2, and the last one is the Benders decomposition developed in Section 3.2.3.

Instance	Time [s] / Gap [%]			Saving [%]		
	Heuristic	MIP	Benders	Heuristic	MIP	Benders
Zeil	1.91 s	44.46%	**32.99%**	9.26%	20.55%	**36.95%**
Bayreuth Hbf	23.80 s	21.37%	**15.49%**	5.99%	19.29%	**27.02%**
Passau	13.88 s	27.65%	**8.75%**	15.18%	8.13%	**26.69%**
Jena Paradies	201.77 s	0.73%	**0.08%**	8.47%	14.93%	**15.33%**
Lichtenfels	6.67 s	5.29%	**84.24 s**	22.71%	22.37%	**26.31%**
Erlangen	5.36 s	98.23%	**0.01%**	19.34%	0.00%	**25.98%**
Bamberg	284.79 s	20.96%	**0.05%**	16.27%	0.00%	**20.84%**
Aschaffenburg	0.09%	3.42%	**0.16%**	8.90%	13.01%	**15.84%**
Kiel Hbf	63.84 s	99.50%	**0.39%**	17.51%	0.00%	**24.16%**
Leipzig Hbf (tief)	0.69 s	0.57%	**294.47 s**	3.88%	19.96%	**20.12%**
Würzburg Hbf	0.05%	18.53%	**0.08%**	14.79%	0.00%	**18.42%**
Dresden	0.05%	98.30%	**0.89%**	13.37%	0.00%	**22.44%**
Ulm Hbf	0.03%	96.56%	**0.18%**	9.31%	0.00%	**17.41%**
Stuttgart Hbf (tief)	9.96 s	4.37%	**55.55 s**	6.41%	7.54%	**9.88%**
Berlin Hbf (S-Bahn)	12.90 s	**0.25%**	0.86%	2.35%	**9.22%**	8.93%
Hamburg-Altona(S)	10.41 s	11.73%	**0.28%**	3.40%	4.63%	**10.98%**
Frankfurt(Main)Hbf	0.01%	15.47%	**272.46 s**	**15.46%**	0.00%	**15.46%**
Nürnberg	380.17 s	98.45%	**0.02%**	13.94%	0.00%	**14.22%**
S-Bahn Hamburg	19.49 s	16.84%	**0.30%**	9.41%	0.00%	**13.68%**
Regio Nord	0.01%	97.64%	**4.14%**	-15.41%	0.00%	**11.67%**
Regio Nordost	0.02%	99.74%	**5.18%**	4.82%	0.00%	**25.63%**
Regio Hessen	12.44 s	97.77%	**2.08%**	-10.61%	0.00%	**15.54%**
Regio Südwest	0.03%	99.98%	**7.75%**	-8.26%	0.00%	**11.16%**
Regio Südost	126.14 s	99.38%	**5.10%**	3.46%	0.00%	**21.53%**
Regio BW	600.06 s	-	**5.69%**	7.46%	-	**18.23%**
S-Bahn Berlin	87.49 s	-	**1.70%**	-2.13%	-	**4.33%**
Regio NRW	239.64 s	-	**3.85%**	3.74%	-	**15.53%**
Regio Bayern	65.08 s	-	**2.40%**	7.22%	-	**20.35%**
Fernverkehr	0.01%	97.26%	**0.02%**	**10.24%**	0.00%	**10.24%**
Regionalverkehr	-	-	-	-	-	-
Deutschland	-	-	-	-	-	-

Table 5: Computational results for (TTMAP) solved via a heuristic, an MIP and Benders decomposition for a time limit of 10 minutes

As expected, the heuristic had the lowest solution times or gaps on almost all instances. However, it generally led to largely suboptimal solutions, sometimes even much worse than the initial timetable. Only in two cases (*Frankfurt* and *Fernverkehr*) it was able to produce the optimal solution, and for *Nürnberg* a solution at least close to optimal. The high proportion of long-distance traffic in these instances might have helped the heuristic, as overall negative consumptions are not as relevant here.

Among the two exact methods, the Benders decomposition performed much better. It had the smaller solution time or gap on all but one of the local instances and on all of the regional instances. For these two sets of instances the attained reduction in peak consumption was much higher accordingly, very often by a factor of 2 or more. From the national instances, we only obtained an improved solution for instance *Fernverkehr* within 10 minutes. The Benders decomposition almost solved it to optimality, vastly outperforming the pure MIP solution, where the latter was not able to reduce peak consumption.

We repeated the experiment with an extended time limit of 10 hours for all instances where the Benders decomposition left an optimality gap of more than 5% (a reasonable optimality target in large-scale practical problems) as well as for all national instances. The result is shown in Table 6.

	Time [s] / Gap [%]			Saving [%]		
Instance	Heuristic	MIP	Benders	Heuristic	MIP	Benders
Zeil	1.88 s	29.86%	**26.74%**	9.26%	36.53%	**40.04%**
Bayreuth Hbf	23.90 s	**5.98%**	10.77%	5.99%	29.10%	**29.32%**
Passau	13.78 s	6.03%	**4.50%**	15.18%	27.59%	**29.14%**
Regio Nordost	6123.60 s	**0.34%**	1.29%	4.41%	**28.55%**	28.19%
Regio Südwest	3273.60 s	**0.18%**	1.08%	-7.32%	**16.66%**	16.21%
Regio Südost	126.62 s	**0.11%**	0.65%	3.46%	**25.07%**	24.76%
Regio BW	604.91 s	**0.66%**	0.82%	7.46%	21.89%	**21.96%**
Fernverkehr	17,717.13 s	0.10%	**26458.22 s**	**10.24%**	10.16%	**10.24%**
Regionalverkehr	6122.06 s	-	**0.90%**	0.29%	-	**16.36%**
Deutschland	9315.04 s	-	**3483.42 s**	7.67%	-	**9.38%**

Table 6: Computational results for (TTMAP) solved with a heuristic, an MIP and Benders decomposition with a time limit of 10 hours for selected instances

The heuristic does not become much more viable as solution times in the previously unsolved instances are relatively high and solution does not improve much. The MIP and Benders decomposition perform much better than the heuristic and, for the local and regional instances, about equally well among each other. For the national instances, Benders decomposition again shows its strength, solving *Fernverkehr* and *Deutschland* to optimality as well well as *Regionalverkehr* to an optimality gap below 1%. In contrast, MIP solves none of them to optimality and still does not find improved solutions for *Regionalverkehr* and *Deutschland*.

We highlight the effects of the optimization for instance *Deutschland* in Figure 5, based on the globally optimal solution obtained by Benders decomposition. Average peak consumption

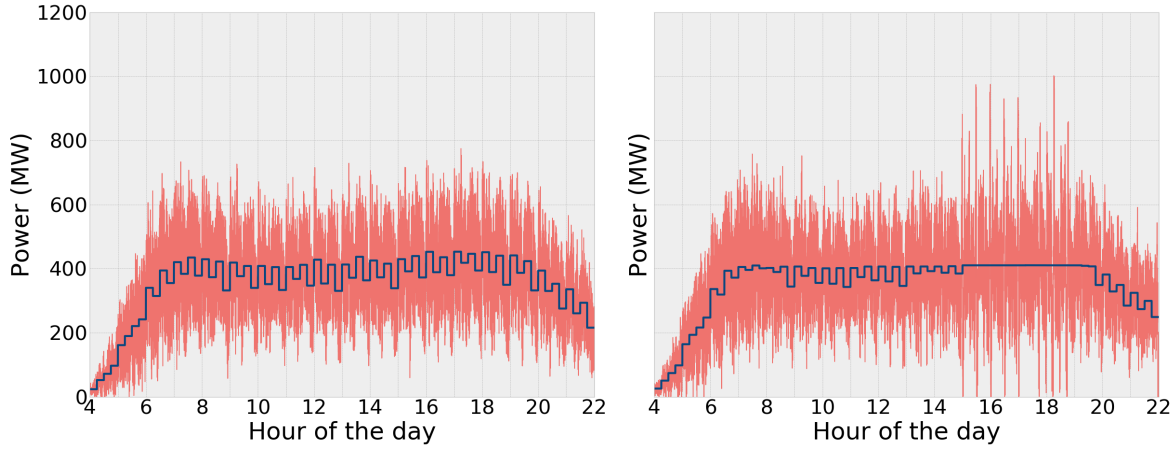

Figure 5: Power consumption profile of instance *Deutschland* before (left) and after (right) optimization according to (TTMAP)

reduced from 452.69 MW to 410.23 MW, which is an improvement of 42.46 MW or over 9%; again considerable savings on a countrywide scale. Although we cannot provide a similar official cost factor as for the TOCs, the savings in energy cost would again be significant. What

we also observe is that instantaneous peaks might be higher in the optimized solution according to Problem (TTMAP). This is why we also study Problem (TTMAPI), which limits these instantaneous peak to the maximum such peak in the initial timetable.

4.4 Computational Results for Problem (TTMAPI)

Finally, we present the results for Problem (TTMAPI), which, in addition to minimizing maximum average peak consumption, also keeps instantaneous peak consumption under control. The heuristic from before is not feasible here, because it cannot guarantee not to exceed the maximum instantaneous peak consumption induced by the initial timetable. Thus, Table 7 only compares the results obtained by MIP and Benders decomposition for a time limit of 10 minutes.

	Time [s] / Gap [%]		Saving [%]	
Instance	MIP	Benders	MIP	Benders
Zeil	44.47%	**32.48%**	20.55%	**36.86%**
Bayreuth Hbf	22.94%	**15.02%**	17.70%	**27.29%**
Passau	33.54%	**10.23%**	0.00%	**25.77%**
Jena Paradies	0.77%	**0.15%**	14.88%	**15.28%**
Lichtenfels	97.34%	**149.15 s**	0.00%	**26.31%**
Erlangen	97.79%	**0.01%**	0.00%	**25.98%**
Bamberg	99.51%	**0.06%**	0.00%	**20.83%**
Aschaffenburg	94.48%	**0.13%**	0.00%	**15.87%**
Kiel Hbf	99.50%	**0.65%**	0.00%	**23.98%**
Leipzig Hbf	0.73%	**0.01%**	19.84%	**20.12%**
Würzburg Hbf	96.84%	**0.13%**	0.00%	**18.36%**
Dresden	98.30%	**1.22%**	0.00%	**22.22%**
Ulm Hbf	95.52%	**0.20%**	0.00%	**17.41%**
Stuttgart Hbf (tief)	4.54%	**65.07 s**	7.48%	**9.88%**
Berlin Hbf (S-Bahn)	**0.10%**	0.74%	**9.30%**	8.87%
Hamburg-Altona(S)	15.98%	**0.36%**	0.00%	**10.99%**
Frankfurt(Main)Hbf	99.41%	**418.39 s**	0.00%	**15.46%**
Nürnberg	98.33%	**0.02%**	0.00%	**14.22%**
S-Bahn Hamburg	97.50%	**2.21%**	0.00%	**13.51%**
Regio Nord	97.64%	**16.46%**	0.00%	**0.06%**
Regio Nordost	99.74%	**16.51%**	0.00%	**16.16%**
Regio Hessen	97.74%	**1.94%**	0.00%	**15.45%**
Regio Südwest	99.98%	**18.38%**	**0.00%**	**0.00%**
Regio Südost	-	**13.98%**	-	**13.84%**
Regio BW	-	**11.79%**	-	**12.70%**
S-Bahn Berlin	-	**1.38%**	-	**4.56%**
Regio NRW	-	**4.03%**	-	**15.38%**
Regio Bayern	-	**7.37%**	-	**16.07%**
Fernverkehr	97.26%	**0.01%**	0.00%	**10.24%**
Regionalverkehr	-	-	-	-
Deutschland	-	-	-	-

Table 7: Computational results for (TTMAPI) solved with an MIP and Benders decomposition with a time limit of 10 minutes

As for (TTMAP), Benders decomposition clearly outperforms MIP, both in solution time /

gap and peak reduction. It is very interesting that the additional constraint in (TTMAPI) only moderately increases the computational complexity. Furthermore, it still allows for the majority of the savings that could be obtained without it.

We again allowed a higher computation time of 10 hours, selecting the instances according to the same criterion as above. The results in Table 8, lead to the same conclusions as before: both methods are equally able to use the additional time to improve the obtained reductions in peak consumption, and Benders is the only method able to solve the national instances (close) to optimality.

Instance	Time [s] / Gap [%] MIP	Benders	Saving [%] MIP	Benders
Zeil	37.11%	**26.05%**	29.56%	**40.85%**
Bayreuth Hbf	**7.83%**	10.22%	28.27%	**29.69%**
Passau	6.01%	**5.82%**	27.74%	**28.29%**
Regio Nord	**0.31%**	2.85%	**14.27%**	14.04%
Regio Nordost	**0.36%**	1.33%	**28.47%**	28.06%
Regio Südwest	**0.19%**	1.05%	**16.57%**	16.10%
Regio Südost	**0.22%**	0.73%	**24.87%**	24.59%
Regio BW	**0.61%**	0.80%	**21.91%**	**21.91%**
Regio Bayern	**0.08%**	0.13%	**21.66%**	**21.66%**
Fernverkehr	0.02%	**2581.79 s**	10.23%	**10.24%**
Regionalverkehr	-	**1.88%**	-	**15.37%**
Deutschland	-	**5132.47 s**	-	**9.32%**

Table 8: Computational results for (TTMAPI) solved with an MIP and Benders decomposition with a time limit of 10 hours for selected instances

We showcase the benefits of the optimization for instance *Deutschland* with all German long-distance and regional passenger traffic operated by Deutsche Bahn. Peaks in consumption are

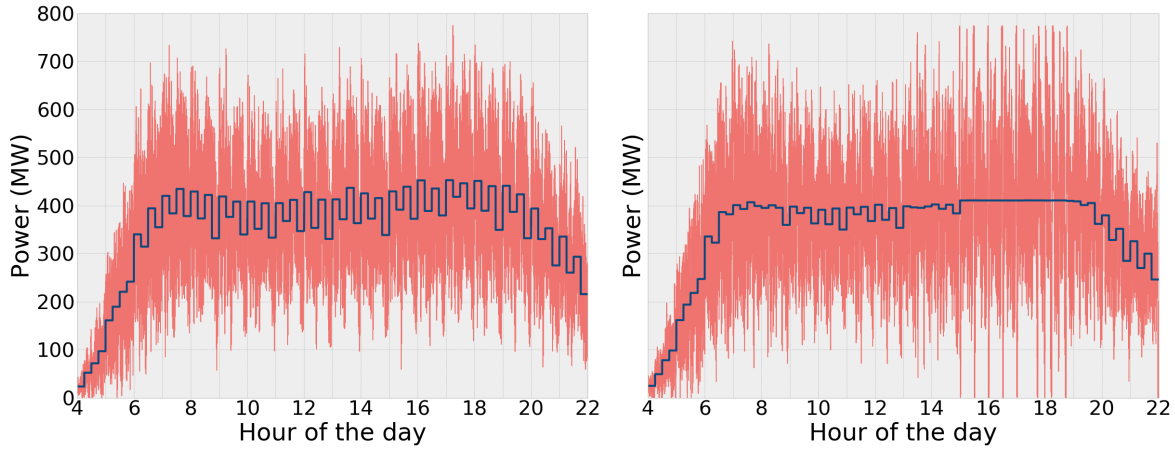

Figure 6: Power consumption profile of instance *Deutschland* before (left) and after (right) optimization according to (TTMAPI) after 10 hours of computation via Benders decompostion

smoothed in the morning and, to a larger extent, in the evening. We obtain reductions in average peak consumption of 42.18 MW or over 9% while not increasing maximum instantaneous peak consumption.

5 Conclusions

We have presented mathematical methods for the optimization of railway timetables according to energy-efficiency, more precisely for reducing maximum average peak consumption by shifting departure times of the trains. Improving on previous results in this field, we are able to solve the problem for large networks of a countrywide scale, enabling also the improved use of recuperated energy and limiting instantaneous peak consumption. This was possible by deriving tight, compact polyhedral formulations for the set of feasible timetable adaptations and by devising an efficient, specially-tailored Benders decomposition scheme. Our results show that high-quality solutions, in many cases even optimal solutions, can be obtained for networks of local, regional and national instances alike. They point to significant savings in energy costs which are achievable for both the train operating companies as well as the infrastructure manager, which makes the approach very promising for use in practice.

References

Albrecht, T. (2010). Reducing power peaks and energy consumption in rail transit systems by simultaneous train running time control. In Pilo, E., editor, *Power Supply, Energy Management and Catenary Problems*. WIT Press.

Bärmann, A., Gellermann, T., Merkert, M., and Schneider, O. (2018). Staircase compatibility and its applications in scheduling and piecewise linearization. *Discrete Optimization*, 29:111–132.

Bärmann, A., Gemander, P., and Merkert, M. (2020). The clique problem with multiple-choice constraints under a cycle-free dependency graph. *Applied Discrete Mathematics*. to appear.

Bärmann, A., Martin, A., and Schneider, O. (2017). A comparison of performance metrics for balancing the power consumption of trains in a railway network by slight timetable adaptation. *Public Transport*, 9(1-2):95–113.

Cacchiani, V. and Toth, P. (2012). Nominal and robust train timetabling problems. *European Journal of Operational Research*, 219:727–737.

Canca, D. and Zarzo, A. (2017). Design of energy-efficient timetables in two-way railway rapid transit lines. *Transportation Research Part B: Methodological*.

Chen, J.-F., Lin, R.-L., and Liu, Y.-C. (2005). Optimization of an mrt train schedule: Reducing maximum traction power by using genetic algorithms. *IEEE Transactions on Power Systems*, 20(3):1366–1372.

E-Motion (2020). Project E-Motion: Energy-efficient Mobility. `https://edom.mi.uni-erlangen.de/e-motion`.

Feng, X., Zhang, H., Ding, Y., Liu, Z., Peng, H., and Xu, B. (2013). A review study on traction energy saving of rail transport. *Discrete Dynamics in Nature and Society*, 2013. Article ID 156548, doi:10.1155/2013/156548.

Fournier, D., Mulard, D., and Fages, F. (2012). Energy optimization of metro timetables: A hybrid approach. In *Proceedings of the 18th International Conference on Principles and Practice of Constraint Programming*, pages 8–12.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, New York.

Gong, C., Zhang, S., Zhang, F., Jiang, J., and Wang, X. (2014). An integrated energy-efficient operation methodology for metro systems based on a real case of shanghau metro line one. *Energies*, 7(11):7305–7329.

Gurobi Optimization, Inc. (2019). Gurobi optimizer reference manual. `http://www.gurobi.com`.

Hasegawa, D., Nicholson, G. L., Roberts, C., and Schmid, F. (2014). The impact of different maximum speed on journey times, energy use, headway times and the number of trains required for phase one of Britain's high speed two line. In Brebbia, C. A., editor, *Computers in Railways XIV*, pages 485–496.

Kim, K. M., Kim, K. T., and Han, M. S. (2011). A model and approaches for synchronized energy saving in timetabling. In *9th World Congress on Railway Research*.

Kim, K. M., Oh, S. M., and Han, M. S. (2010). A mathematical approach for reducing the maximum traction energy: The case of Korean MRT trains. In *International MultiConference of Engineers and Computer Scientists*, pages 2169–2173.

Kimura, N. . and Miyatake, M. (2014). Strategy of speed restriction allowing extended running times to minimize energy consumption and passenger disutility. In Brebbia, C. A., editor, *Computers in Railways XIV*, pages 733–743.

Li, X. and Lo, H. K. (2014). Energy minimization in dynamic train scheduling and control for metro rail operations. *Transportation Research Part B: Methodological*, 70:269–284.

Liers, F. and Merkert, M. (2016). Structural investigation of piecewise linearized network flow problems. *SIAM Journal on Optimization*, 26(4):2863–2886.

Lorenz, S., Hesse, M., and Fischer, A. (2012). Simulation and optimization of robot driven production systems for peak-load reduction. In Laroque, C., Himmelspach, J., Pasupathy, R., Rose, O., and Uhrmacher, A. M., editors, *Proceedings of the 2012 Winter Simulation Conference*, pages 2875–2886.

Luan, X., Wang, Y., Schutter, B. D., Meng, L., Lodewijks, G., and Corman, F. (2018a). Integration of real-time traffic management and train control for rail networks - part 1: Optimization problems and solution approaches. *Transportation Research Part B: Methodological*, 115:41–71.

Luan, X., Wang, Y., Schutter, B. D., Meng, L., Lodewijks, G., and Corman, F. (2018b). Integration of real-time traffic management and train control for rail networks - part 2: Extensions towards energy-efficient train operations. *Transportation Research Part B: Methodological*, 115:72–94.

Mo, P., Yang, L., Wang, Y., and Qi, J. (2019). A flexible metro train scheduling approach to minimize energy cost and passenger waiting time. *Computers & Industrial Engineering*, 132:412–432.

Nelson, K. F., Uhan, A., Zhao, F., and Sutherland, J. W. (2013). Flow shop scheduling with peak power consumption constraints. *Annals of Operations Research*, 206:115–145.

Pachl, J. (2016). *Systemtechnik des Schienenverkehrs: Bahnbetrieb planen, steuern und sichern*. Springer Vieweg.

Peña-Alcaraz, M., Fernández, A., Cucala, A. P., Ramos, A., and Pecharromán, R. R. (2011). Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226(4):397–408.

Ragunathan, A. U., Wada, T., Ueda, K., and Takahasi, S. (2014). Minimizing energy consumption in railways by voltage control on substations. In Brebbia, C. A., editor, *Computers in Railways XIV*, pages 697–708.

Sansó, B. and Girard, P. (1997). Instantaneous power peak reduction and train scheduling desynchronization in subway systems. *Transportation Science*, 31(4):312–323.

Scheepmaker, G. M., Goverde, R. M. P., and Kroon, L. G. (2017). Review of energy-efficient train control and scheduling. *European Journal on Operational Research*, 257:355–376.

Schwindt, C. and Zimmermann, J., editors (2015). *Handbook on Project Scheduling (Vol. 1 + Vol. 2)*. Springer.

Su, S., Li, X., Tang, T., and Gao, Z. (2013). A subway train timetable optimization approach based on energy-efficient operation strategy. *IEEE Transactions on Intelligent Transportation Systems*, 14(2):883–893.

Vielma, J. P., Ahmed, S., and Nemhauser, G. (2010). Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions. *Operations Research*, 58(2):303–315.

Wang, P. and Goverde, R. M. P. (2019). Multi-train trajectory optimization for energy-efficient timetabling. *European Journal of Operational Research*, 272(2):621–635.

Yin, J., Tang, T., Yang, L., Gao, Z., and Ran, B. (2016). Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach. *Transportation Research Part B: Methodological*, pages 178–210.

Yin, J., Tang, T., Yang, L., Xun, J., Huang, Y., and Gao, Z. (2017a). Research and development of automatic train operation for railway transportation systems: A survey. *Transportation Research Part C: Emerging Technologies*, 85:548–572.

Yin, J., Yang, L., Tang, T., Gao, Z., and Ran, B. (2017b). Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. *Transportation Research Part B: Methodological*.

Zhou, L., Tong, L. C., Chen, J., Tang, J., and Zhou, X. (2017). Joint optimization of high-speed train timetables and speed profiles: A unified modeling approach using space-time-speed grid networks. *Transportation Research Part B: Methodological*, 97:157–181.

Zhu, X. (2014). Personal communication.

A Further Computational Statistics

In this appendix, we give some further computational statistics on the solution process of our models and algorithms as well as results for varying the upper bound on instantaneous peak consumption in Problem (TTMAPI).

Tables 9 and 10 detail the computational behaviour of formulations NA, TU and DF for Problem (TTMAPBR), distinguishable by Column *Form.*. Column *BB* shows the number of branch-of-bounds nodes solved while Column *Root-IP gap* shows the gap between the value of the root LP relaxation and the best integer solution found within a time limit of 10 minutes in per cent. Column *Next sol.* shows the time which passed until Gurobi finds a solution improving upon the initial timetable, which was passed to Gurobi as the first incumbent. Column *Next*

sav. then shows the savings achieved with this improved solution, and, finally, Column *Best sav.* shows the best savings achieved by the best solution found within 10 minutes.

We see that the gap between the LP relaxation and the best solution found within 10 minutes by the two totally unimodular formulations TU and DF is always smaller than the produced by NA. On several of the regional and national instances, (*Regio Nordost*, *Regio Südost*, *Regio BW*, *S-Bahn Hamburg* and *Deutschland*), a significant gap remains when using formulation NA, whereas TU and DF close the gap (almost) completely. Furthermore, it becomes apparent the next better found solution for all formulations already brings large savings in most cases and is almost always already close to the savings in the optimal solution. However, we see that the totally unimodular formulations generally produce these improved solutions much faster in comparison to NA, with DF being far superior. The same holds for the number of solved branch-and-bound nodes. Note that the sparsity of DF gives it an additional edge in both respects, as Gurobi's heuristics take less time to execute and the LP relaxations can be solved much faster. In summary, these statistics explain very well why DF performed best in our computations.

Instance	Form.	#BB	Root-IP gap [%]	Next sol. [s]	Next sav. [%]	Best sav. [%]
Zeil	NA	1,297	0.33	0	7.57	14.89
Zeil	TU	19	0.32	0	6.80	14.89
Zeil	DF	14	0.32	0	11.04	14.89
Bayreuth Hbf	NA	1	0.65	0	18.24	22.18
Bayreuth Hbf	TU	31	0.65	0	18.99	22.18
Bayreuth Hbf	DF	1	0.65	0	19.71	22.18
Passau	NA	40,638	1.57	0	6.54	14.48
Passau	TU	65,317	0.30	0	10.95	14.48
Passau	DF	4,806	0.30	0	11.82	14.48
Jena Paradies	NA	12,068	0.56	2	6.90	12.46
Jena Paradies	TU	5,294	0.13	0	10.98	12.46
Jena Paradies	DF	2,887	0.13	0	10.94	12.46
Lichtenfels	NA	14,750	0.90	1	9.60	15.25
Lichtenfels	TU	10,117	0.29	1	13.76	15.25
Lichtenfels	DF	2,016	0.29	0	14.34	15.25
Erlangen	NA	1,384	2.45	10	7.13	14.83
Erlangen	TU	21,091	0.18	6	14.29	15.28
Erlangen	DF	3,523	0.18	2	14.90	15.28
Bamberg	NA	972	0.89	6	4.95	12.97
Bamberg	TU	5,203	0.05	4	11.46	13.07
Bamberg	DF	1,496	0.05	1	12.69	13.08
Aschaffenburg	NA	6,233	0.02	5	11.51	12.95
Aschaffenburg	TU	1	0.01	2	12.54	12.95
Aschaffenburg	DF	1	0.01	1	12.40	12.95
Kiel Hbf	NA	3,440	0.07	2	4.20	11.20
Kiel Hbf	TU	1	0.01	2	10.54	11.19
Kiel Hbf	DF	1	0.01	0	10.84	11.20
Leipzig Hbf (tief)	NA	1	2.25	3	5.13	6.40
Leipzig Hbf (tief)	TU	1	0.69	3	5.13	6.40
Leipzig Hbf (tief)	DF	1	0.58	0	5.13	6.40
Würzburg Hbf	NA	1,840	0.60	16	0.97	7.85
Würzburg Hbf	TU	12,089	0.03	8	8.08	8.31
Würzburg Hbf	DF	34,801	0.02	3	8.00	8.32
Dresden	NA	1	0.82	123	6.35	8.81
Dresden	TU	21,596	0.06	30	8.48	9.29
Dresden	DF	5,330	0.05	3	8.48	9.30
Ulm Hbf	NA	26	0.02	26	5.95	11.22
Ulm Hbf	TU	1	0.01	7	10.77	11.23
Ulm Hbf	DF	1	0.01	3	11.03	11.23
Stuttgart Hbf (tief)	NA	1	1.10	538	0.70	0.93
Stuttgart Hbf (tief)	TU	75	0.13	121	0.66	0.93
Stuttgart Hbf (tief)	DF	23	0.13	9	0.80	0.93
Berlin Hbf (S-Bahn)	NA	0	0.00	19	2.97	2.97
Berlin Hbf (S-Bahn)	TU	0	0.00	23	2.97	2.97
Berlin Hbf (S-Bahn)	DF	0	0.00	13	2.97	2.97
Hamburg-Altona(S)	NA	1	1.73	533	0.44	1.29
Hamburg-Altona(S)	TU	1	0.06	129	1.25	1.29
Hamburg-Altona(S)	DF	1	0.06	9	1.25	1.29
Frankfurt(Main)Hbf	NA	81	1.14	136	4.10	9.34
Frankfurt(Main)Hbf	TU	1,213	0.01	23	10.09	10.28
Frankfurt(Main)Hbf	DF	32	0.01	5	10.09	10.28
Nürnberg	NA	0	-	-	-	-
Nürnberg	TU	1	0.01	43	7.10	7.10
Nürnberg	DF	1	0.01	5	6.92	7.10

Table 9: Statistics on the solution process for the local instances for the different formulations of Problem (TTMAPBR). A dash ('-') indicates hitting the time limit of 10 minutes and no feasible solution and/or bound found.

Instance	Form.	#BB	Root-IP gap [%]	Next sol. [s]	Next sav. [%]	Best sav. [%]
S-Bahn Hamburg	NA	0	4.47	-	-	-
S-Bahn Hamburg	TU	1	0.03	273	2.28	2.36
S-Bahn Hamburg	DF	1	0.04	18	2.27	2.36
Regio Nord	NA	1	0.12	68	10.94	12.79
Regio Nord	TU	1	0.01	20	12.78	12.79
Regio Nord	DF	1	0.01	7	12.43	12.79
Regio Nordost	NA	47	15.59	-	-	-
Regio Nordost	TU	1	0.01	28	15.50	15.50
Regio Nordost	DF	1	0.01	9	15.39	15.50
Regio Hessen	NA	1	0.01	26	5.64	5.64
Regio Hessen	TU	1	0.01	31	5.12	5.65
Regio Hessen	DF	1	0.01	10	5.61	5.64
Regio Südwest	NA	1	0.06	85	12.98	13.00
Regio Südwest	TU	1	0.01	65	12.95	13.00
Regio Südwest	DF	1	0.01	10	12.88	13.00
Regio Südost	NA	5,930	8.96	-	-	-
Regio Südost	TU	1	0.00	72	8.81	8.96
Regio Südost	DF	1	0.01	17	8.86	8.95
Regio BW	NA	0	14.30	-	-	-
Regio BW	TU	1	0.01	286	13.09	13.35
Regio BW	DF	1	0.00	22	13.34	13.36
S-Bahn Berlin	NA	1	0.01	84	1.30	1.73
S-Bahn Berlin	TU	1	0.01	195	1.55	1.73
S-Bahn Berlin	DF	1	0.01	83	1.67	1.73
Regio NRW	NA	0	-	-	-	-
Regio NRW	TU	0	-	-	-	-
Regio NRW	DF	1	0.01	28	5.08	5.13
Regio Bayern	NA	1	0.21	513	10.65	10.73
Regio Bayern	TU	1	0.01	94	10.72	10.72
Regio Bayern	DF	1	0.01	41	10.60	10.72
Fernverkehr	NA	1	0.01	59	2.93	5.38
Fernverkehr	TU	1	0.01	17	5.38	5.38
Fernverkehr	DF	1	0.01	11	5.24	5.38
Regionalverkehr	NA	0	-	-	-	-
Regionalverkehr	TU	0	-	-	-	-
Regionalverkehr	DF	0	-	-	-	-
Deutschland	NA	0	5.06	-	-	-
Deutschland	TU	0	-	-	-	-
Deutschland	DF	1	0.01	481	5.02	5.05

Table 10: Statistics on the solution process for the regional and national instances for the different formulations of Problem (TTMAPBR). A dash ('-') indicates hitting the time limit of 10 minutes and no feasible solution and/or bound found.

Tables 11 and 12 give the same statistics as before for Problem (TTMAPI), comparing methods MIP and Benders over a time limit of 10 hours. The results can be seen as exemplary for (TTMAP) as well. We see that Benders can always produce the first improved solution faster than MIP, often by one or two orders of magnitude. For the local instances, Benders finds superior first improving solutions in addition. On the other hand, the first improving solution found by MIP for the regional instances is usually much better than the one found by Benders. For instances *Regionalverkehr* and *Deutschland*, we see that only Benders is able to produce an improving solution at all. In general, Benders can afford to solve many more branch-and-bound nodes within the time limit due to the reduced problem size. For the local instances, this leads to a better overall performance. For the regional instances, MIP takes a long time to solve the root LP relaxation, but then the information gained there is superior to that obtained by Benders, and better first improving solutions are found. However, for instances *Regionalverkehr*

and *Deutschland*, MIP cannot solve the root relaxation within 10 hours, while Benders performs much better. Altogether, the reduction of problem size by Benders enables us to find very good LP bounds and solutions early on, while after 10 hours the MIP is able to catch up on all but the largest two instances.

Instance	Method	#BB	Root-IP gap [%]	Next sol. [s]	Next sav. [%]	Best sav. [%]
Zeil	MIP	149	37.42	257	7.93	29.56
Zeil	Benders	1,076,763	29.98	2	13.96	40.85
Bayreuth Hbf	MIP	18,549	12.42	49	15.92	28.27
Bayreuth Hbf	Benders	3,616,285	16.15	2	7.23	29.69
Passau	MIP	9,372	8.02	674	8.13	27.74
Passau	Benders	966,406	9.71	18	10.02	28.29
Jena Paradies	MIP	41,932	0.41	307	12.17	15.23
Jena Paradies	Benders	1,461,635	0.43	1	13.52	15.32
Lichtenfels	MIP	3,970	0.29	1,347	21.72	26.31
Lichtenfels	Benders	37,287	0.49	1	18.91	26.31
Erlangen	MIP	4,746	0.37	1,552	21.90	25.98
Erlangen	Benders	1,337,382	0.50	6	22.32	25.98
Bamberg	MIP	42,640	0.15	2,857	18.26	20.84
Bamberg	Benders	1,356,763	0.22	4	17.23	20.85
Aschaffenburg	MIP	22,916	0.15	5,710	13.01	15.85
Aschaffenburg	Benders	2,115,654	0.17	11	13.91	15.90
Kiel Hbf	MIP	23,397	0.63	2,772	14.12	24.07
Kiel Hbf	Benders	3,112,494	0.57	8	21.01	24.24
Leipzig Hbf (tief)	MIP	6,422	1.08	169	10.77	20.12
Leipzig Hbf (tief)	Benders	34,916,658	3.04	5	5.71	20.12
Würzburg Hbf	MIP	13,659	0.22	8,174	15.83	18.33
Würzburg Hbf	Benders	1,283,635	0.24	15	17.09	18.37
Dresden	MIP	5,780	0.59	11,717	18.79	22.64
Dresden	Benders	556,324	0.84	33	12.19	22.80
Ulm Hbf	MIP	7,687	0.15	15,153	13.30	17.45
Ulm Hbf	Benders	1,264,683	0.13	11	14.13	17.52
Stuttgart Hbf (tief)	MIP	363	2.00	596	7.48	9.88
Stuttgart Hbf (tief)	Benders	37	3.05	18	6.54	9.88
Berlin Hbf (S-Bahn)	MIP	15,272	0.42	103	7.74	9.30
Berlin Hbf (S-Bahn)	Benders	22,488,370	4.50	20	2.10	9.30
Hamburg-Altona(S)	MIP	3,811	5.61	631	4.39	10.99
Hamburg-Altona(S)	Benders	1,461,544	9.80	19	3.32	10.99
Frankfurt(Main)Hbf	MIP	4,034	0.01	2,600	15.13	15.46
Frankfurt(Main)Hbf	Benders	6,230	0.01	30	15.43	15.46
Nürnberg	MIP	23,062	0.04	3,696	13.18	14.21
Nürnberg	Benders	394,527	0.03	31	13.49	14.22

Table 11: Statistics on the solution process for the regional and national instances for the different formulations of Problem (TTMAPBR) over a time limit of 10 hours

Instance	Method	#BB	Root-IP gap [%]	Next sol. [s]	Next sav. [%]	Best sav. [%]
S-Bahn Hamburg	MIP	1,428	3.63	1,321	10.30	13.70
S-Bahn Hamburg	Benders	71,830	7.51	63	11.01	13.70
Regio Nord	MIP	5,854	0.31	5,262	11.95	14.27
Regio Nord	Benders	260,334	2.98	283	0.06	14.04
Regio Nordost	MIP	5,863	0.36	7,093	28.17	28.47
Regio Nordost	Benders	91,600	2.87	357	0.83	28.06
Regio Hessen	MIP	23,000	0.20	1,789	13.86	16.52
Regio Hessen	Benders	358,009	1.94	85	5.11	16.44
Regio Südwest	MIP	5,451	0.19	6,944	14.38	16.57
Regio Südwest	Benders	22,069	3.13	663	2.78	16.10
Regio Südost	MIP	1,045	0.22	10,081	23.99	24.87
Regio Südost	Benders	18,994	1.71	362	5.22	24.59
Regio BW	MIP	5	0.61	10,854	19.71	21.91
Regio BW	Benders	9,693	1.40	191	5.64	21.91
S-Bahn Berlin	MIP	75,450	0.12	1,192	4.46	5.07
S-Bahn Berlin	Benders	1,077,974	8.31	192	1.13	5.04
Regio NRW	MIP	5,972	0.29	9,786	14.14	18.18
Regio NRW	Benders	340,403	1.06	178	1.98	18.08
Regio Bayern	MIP	3,540	0.08	13,456	20.88	21.66
Regio Bayern	Benders	44,497	1.68	147	4.63	21.66
Fernverkehr	MIP	5,914	0.02	11,572	9.61	10.23
Fernverkehr	Benders	31,656	0.01	118	10.22	10.24
Regionalverkehr	MIP	0	-	-	-	-
Regionalverkehr	Benders	5,803	3.28	6,504	3.66	15.37
Deutschland	MIP	0	-	-	-	-
Deutschland	Benders	1	0.09	3,082	4.22	9.32

Table 12: Statistics on the solution process for the regional and national instances for the different formulations of Problem (TTMAPBR). A dash ('-') indicates hitting the time limit of 10 hours and no feasible solution and/or bound found.

For instance *Deutschland*, we finally present an additional study with different values for the allowable upper bound on instantaneous peak power consumption, see Table 13. Here, we took the largest power consumption in any second of the planning horizon, which was 774.42 MW, and multiplied it by different factors (see Column *Factor*) to obtain the value for U. The resulting maximally allowed instantaneous power consumption is shown in Column *Max U*. In Column *Max U solution*, we compare it to the maximum instantaneous power consumption actually occurring in the resulting solution. Column *Savings* shows the achieved reduction in maximum average peak consumption compared to the initial timetable, and Column *Time* shows the solution time needed by Benders. It becomes clear that it is possible to reduce the maximum instantaneous power consumption significantly while only loosing little in the savings with respect to maximum average power consumption. We also see that with smaller values for U, it becomes more difficult to solve the resulting instances. Our experiments show that it becomes increasingly difficult to find a first feasible solution, which is to be expected, as the search space shrinks considerably. We suspect that for even smaller factors than 0.87, the problem is infeasible.

Figure 7 shows that, indeed, both aims of reducing maximum peak consumption and instantaneous consumption can be realized in Germany-wide railway traffic. The depicted solution reduces the former by 9.16% and the latter by 13%.

Instance	Factor	Max U [MW]	Max. U solution [MW]	Savings [%]	Time [s]
Deutschland	0.87	673.75	673.71	9.16	19,339
Deutschland	0.88	681.49	681.46	9.18	15,593
Deutschland	0.90	696.98	696.97	9.22	12,627
Deutschland	1.00	774.42	774.39	9.32	5,132
Deutschland	1.10	851.87	851.75	9.36	4,216
Deutschland	1.20	929.31	928.95	9.38	3,679
Deutschland	no limit	-	1,001.91	9.38	3,483

Table 13: Statistics for the solutions of Problem (TTMAPI) for instance *Deutschland* for different values of U using Benders decomposition.

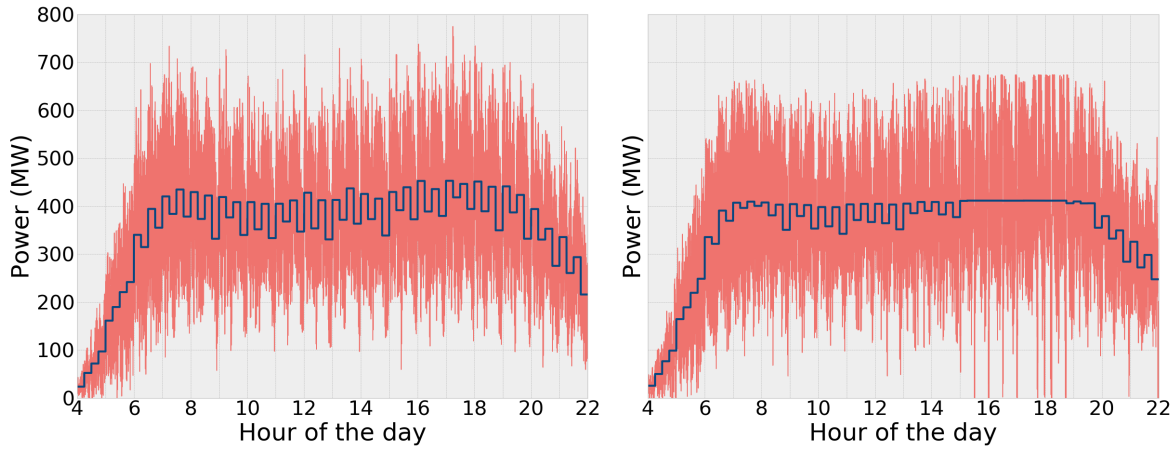

Figure 7: Power consumption profile of instance *Deutschland* before (left) and after (right) optimization according to (TTMAPI) with a peak reduction factor of 0.87 after 10 hours of computation via Benders decomposition

Acknowledgements

We gratefully acknowledge the computing resources provided by the group of Michael Jünger at the University of Cologne. In particular, we thank Thomas Lange for his technical support. We also thank Xiaojun Zhu for sharing his insights on the complexity of the timetable optimization problems and appreciate very much our fruitful discussions with Rodrigo Alexander Castro Campos, Sergio Luis Pérez Pérez, Gualberto Vazquez Casas as well as Francisco Javier Zaragoza Martínez on the decomposition approach presented here. Furthermore, we acknowledge financial support by the German Ministry of Education and Research (BMBF) under grant 05M13WEE (see E-Motion (2020) for the homepage of the corresponding project *E-Motion*). Moreover, we acknowledge financial support by the Bavarian Ministry of Economic Affairs, Regional Development and Energy through the Center for Analytics – Data – Applications (ADA-Center) within the framework of "BAYERN DIGITAL II". Last but not least, we thank the anonymous referees for their helpful comments.