

Traffic signal optimization: combining static and dynamic models

Ekkehard Köhler

Martin Strehler

Brandenburg University of Technology, Mathematical Institute,
P.O. box 10 13 44, 03013 Cottbus, Germany
{ekkehard.koehler,martin.strehler}@b-tu.de

March 22, 2018

In this paper, we present a cyclically time-expanded network model for simultaneous optimization of traffic assignment and traffic signal parameters, in particular offsets, split times, and phase orders. Since travel times are of great importance for developing realistic solutions for traffic assignment and traffic signal coordination in urban road networks, we perform an extensive analysis of the model. We show that a linear time-expanded model can reproduce realistic travel times especially for use with traffic signals and we verify this by simulation. Furthermore, we show how exact mathematical programming techniques can be used for optimizing the control of traffic signals. We provide computational results for real world instances and demonstrate the capabilities of the cyclically time-expanded by simulation results obtained with state-of-the-art traffic simulation tools.

1 Motivation

Traffic signals can be seen as a backbone in the control of traffic flows in urban areas. Since the appearance of the first signalized intersections over 100 years ago, the improvement of signal control strategies has been an important subject for research. Friedrich [13] and Papageorgiou et al. [24] provide overviews of the main research lines over the last years.

Recently, we presented a model for traffic signal coordination based on cyclically time-expanded networks (see [20, 21, 32]). In contrast to most of the previous mathematical approaches, we focused not only on optimizing the offsets of signals, but also optimizing the implied traffic assignment. Each change in the signal parameters may influence the travel times in the network. Consequently, road users will quickly adapt to a new signal coordination and they will switch to faster routes if they suffer from the last intervention

in the signal plans. The traffic assignment, i.e., the distribution of traffic in the road network, may change significantly.

Contrary to previous practical approaches based on non-linear models and heuristics, we focus on a model which allows for the use of exact programming techniques. Hence, proving optimality of a solution or bounding the gap between primal and dual solution was given a higher priority than modeling every effect of real traffic. Thus, this paper is particularly addressing the underlying theoretical questions of traffic signal optimization and the corresponding hardness.

In this paper, we present an extensive analysis of the cyclically time-expanded model. It is a hybrid between a static, i.e., time independent, and a dynamic, i.e., time dependent, network flow approach. Although this time-expanded model is based on a linear program, it can reproduce very realistic non-linear flow-dependent travel times for urban traffic networks with signal control. In particular, both typical convex link performance functions and the time-dependent behavior of traffic, necessary for traffic signal coordination, are realistically mapped in this model. Additionally, the propagation of platoons of cars that is essential for the minimization of waiting times in signal controlled road network is modeled. This is achieved by a lifting method. Whereas previous non-linear approaches use only a small number of variables to describe flow and travel times on a single road segment, we use a linear but high-dimensional model. Moreover, we do not use the common approach of linearizing an existing non-linear model. Instead, our linear model is built from scratch and it is compared to other approaches afterwards.

Related Work Traffic signal optimization is studied since the 1920s. Several models and methods have been proposed. Focussing only on the most theory-driven approaches, we refer to , e.g., Gartner et al. [15], Serafini and Ukovich [27] or Köhler et al. [18]. These approaches rely on abstract models, but exact solution techniques can be applied. However, these approaches do not consider feedback to the traffic assignment. Other approaches use more sophisticated models, but afterwards heuristic algorithms have to be applied to carry out the optimization of signal settings. The most well-known example is the TRANSYT package, which is developed since in the 1960s beginning with the work of Robertson [25]. TRANSYT is based on an accurate traffic simulation and a genetic algorithm is used to optimize signal parameters.

Traffic assignment on its own is also a well studied problem. There are strict mathematical approaches using network flow theory. This branch of research started with the seminal work of Ford and Fulkerson [11] in 1956 and recent result especially focus on various aspects of traffic, e.g., a new concept of fairness in networks with congestion (Jahn, Möhring, Schulz, & Stier Moses, see [17]) and flows over time with load dependent transit times (Köhler & Skutella, see [19]). More application-oriented concepts include dynamic traffic assignment (DTA) techniques, see, e.g., Szeto & Lo [33] and Chiu et al. [7], and simulation based solutions (Nagel & Flötteröd, see [23]).

In contrast, the combined problem of traffic signal optimization and traffic assignment is rarely studied. Allsop and Charlesworth [2] recognized the feedback between an optimized coordination and traffic assignment. They proposed an iterative approach

where signal timings are optimized with TRANSYT. Afterwards, an equilibrium traffic assignment is computed. These steps are repeated until no change in the coordination occurs. Using heuristic methods iteratively, one may only hope to derive a local optimum. Even worse, this approach may lead to a decline of network performance as shown by Dickson [10]. Despite the strong interaction of signal coordination and traffic assignment, only some research was done in the area of integrated optimization of these two aspects of traffic. For example, Chiou [6] presented a bilevel formulation based on the approach of Allsop and Charlesworth. Smith [30] also suggested a bilevel optimization, where the equilibrium property of the flow is preserved during the iterations. Recent results were also made by Bell and Ceylan [3] as well as Teklu et al. [34] using genetic programming. Van den Berg et al. [4] proposed a hybrid approach using both methods from optimal control and mixed integer linear programming. However, in a very recent paper, Smith [31] summarizes the situation as follows: 'At the moment, in practice, traffic signal timings are designed or optimised without systematically seeking to influence route choices beneficially.'

This short overview also reveals one of the main conflicts in traffic optimization. On the one hand, the practitioner favors very realistic and dynamic traffic models. However, these sophisticated models do usually not allow for strict mathematical optimization techniques for solving instances of relevant size. Instead, heuristics, e.g., genetic algorithms or line search strategies in non-convex settings, have to be used. Yet, they are very sensitive to local optima and neither provide guarantees on the gap towards the optimal solution nor on the convergence ratio. Thus, one may only hope to improve the present solution, but the actual optimal solution remains unknown. On the other hand, strict mathematical optimization approaches like linear programming often require several assumptions on the problem formulations which lead to very simplified traffic models. However, these strict mathematical strategies often provide much more insight in the underlying problems, e.g., via the dual problem formulation. That is, out of the solution we can directly derive quantified suggestions for an improvement of the network performance. Therefore, one has to find a compromise between a qualified optimality result and the model's accuracy.

Our contribution In this paper we aim at the crossover between static and dynamic models for traffic assignment. Our main result is a combined optimization approach for simultaneous traffic assignment and traffic signal coordination with realistic load dependent travel times which still allows applying strict mathematical optimization techniques.

Firstly, we shortly present static and dynamic traffic models and introduce the main concept of our cyclically time-expanded network model. In Section 3, we show that this new discretized and linearized model yields load dependent travel times which resemble main properties of common link performance functions. Moreover, our model is also capable of capturing platoons of cars of quickly varying density as well as changing phases of traffic signals. We verify the applicability of our model by simulation with state-of-the-art simulation tools, namely VISSIM and MATSim. Furthermore, we present an extension of our cyclically time-expanded model in Section 4, which allows for the

simultaneous optimization of traffic signal parameters to design fixed time or time of day signal timings. Whereas the previous version of our model was only capable of optimizing offsets, we extend our approach to the simultaneous optimization of split times and phase orders. Afterwards, we study solutions of real-world instances in Section 5. Finally, we use the traffic simulation MatSim to investigate the price of anarchy for our scenarios, i.e., the gap between system optimum and user equilibrium.

2 A Cyclically Time-Expanded Model for Traffic Assignment

2.1 Static and dynamic models for traffic assignment

Intersections and connecting roads can be represented by the set V of nodes and the set A of links of a (directed) network $G = (V, A, u)$, where $u : A \rightarrow \mathbb{R}_0^+$ assigns a capacity to each of the arcs in the network. In a static network flow model, we assign flow values $f : A \rightarrow \mathbb{R}_0^+$, such that capacities are not exceeded and flow conservation is satisfied. That is, the amount of incoming flow equals the amount of outgoing flow. First results on efficiently computing maximum flows were obtained by Ford and Fulkerson [11] already in the 1950s. Due to the abundance of applications ranging from traffic flows to image processing, numerous algorithms and modifications have been developed.

However, many applications require that the flow is a function of time, i.e., flow values are changing over time. Generally, a (traffic) *flow over time* or dynamic flow can formally be defined as a function $f : A \times [0, T) \rightarrow \mathbb{R}_0^+$ on a (directed) network $G = (V, A, u)$, where $[0, T)$ is the time interval under consideration. Flow particles entering $e \in A$ at time t arrive at the head of e at time $t + t_e(f(e, t))$, where $t_e : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ is the travel time function or *link performance function* of link e . The amount of flow that passes a link e can be calculated by $\int_0^T f(e, t) dt$. Furthermore, *flow conservation* has to ensure that flow cannot leave a node before it arrives there. Additionally, one may assume that flow can be stored in a node for a certain time span representing queues at intersections. In detail, for load-independent travel times t_e , for each $\tau \in [0, T)$, and each non-terminal $v \in V$, $\sum_{e \in \delta^-(v)} \int_0^{\tau - t_e} f(e, t) dt \geq \sum_{e \in \delta^+(v)} \int_0^{\tau} f(e, t) dt$, where $\delta^-(v)$ and $\delta^+(v)$ refers to the set of incoming and outgoing arcs of v .

Flows over time are widely studied in the mathematical literature; Skutella provides a good overview of recent results [29]. Many dynamic flow problems with travel times independent of the flow can be solved using *time-expanded networks*. Already Ford and Fulkerson introduced time-expanded networks in their seminal work on network flow theory [11]. To create an expanded network out of a simple network, for every node, several copies of this node are added to the graph, one for each desired time step. These nodes are connected by arcs, where the various copies of the vertices are connected according to the travel times of the original arcs. Additionally, arcs connecting consecutive copies of the same node model waiting at this node. Yet, time-expanded networks are rather inefficient if the time horizon T is large, since the number of time steps determines the number of network copies that have to be provided.

Ford and Fulkerson's time-expanded networks were also applied in traffic flow theory, often called *space-time expanded networks* (STEN) in this context [37]. *Dynamic traffic*

assignment (DTA) combines flow over time with variable departure times and non-linear link performance functions. It is used to model flow dependent travel times and to compute user equilibria according to Wardrop's principle. This approach allows the usage of solving techniques of Control Theory or Calculus of Variations. However, even for medium size networks it is very difficult to derive analytical solutions. Therefore, to be able to compute at least numerical solutions in an iterative approach, time is discretized again and several other parameters like the number of routes are limited.

2.2 The cyclically time-expanded network

For traffic signals and their coordination, a time-dependent model, capable of describing the time-offset between consecutive intersections, is a vital ingredient. The impact of *red* signals on the travel times is essential.

Link performance functions, as suggested by the Bureau of Public Roads (BPR), are a widely accepted approach to model load-dependent travel times. However, these static link performance functions describe traffic in a rather statistical manner, i.e., they support an a priori estimation of the expected average delay. Especially without standard deviations, it is hardly possible to deduce the travel time of an individual driver. Hence, link performance functions do not provide enough information to be used directly for traffic signal coordination. Furthermore, to the best of our knowledge, approaches like DTA are not used to model dynamic flow in realistic large scale scenarios on such a fine timely level as needed for traffic signals.

Indeed, one should look into the causes for flow dependence of travel times in urban areas. In inner-city road networks, the individual behavior of road users is of minor importance for the travel times. In comparison, traffic signals and the rather low speed limit have a much greater influence on travel times. Thus, for urban networks, we suggest to decompose the travel time into *pure transit time* and *waiting time* at the intersections. In the following, we assume that the *transit time is constant* and the *flow-dependent component of travel time is caused only by the delay* at traffic signals.

With constant transit times, time-expanded networks come into play again. Even better, since pretimed traffic signals have a periodic behavior, it is not necessary to use a full time horizon expansion. Instead, we suggest a *cyclic time-expansion* where we expand only a time interval of size of the cycle time Γ of the traffic signals and use only $k \in \mathbb{N}$ time steps of size $t = \frac{\Gamma}{k}$.

Definition 1 (Cyclically time-expanded network). *Let $G = (V, A, u)$ be a network with capacities $u : A \rightarrow \mathbb{N}$ and non-negative integral transit times t_e for each $e \in A$. For a given number k of time steps of length $t = \frac{\Gamma}{k}$, the corresponding cyclically time-expanded network $G^T = (V^T, A^T, u^T)$ is constructed as follows.*

- For each node $v \in V$, we create k copies v_0, v_1, \dots, v_{k-1} , thus $V^T = \{v_t | v \in V, t \in \{0, \dots, k-1\}\}$.
- For each link $e = (v, w) \in A$, we create k copies e_0, e_1, \dots, e_{k-1} where arc e_t connects node v_t to node $w_{(t+t_e) \bmod k}$. These arcs are called *transit arcs* and e_t has capacity $u(e_t) := u(e)$.

- Additionally, there are waiting arcs from v_t to $v_{t+1} \forall v \in V$ and $\forall t \in \{0, \dots, k-2\}$ and from v_{k-1} to v_0 .

Please note that there is almost no difference between transit arcs and waiting arcs in the model. On waiting arcs one only moves in time whereas on transit arcs one moves in space and time. For all arcs the cost is the travel time on this arc. Thus, waiting time is travel time on waiting arcs. Hence, both kinds of arcs are treated in the same way and we do not need to explicitly distinguish between them in the modeling of traffic assignment. In practice, we will compute the transit time via the speed limit and the length of the road segment, whereas the waiting time on a waiting arc is simply determined as the duration of a time step in the expansion.

Now, traffic signals with their *green* and *red* phases can be modeled by setting the capacity of a transit arc starting during the *red* phase to zero, i.e., $u(e_t) := 0$ for some t . A simple example of a cyclically time-expanded network consisting of two links and one traffic signal is shown in Figure 1. Additionally, capacities of the waiting arcs correspond to the maximum queue length on a link and they can be chosen accordingly. This also allows the modeling of vehicle spillback. If the capacity of a waiting arc is exhausted and the signal is still *red*, no more flow can enter the corresponding transit arc due to the flow conservation constraint. Thus, flow on the upstream link has to use the waiting arc there, although the associated traffic light might be *green*.

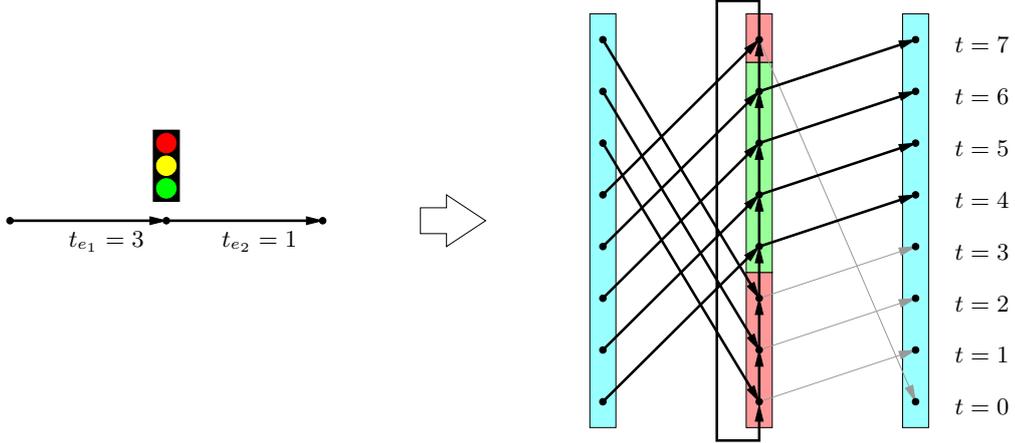


Figure 1: A small network with two links and one traffic signal (left) and its cyclic time-expansion for $k = 8$ (right). The thin, gray arcs starting at the red phase of the signal are switched off and, thus, they have zero capacity.

Consequently, the cyclically time-expanded network model is a hybrid inbetween static and dynamic models. On the one hand, it provides several advantages of static flow models. Due to its linearity and limited size, standard approaches for computing (static) multi-commodity flows can be applied. On the other hand, it provides sufficient time resolution to account for traffic signals. Yet, only a narrow time slice is considered, hence it is not a fully dynamic model.

3 Inherent Link Performance in the Cyclically Time-Expanded Network

So far, we have introduced a model with constant travel times on transit arcs and waiting arcs. Consequently, travel times are independent of the flow on the link. From a practitioner's point of view, this might seem too restrictive in the first moment. However, there is in fact a flow-dependent behavior and it is hidden in the time expansion. We will now study this flow dependency by an analysis of our model for a single road with a traffic signal. Furthermore, we support the observations by a simulation of traffic flow on this link. As a main insight, we will see that our high-dimensional linear model can reproduce the same effects as a classical low-dimensional non-linear traffic flow model.

In the cyclically time-expanded model, we can look at traffic flow in two ways. *Firstly*, we can examine each of the copies of an original link and the adjacent waiting arcs on their own. This provides a dynamic and detailed view on the traffic flow and we can observe flow particles moving through the network in a timely fashion. Considering only one road segment, which is modeled by k transit arcs and k waiting arcs, we have a $2k$ -dimensional linear relation between flow and travel time.

Secondly, we can virtually contract this flow over time back to the original link, i.e., we regard all copies of the links and the corresponding waiting arcs as a single link and compute the average travel time over the whole time horizon. For a single link, we can imagine this as a projection of the $2k$ -dimensional space, which yields a piecewise linear relation between total flow and average travel time. In that way, we have a more static view on traffic flow on this contracted link and temporal details are lost, but we can compare the travel time to other static models, now.

In the following, we will use the first point of view to compute the average travel times for the second perspective. We will see that travel time on a contracted link depends on numerous parameters, e.g., the distribution of incoming flow values over time. Furthermore, flow particles on the same (contracted) link at different times will experience different travel times which is very obvious in the time-expanded model. We use *link performance* to denote the average travel time on a contracted link. However, we will not present a closed function for computing travel times due to the enormous number of parameters. Nevertheless, we use the term *inherent link performance function* according to related static traffic assignment models with flow dependent travel times. In the following, we fix most of the parameters and restrict the inherent link performance function to a small subset of its domain to at least demonstrate the most important properties of this implicit link performance in the cyclically time-expanded model.

3.1 Evolving flow-dependent waiting times

We consider a very small network which consists of a single road with a traffic signal in the middle similar to the network in Figure 1. Assume that a set of flow values is assigned to the incoming transit arcs in the cyclically time-expanded network of this scenario and a fixed signal setting is given. The flow values of the other arcs are chosen such that flow conservation is fulfilled and flow is assigned to a waiting arc if and only

if the capacity of the outgoing transit arc at this node is exhausted. That is, flow units only wait when the outgoing road is already blocked by other flow units.

The total waiting time in this small network can be determined by summing up all flow value of the waiting arcs and multiplying the result by the length of one time step. Now, we increase the flow values on the incoming transit arcs. This will increase the waiting time, because more flow will be assigned to the waiting arcs. Furthermore, due to the bounded capacities, the flow units on the waiting arcs may not leave completely at the first green outgoing transit arc if this link does not provide enough capacity. Instead, the flow will have to use more waiting arcs until the accumulated flow is drained off. This relation is illustrated in Figure 2. Therefore, if the incoming flow is raised linearly on all copies of the incoming transit arc, then the growth of the waiting time will not be linear but quadratic. More precisely, the obtained function is piecewise linear, but converges to a quadratic function if the length of the time steps tends to zero.

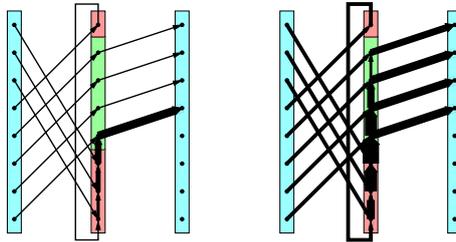


Figure 2: The same single road network with two different flows. The thickness of the arcs depicts the flow value. On the left side, incoming flow is low. All flow accumulated during the *red* phase can leave the signal directly when it becomes *green*. On the right side, the incoming flow is tripled. Thus, much more flow is accumulated on the waiting arcs. Furthermore, due to the capacity bound of the outgoing street, only a fraction of flow can leave on the first outgoing transit arc and more waiting arcs have to be used. This yields an increase of waiting time by a factor of 4.8.

In Figure 3, we present the relation between flow and average waiting time in the cyclically time-expanded network with a finer time granularity. A cycle time of 60 seconds and 60 time steps are used. The incoming traffic is uniformly distributed on all transit arcs, a traffic signal which is *red* for 20 seconds is put at the end of the road, and a free speed travel time of 10 seconds is assumed. Additionally, a capacity reduction from two lanes to one lane at the traffic signal was used. We now compute the average travel time in this scenario with respect to the flow value. The resulting travel time in Figure 3 demonstrates the capability of our waiting arc model: although using only constant travel times the inherent, implicit link performance functions of the model are not linear. Even better, the implicit link performance resembles a common standard link performance function. Please note that Figure 3 visualizes the average travel time of a flow particle on the ‘contracted’ link. The individual travel time of a single flow unit depends on its arrival time at the traffic signal. It ranges from 10 seconds for flow

units arriving at *green* with no waiting queue at the signal up to 30 seconds for road users arriving at the beginning of the *red* phase.

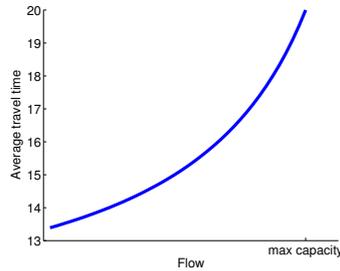


Figure 3: Relation between average travel time and flow value on a single link in the cyclically time-expanded network. Incoming traffic is uniformly distributed over time.

This example implies that our model uses flow dependent travel times even though they are not explicitly specified. But within this analysis, we assumed traffic to be uniformly distributed over time. Obviously, this is an unrealizable assumption when the steady flow is disturbed by traffic signals. Hence, we extend our study to flows with quickly changing traffic density.

3.2 Travel times for platoons of cars

In inner-city traffic, one can often observe platoons of cars. Traffic signals cause an accumulation of cars. During one cycle of the signals, there are periods with cars nose-to-tail and periods with empty streets. In the cyclically time-expanded network, platoons of cars can be easily modeled by different flow values on the copies of each particular transit arc. Some copies may be used at full capacity, other arcs may not be used at all. Varying flow values can be interpreted as platoons of different lengths and densities. Thus, platoons are also created, splitted, merged, compressed, or stretched in our model in a very natural and self-acting manner.

Even in our small example of a single road with one traffic signal, a platoon can change the characteristics of the average waiting time dramatically. Let us describe a platoon of cars by its length and its density. Suppose for the moment the platoon to be homogeneous, i.e., the density is not varying inside the platoon. Therefore, the platoon length can be considered to be proportional to the average link flow. For simplicity, we measure the length of a platoon as time elapsing between passing of head and tail of this platoon. This can be easily transformed to length in meter or number of cars, when speed or distance between cars is known. Average travel time now depends on two parameters: the length of the platoon and the arrival time of the head of the platoon at the traffic signal.

To emphasize the occurring effects even more, we change the *green* period to 30 seconds and the capacity of the outgoing road is set to twice the capacity of the incoming road, e.g., the number of lanes are doubled. The scenario is visualized in Figure 4.



Figure 4: The described scenario was modeled in VISSIM (screenshot detail). Travel time is measured between A shortly after the upstream signal and B shortly after the actual signal. Due to the first signal, platoons are generated.

In Figure 5, the calculated average travel time is plotted versus the platoon length. In Figure 5 (left), the first flow unit arrives at the intersection 10 seconds before the signal turns *green*. In Figure 5 (right), the head of the platoon arrives 20 seconds after the signal turned *green*. As a consequence, the traffic signal is used to densify the traffic in the first setting. In the second setting, long platoons are split into two smaller platoons, of which the second one has to wait at the signal. The different characteristics of the functions in Figure 5 demonstrate the impact of these two offsets on our travel times.

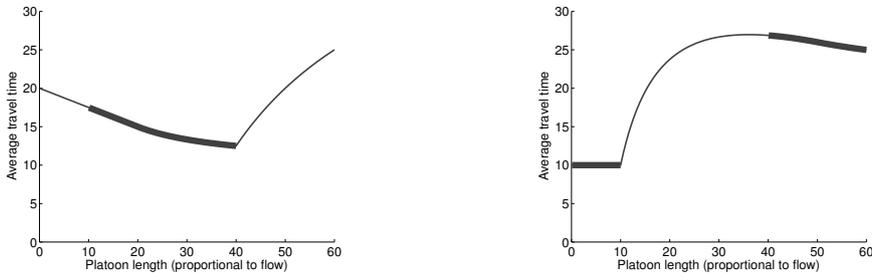


Figure 5: The plots show the inherent link performance function for a platoon of cars in the cyclically time-expanded model for two different offsets. The thick line also visualizes the *green* period, i.e., that part of the platoon arriving at a *green* signal. In the left diagram, the head of the platoon is arriving 10 seconds before the signal turns *green*. Due to the higher capacity of the outgoing road, that is, the cars may leave on parallel lanes, the platoon is densified. The first car of the platoon has to wait for the longest time, subsequent cars may even pass the signal without stopping. On the right side, the first flow particle arrives 20 seconds after the signal turned *green*. Thus, small platoons can pass without stopping, long platoons are splitted by the traffic signal which turns *red* 10 seconds after the head of the platoon has passed.

The implicit travel times for platoons show interesting or even unexpected properties. The travel times are decreasing or concave in some intervals, so they do not resemble standard static link performance functions found in traffic literature (e.g., [9, 16, 28]). There, link performance functions are usually assumed to be convex and monotonically increasing with the amount of flow on the link. However, this is no inaccuracy or disadvantage of our model. As we will show, these characteristics of travel time functions can in fact be observed in state-of-the-art traffic simulation tools which supports our

model.

To evaluate the travel times computed by the cyclically time-expanded model, we use two well established simulation tools, namely VISSIM and MATSim. VISSIM, developed by ptv group, is a state-of-the-art micro-simulation, featuring nearly every aspect of urban traffic [35]. MATSim, mainly developed by TU Berlin and ETH Zurich, is an agent-based simulation capable of computing large scale traffic simulations. MATSim uses a simpler queuing model, but it supports the computation of equilibrium assignments for thousands of traffic participants [22]. Figures 4 and 6 present the described scenario and the corresponding measured travel times, respectively. To create the platoons, we use an additional upstream traffic signal (in front of A in Figure 4). The arrival time of the platoon at the signal of interest can be influenced by changing the offset between both signals. The length of the platoon is determined by the *green* period of the upstream signal.

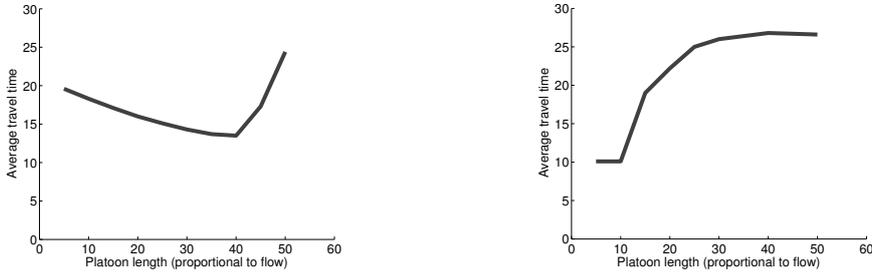


Figure 6: Link performance for the same scenario as in Figure 5, but now simulated and measured with VISSIM. Note that only the average travel time is displayed. The actual time depends on the position of the car within the platoon.

As one can clearly see, the predicted travel times in Figure 5 fit remarkably well to the simulated travel times in Figure 6 even without a careful calibration of our model. Hence, one can conclude that the cyclically time-expanded model can also capture time dependency of flow on a very fine level, especially for platoons of cars and traffic signals. In contrast, static link performance functions are the best choice for describing traffic flow in rural areas or on highways, but they are not accurate enough for signalized inner-city traffic. This also affects user equilibrium flows. Here, the common assumptions on convexity and monotonicity are crucial for the existence of unique equilibria.

However, we cannot present a closed formulation of travel times for platoons and signals here, since this depends on various parameters, e.g., number, length and density of platoons, traffic signal settings, and capacity changes of the link. If most parameters are fixed, the average travel time of a flow unit of a single platoon in our scenario is depicted in Figure 7 with respect to both parameters platoon length (i.e., traffic flow) and arrival time at the intersection (i.e., offset of the signal).

Summarizing, our cyclically time-expanded network provides a piecewise linear approximation of classical non-linear link performance functions, but it can also keep track of platoons of cars and varying traffic density. To avoid any misunderstanding, please

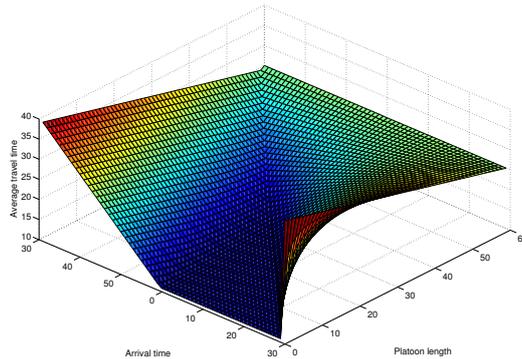


Figure 7: Calculated average travel time for a road user in a platoon with respect to platoon length and arrival time at the traffic signal. The signal turns *green* at $t = 0$ and *red* at $t = 30$. The inherent link performance functions of Figure 5 are obtained as profile for $t = 50$ and $t = 20$, respectively. Note that for fixed platoon length one obtains waiting time functions very similar to those used by Wunsch [36].

note that the above results are just used to compare our model to other non-linear approaches. The non-linear link performance functions that were studied in this section are not used in the remainder of this paper. All further results are obtained by using the original cyclically time-expanded network with its constant transit times.

4 Traffic Signal Optimization in the Cyclically Time-Expanded Model

Up to now, we analyzed properties of the cyclically time-expanded model. In the following, we show how our model can be used for traffic signal optimization. While we only considered offset optimization in our previous paper [20], we now present an extended approach which allows simultaneous optimization of traffic assignment, traffic signal offsets, split times and phase order.

4.1 Efficient computation of traffic assignments

Building up on the cyclically time-expanded network, we can now consider the traffic assignment problem within this framework. Although flow can be considered to travel in a time-dependent manner through the cyclically time-expanded network, one should rather see this model as a static model that just captures some time-dependent aspects of traffic flow. Due to the cyclic repetition of the vertex and arc copies, a flow particle traveling through this network can be seen as a representative of a whole set of temporally repeated particles at every multiple of the cycle time.

There is a closely related interpretation of static network flow for traffic networks. In this interpretation one considers a flow-carrying path in the static network as a mapping of a corresponding amount of flow particles traveling over time through the traffic network at the corresponding flow rate. In other words, a flow-carrying path in the static network represents a constant rate of flow on this path in the ‘real’ time-dependent traffic network.

This interpretation suggests how to put together the two models, the cyclically time-expanded network on the one hand and the static traffic assignment model on the other hand. Basically, the demands for different commodities¹ have to be subdivided to the number of layers/time steps in the cyclically time-expanded network.

The cyclically time-expanded model is a fully linear model with constant travel times on the expanded links. Also with the additional extensions in the following sections, the model is still a linear one. Thus, we can use standard network flow algorithms to efficiently compute a system optimal traffic assignment, i.e., fast and exact combinatorial and mathematical programming tools can be applied. A very fine time discretization of time steps less than one second is easily achievable and quickly solvable even for large scenarios.

Theorem 1 ([20]). *Using the cyclically time-expanded network the traffic assignment problem for a fixed traffic signal coordination and for a fixed time granularity can be solved efficiently.*

Proof. The assignment problem can be formulated as a linear program (see [1]), because all travel times in the expanded network are constant. Since the LP has polynomial size with respect to the input, i.e., the cyclically time-expanded network, it can be solved in polynomial time, e.g., by using the ellipsoid method. \square

4.2 Detailed modeling of intersections

In Section 2.2, we modeled traffic signals by setting capacities of outgoing links to zero. At an intersection, we have to consider several lights and conflicts between crossing directions. Thus, to model intersections with different lanes, turning directions, and interior traffic signal offsets, we initially use a standard approach from traffic networks. Before the time expansion is employed, every intersection node is split up into several nodes for *incoming* and *outgoing traffic*; for each turning direction an interior arc connects one incoming node with one outgoing node (see Figure 8). Each of these interior arcs is assigned to one traffic light.

We will refer to the subset of all interior arcs of intersections by $E \subset A$. By choosing appropriate capacities $u(e)$ and transit times $t(e)$, $e \in E$, these interior links limit the outgoing flow of the waiting arcs of the several turning directions. Hence, we can also influence the creation of queues and back spill with these parameters.

¹According to mathematical network flows, the term commodity denotes an origin-destination pair.

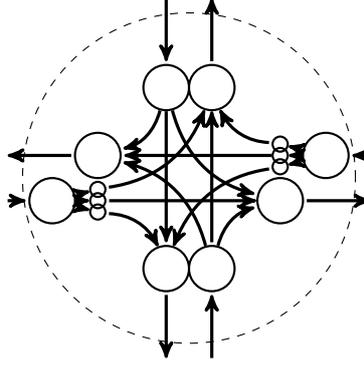


Figure 8: Standard approach of an expanded intersection with arcs for each turning alternative. The incoming nodes of the horizontal road are subdivided into three nodes to model different lanes and queues for the turning directions. The vertical road is smaller, thus, all cars are waiting in the same queue.

4.3 Modeling traffic signals

Instead of setting fixed capacities to interior arcs of intersections, traffic signals can be modeled by binary decision variables that switch the capacities of the interior arcs on or off by multiplying the capacity with an appropriate binary variable. For each turning direction $e \in E$ we introduce k binary variables $b^e = (b_0^e, \dots, b_{k-1}^e) \in \{0, 1\}^k$, one for each time step. Now, we multiply the binary variables with the capacity constraint of the transit arcs corresponding to e in the cyclically time-expanded network. This yields k inequalities for the flow in the expanded network, namely $f(e_t) \leq b_t^e u(e_t) \quad \forall t \in \{0, \dots, k-1\}$. If $b_t^e = 0$, then the effective capacity of e_t is zero. Thus, turning direction e cannot be used in time step t , i.e., the corresponding signal group shows a *red* light. If $b_t^e = 1$, then the capacity remains unchanged, hence we may think of a *green* traffic light at time step t .

With these binary variables, we can permit or block flow on the interior arcs of an intersection. However, without further constraints, each signal group would be *green* all the time. To ensure the interior logic of traffic lights at a single intersection, we have to couple these binary variables as in logic-based signal controls [13]. We concentrate on the following constraints:

- Each signal group should switch to *green* only once per cycle.
- Crossing directions must not have *green* at the same time.
- Some directions must have *green* at the same time.
- There has to be enough time to clear the intersection at phase changes before any other direction gets *green*.
- There has to be a minimum duration of each *green* and *red* phase.

In reality, many more constraints can occur, e.g., due to pedestrians. For simplicity, we only discuss here a small standard intersection with four legs and four signal groups W, X, Y, Z as depicted in Figure 9. The common cycle time is Γ and we use a cyclical expansion with k time steps. Thus, in the following, all indices apply modulo k . As described above, we introduce k binary variables $b^W = (b_0^W, \dots, b_{k-1}^W)$ for signal group W and similarly for X, Y , and Z that represent the status of the corresponding light directly. We refer to these variables as *status variables*.

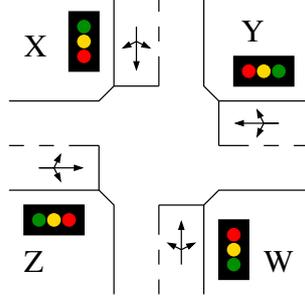


Figure 9: Simple intersection with four signal groups W, X, Y and Z .

Now, we demand that the signal group W switches from *red* to *green* and from *green* to *red* only once during one cycle. To force this, we use $2k$ binary variables $b^{W,\text{on}} = (b_0^{W,\text{on}}, \dots, b_{k-1}^{W,\text{on}})$ and $b^{W,\text{off}} = (b_0^{W,\text{off}}, \dots, b_{k-1}^{W,\text{off}})$, respectively. These variables act like the decision variables for the offset in our previous model [20] and consequently, we refer to them as *decision variables*. Thus, $b_i^{W,\text{on}} = 1$ means that signal group W switches from *red* to *green* at time step i . The following constraints fix the number of switching operations per cycle to 1.

$$\sum_{i=1}^k b_i^{W,\text{on}} = 1 \text{ and } \sum_{i=1}^k b_i^{W,\text{off}} = 1$$

Now, we link the decision variables to the status variables (remember to regard time modulo k).

$$b_i^W - b_{i-1}^W \leq b_i^{W,\text{on}} \quad \forall i \in \{0, \dots, k-1\}$$

Thus, if $b_i^{W,\text{on}} = 1$, then the signal may switch to *green*, i.e., $b_{i-1}^W = 0$ and $b_i^W = 1$. If $b_i^{W,\text{on}} = 0$, then $b_{i-1}^W \geq b_i^W$, that is, the status is not changed or the signal switches to *red*. Similarly, we formulate the constraints for switching the signal to *red*.

$$b_i^W - b_{i-1}^W \geq -b_i^{W,\text{off}} \quad \forall i \in \{0, \dots, k-1\}$$

However, it is not guaranteed that the signal group switches at all. We combine this with a constraint for minimum *green* and *red* periods. If the signal group should be *green* for at least g time steps, we require:

$$\sum_{i=1}^k b_i^W \geq g.$$

If the signal group should show *red* for at least r time steps, we require:

$$\sum_{i=1}^k b_i^W \leq k - r.$$

Therefore, if the signal group is *green* for at least one time step and it is *red* for at least one time step, two switching operations are forced.

So far, we can switch traffic lights and we can control capacities. We continue with the safety constraints. Crossing directions must not have *green* at the same time. However, *red* for both directions is possible. Obviously, the following constraints ensure safe crossing for two signal groups W and X (cf. Figure 9).

$$b_i^W + b_i^X \leq 1 \quad \forall i \in \{0, \dots, k-1\}$$

Furthermore, we have to guarantee a clearance time. Together with the previous inequality, a clearance time of at least one time step is forced by the following constraint.

$$b_i^{W,\text{off}} + b_i^{X,\text{on}} + b_{i+1}^{X,\text{on}} \leq 1 \quad \forall i \in \{0, \dots, k-1\}$$

Thus, if signal group W switches to *red* in time step i , then X cannot switch to *green* at the same time step or at time step $i+1$. Longer clearance times are possible. One may even insert another phase. Hence, clearance time constraints have to be formulated for each pair of signal groups. In contrast, we may also fix the clearance time and the phase order using the next equation.

$$b_i^{W,\text{off}} = b_{i+2}^{X,\text{on}} \quad \forall i \in \{0, \dots, k-1\}$$

Finally, if two signal groups should switch to *green* (or *red*) at the same time, e.g., group W and Y in Figure 9, we may use the following constraints.

$$b_i^{W,\text{on}} = b_i^{Y,\text{on}} \quad \forall i \in \{0, \dots, k-1\}$$

4.4 Simultaneous optimization of assignment and signals

Each change in the traffic signal settings changes the travel times in the network. Thus, traffic signals and assignment have to be optimized simultaneously. We will refer to this problem as COMBINED TRAFFIC ASSIGNMENT TRAFFIC SIGNAL COORDINATION problem, i.e., find a traffic signal setting and an traffic assignment such that the overall travel time of all road users is minimized.

The choice of the objective is motivated by the following consideration. When optimizing the traffic assignment problem and the traffic signal coordination simultaneously, measuring delays and stoppages does not characterize efficient coordinations. Traffic

participants may choose arbitrary routes. Thus, taking only stops and delay into account, a road user may choose a very long detour through the network just to avoid stopping or waiting in front of a *red* traffic light. This is rather unrealistic as most road users are interested in the fastest way to their destination. Therefore, we use the total or average travel time (pure transit time + delay) as the measure of quality of the solution.

Summarizing the results of the previous subsection, we have introduced several linear equations or inequalities for the most important safety constraints at a traffic signal. Let $G^T = (V^T, A^T, u^T)$ be a cyclically time-expanded traffic network and commodities $\theta \in \Theta$, $\theta = (s_\theta, z_\theta, d_\theta)$ (with origin $s_\theta \in V$, destination $z_\theta \in V$, and demand $d_\theta \in \mathbb{N}$), capacities $u : A \rightarrow \mathbb{N}$, a set $E \subset A$ of interior arcs at intersections with associated status variables b^e for each $e \in E$, constant travel times $t(e)$ for each link $e \in A$ and flow functions $f_\theta : A \rightarrow \mathbb{R}$ for each commodity.

We extend a common multicommodity min-cost circulation program (i.e., use a backward arc from z_θ to s_θ for each commodity $\theta \in \Theta$) for the cyclically time-expanded network by adding the binary variables and capacity constraints above. Adding these binary variables and constraints to the linear program leads to the following mixed integer program for the COMBINED TRAFFIC ASSIGNMENT TRAFFIC SIGNAL COORDINATION problem:

$$\min \sum_{e \in A} \sum_{\theta \in \Theta} t_e f_\theta(e)$$

$$\text{s. t. } 0 \leq \sum_{\theta \in \Theta} f_\theta(e) = f(e) \leq u(e) \quad \forall e \in A \setminus E \quad (1)$$

$$\sum_{e \in \delta^+(v)} f_\theta(e) = \sum_{e \in \delta^-(v)} f_\theta(e) \quad \forall \theta \in \Theta \quad \forall v \in V \quad (2)$$

$$f_\theta((z_\theta, s_\theta)) = d_\theta \quad \forall \theta \in \Theta \quad (3)$$

$$f(e) \leq b^e u(e) \quad \forall e \in E \quad (4)$$

$$\text{Constraints of Section 4.3} \quad (5)$$

$$f(e) \geq 0, b^e \in \{0, 1\}^k$$

The constraints of type (1) fix the capacity bounds, type (2) implements the flow conservation and the constraints (3) force the circulation. Equation (4) permits flow only on arcs that are switched on by the binary variables and in (5), we insert all necessary constraints for these status variables. Due to the numerous equations for every intersection and every lane we omit the details here.

4.5 Solving the MIP

The mixed integer linear program for the COMBINED TRAFFIC ASSIGNMENT TRAFFIC SIGNAL COORDINATION problem can be passed to every appropriate solver, for example CPLEX or GUROBI. Due to the time-expansion and the decision variables, the MIP-formulation of our model is rather large. In practice (see also Section 5), good solutions

are found quite fast, but it takes a lot of time to close the gap between primal and dual solution. In some larger scenarios, a gap of 10 to 30 % remains even after hours of computation.

On the one hand, some further efforts are necessary to get fast and provable good solutions with help of the presented model. On the other hand, the COMBINED TRAFFIC ASSIGNMENT TRAFFIC SIGNAL COORDINATION problem is an \mathcal{NP} -hard problem and we may not hope for efficient algorithms (see, e.g, [14] for a guide to \mathcal{NP} -completeness). Even approximate solutions for a constant approximation ratio will be hard to get as shown in the following theorem.

Theorem 2. *For traffic signals which may only switch at discrete points in time, there exists no polynomial-time α -approximation algorithm for any constant $\alpha \geq 1$ for the COMBINED TRAFFIC ASSIGNMENT TRAFFIC SIGNAL COORDINATION problem unless $\mathcal{P} = \mathcal{NP}$.*

Proof. We use a reduction from the k -DIRECTED VERTEX DISJOINT PATHS problem. This problem is defined as follows. Given a directed graph G with k terminal pairs (s_1, t_1) to (s_k, t_k) , are there k node-disjoint paths that connect the corresponding terminals? Even for $k = 2$, this problem is strongly \mathcal{NP} -complete [12].

Given an arbitrary instance of the 2-DIRECTED VERTEX DISJOINT PATHS-problem, we construct the following combined traffic assignment traffic signal coordination problem. The network is formed by assigning unit capacities to the arcs, i.e., $u \equiv 1$, and zero transit times, i.e., $t_e = 0 \quad \forall e \in E$. We use two commodities, each with demand $d_{1/2} = 1$. Commodity i ($i \in \{1, 2\}$) starts at s_i and ends at t_i . Each node of the graph is transformed to a signalized intersection and we assume that each traffic signal may only switch at integral points in time. At each signal, each turning direction has a separate traffic light and only one traffic light is allowed to be *green* at each time step. That is, for all interior intersection arcs $\sum_X b_i^X \leq 1$ for all time steps i . Further, after each *green* period of the signal there is a clearance time of at least $\lceil 2\alpha \rceil$ time steps with all lights *red*.

Consequently, each single commodity can be routed in one time step. Due to the signal settings, commodities cannot split without waiting. Hence, both commodities can be routed in one time step if and only if there exist two vertex disjoint paths in the original network. In contrast, if there exists no pair of vertex disjoint paths, both commodities have to meet at an intersection. They have to cross this intersection one after the other. Thus, half of the flow needs at least $\lceil 2\alpha \rceil$ time steps to reach its destination.

Therefore, any solution of the COMBINED TRAFFIC ASSIGNMENT TRAFFIC SIGNAL COORDINATION problem that needs less than $\lceil 2\alpha \rceil$ time steps (or less than $\lceil \alpha \rceil$ total travel time) implies the existence of two vertex disjoint paths. This yields the claim. \square

Obviously, the construction in the proof is somewhat extreme. Thus, one should not be discouraged by the above result, since one may get quite good solutions for realistic traffic networks in an acceptable time. However, even for traffic signals which may switch continuously, i.e., at arbitrary points in time, and without clearance times as

well as planar networks with a maximum node degree of 4, the problem remains \mathcal{NP} -hard (see [32]). Consequently, mixed integer programming is a reasonable approach for solving the COMBINED TRAFFIC ASSIGNMENT TRAFFIC SIGNAL COORDINATION problem, since we may not hope for an efficient algorithm unless $\mathcal{P} = \mathcal{NP}$.

4.6 System optimum vs. user equilibrium

Since the famous example of Braess [5], one should be aware that road users may behave in a selfish way that is not optimal for the system as a whole.

However, up to here, we only computed system optimal assignments based on the cyclically time-expanded model. This has two reasons. Firstly, we use a network with constant travel times. Thus, the system optimum is also a user equilibrium by definition. None of the road users can switch to a faster path towards the destination. Secondly, we use a network with capacities. Hence, [8] implies that the user equilibrium is not unique anymore and that different equilibria may occur.

The theoretical results on the gap between system optimum and user equilibrium in static models are based upon properties of the link performance functions [26]. In the cyclically time-expanded model, these assumptions are not fulfilled. Instead, we have seen in Section 3 that we can capture much more details of the time dependent behavior. Thus, it is not possible to compare the (system optimal or user equilibrium) solutions of the static model and the cyclically time-expanded model directly.

Furthermore, we have seen non-convex and non-monotone travel time functions in Section 3, which were created due to signals and platoons. Hence, considering signals and platoons in the static case with help of more sophisticated link performance functions would not meet the conditions which ensure the existence and uniqueness of an user equilibrium, neither.

Since we cannot provide a positive theoretical result here, we move the discussion of user equilibria to Section 5 where we compare our system optimal solutions to equilibria obtained with a multi-agent simulation.

5 Evaluation of real world scenarios

In this section, we will use two real-world scenarios to demonstrate the capabilities of the cyclically time-expanded model. The first instance is a small part of the inner-city of Brunswick, Germany. Here, the focus is the optimization of offsets, split times and phase orders. In the second scenario, the inner-city of Cottbus, Germany, we investigate the interplay of offset optimization and traffic assignment. Afterwards, simulation tools are applied to study the gap between system optimum and user equilibrium.

5.1 Brunswick

The first scenario consists of five signalized intersections and two pedestrian crossings in the inner-city of Brunswick. This scenario was provided by Sándor Fekete from Technical University of Brunswick. It consists of the Bohlweg, a main arterial street through

the city center connecting university and main station between Fallersleber Straße and Schloßplatz.

The network is presented in Figure 10. We study three commodities visualized by colored arcs. The orange commodity and purple commodity have each twice the demand of the blue commodity. Since there are no alternative routes, route choice is of no importance in this scenario. As an additional constraint, the pedestrian signals 6 and 7 have to switch twice during one cycle of 84 seconds. Due to a nearby shopping mall and a tram station, a rather long minimum green time was assigned to these pedestrian signals. Consequently, these two signals will most likely be the bottleneck in this scenario.

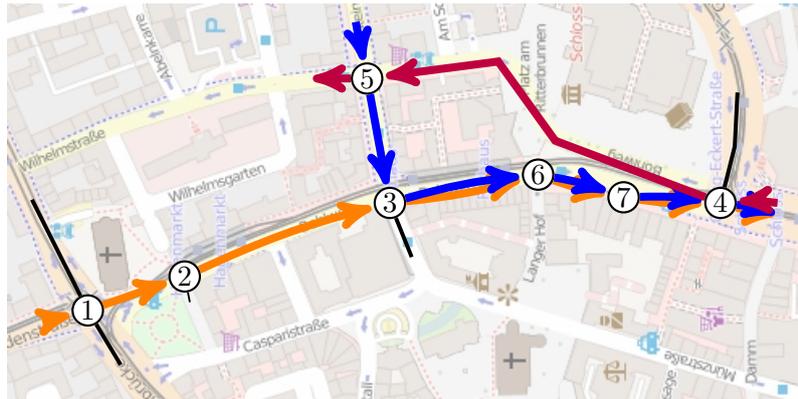


Figure 10: The Brunswick network with 5 signalized intersections (1–5) and two pedestrian signals (6,7). The three commodities under consideration are visualized with arcs in different colors. The underlying map was created with www.openstreetmap.org.

Usually, the optimal solution for instances of this scenario in our framework is found and proved within a few seconds. Solution time depends on demands, signal parameters, optimization parameters in CPLEX or Gurobi, respectively, and destroying symmetry of the cyclic network, e.g., by fixing an offset of one intersection. An exemplary solution is shown in Figure 11.

The path-time diagrams in Figure 11 indicate that road users on the Bohlweg (orange) may pass every signal with a speed of 30 miles per hour (50 km/h) without stopping. Also for the purple commodity, the second signal becomes green exactly at the arrival of the first flow. The main bottleneck are the pedestrian crossings. Here, the blue commodity has to wait at signal 6, but signals 3, 7, and 4 can be passed without stopping. Green splits are automatically adapted to flow values. Summarizing, our approach can be used to optimize traffic signal settings and to find good coordinations of signals.

5.2 Cottbus

Whereas the first example does not allow any route choice, the second scenario is based on a much larger network. Cottbus is a German town just south of Berlin with about 100,000

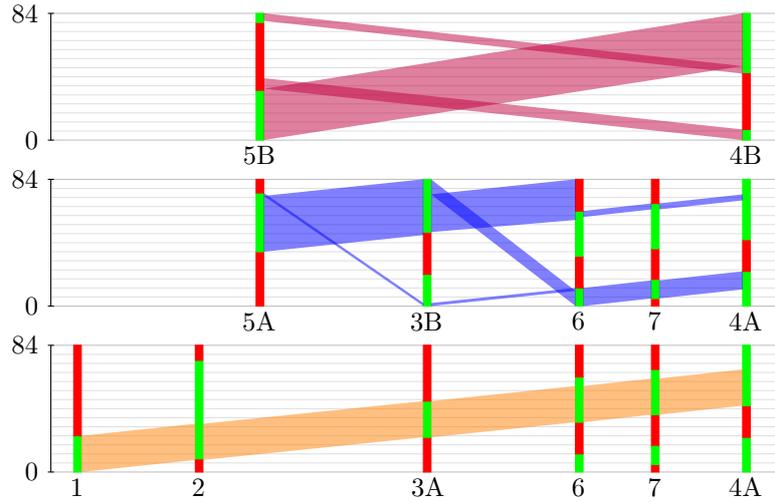


Figure 11: Optimized green bands for the Brunswick scenario. Time and state of the signals is shown on the vertical axis. Signals are labelled with intersection numbers and signal groups, e.g., signal 3A and 3B must not be *green* at the same time. Distances on the horizontal axis are chosen with respect to transit time, the slope of the parallelograms is chosen with respect to free speed. Note that the purple commodity is travelling from right to left and the cyclic overflow is visualized by a negative slope.

inhabitants. The scenario consists of the whole inner-city with an area of 10 square kilometers and 32 signalized intersections. The scenario is presented in Figures 12 to 15.

The network itself was created with data from www.openstreetmap.org. Very small streets, traffic-calmed areas, dead ends, etc. were removed. Traffic demand, i.e. origin-destination pairs, was generated based on anonymised data obtained from the German employment agency. We have access to aggregated home addresses and workplaces of all employees in the area of Cottbus and the surrounding rural district. This allows a good approximation of traffic in the morning and afternoon peak. We reduced the number of commodities by aggregating commodities with similar origins and destinations. That is, we merged all points of interest within a certain neighborhood to a single terminal. Furthermore, we only consider commodities where the shortest path connection passes through the city center. This yields about 250 commodities.

For our first experiments, we used 17 out of these 250 commodities to save computation time. However, demand was scaled back to 100 percent. We also limited our model to offset optimization, all split times and phase orders were fixed to measured values. We started with an already optimized signal setting for a fixed assignment, i.e. the signals provide minimum delay for this fixed assignment. This assignment was computed by VISUM (ptv AG, <http://www.ptvgroup.com>) and we applied our mixed integer program without rerouting to optimize the signal settings. Afterwards, our approach could reduce waiting time in the network by another 11 % by rerouting traffic flow and

resetting traffic signals.

An interesting property of our optimized solutions is the creation of circulations around the traffic center (see Figure 13). This omits a lot of conflicts between opposing traffic streams and allows progressive signal settings for many road users. The computational results were verified by simulation with VISSIM (ptv AG, www.ptvgroup.com) and MATSim (www.matsim.org).

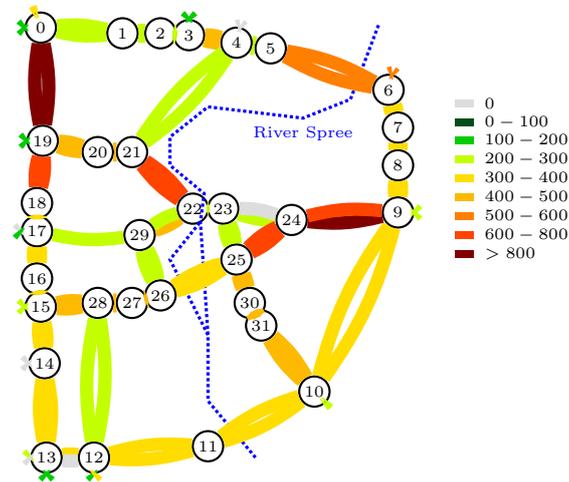


Figure 12: Traffic assignment calculated with VISUM for 17 commodities in the Cottbus scenario. The color encodes the traffic density from dark green (up to 100 cars per hour and lane) to dark red (more than 800 cars per hour and lane) separately for both directions. It can be observed that the traffic is distributed fairly equally in the network. Only a few commodities deviate from their shortest paths due to exhausted capacities. Some changes in traffic density are caused by capacity changes, i.e., a varying number of lanes.

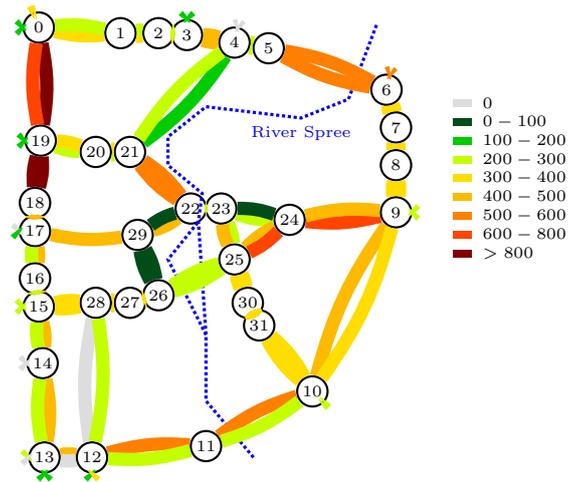


Figure 13: Traffic assignment for the same scenario as in Figure 12, calculated with the cyclically time-expanded model. The traffic on the outer road is increased in clockwise direction.

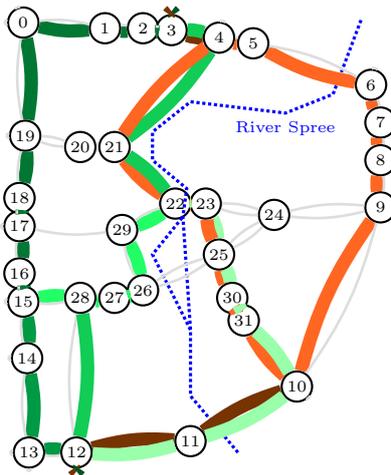


Figure 14: Splitting of two commodities in the scenario of Figure 13. 400 road users per hour travel from a shopping mall in the North (intersection 3) to a housing area in the South (intersection 12). Their paths are displayed in red. The same amount of cars is traveling in the opposite direction (green paths). VISUM assigns all flow units to the shortest path which runs straight through the city center (not shown in the diagram). In our model more than 50% of the road users join the circulation on the outer road. Here, the color brightness encodes the traffic density (dark means higher density).

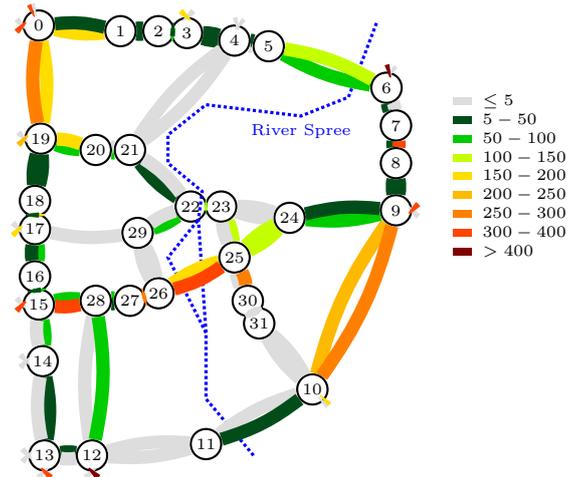


Figure 15: This diagram shows the waiting time integrated over all cars during one cycle at each traffic signal after optimization with our model. Dark green means total waiting time of less than 50 seconds, dark red stands for more than 400 seconds at the specific traffic signal. With the assignment in Figure 13 in mind the waiting time on streets with high traffic volume is rather low. A good coordination is found. The highest waiting times occur at the incoming roads from rural areas where traffic is equally distributed over time.

Unfortunately, a gap of 20 percent between the obtained solution and the lower bound remains when solving this scenario with CPLEX or Gurobi even after hours of computation. Hence, we cannot prove optimality and 20 percent leaves space for improvements. However, experiments with smaller scenarios showed that the dual bounds are much harder to compute. That is, the primal solution converges faster towards the optimal solution than the dual bound does. Thus, the gap between the obtained solution and the optimal solution is most likely much smaller than 20 percent.

In contrast, other approaches for traffic signal optimization and traffic assignment use genetic programming or similar heuristics. Thus, these approaches do not provide such duality gaps at all.

Concluding, one of the main intentions of our model – feedback between traffic assignment and coordination during the optimization process – is achieved. A conventional approach would first determine the assignment, afterwards it would optimize the traffic signals. This could lead to a very uniform distribution of traffic in the network.

In contrast, the computed solutions of the simultaneous optimization with our model are quite different. Traffic is assigned to a few roads. Consequently, the flow rate is high on these roads. However, these routes automatically get very high priority in the signal optimization subproblem and traffic is concentrated in a small number of dense platoons. In artificial scenarios [32], we could save up to 75 percent waiting time by simultaneous optimization compared to the consecutive approach.

These completely different but high-performance assignments characterize the pre-

sented model. We believe that similar efficient results cannot be obtained without simultaneously considering assignment and coordination, i.e., a highly-adaptive traffic signal optimization will have problems to find a competitive solution. These results also suggest that traffic signal coordination could be used to actively route or redirect traffic.

5.3 System optimum vs. user equilibrium revisited

In this section, we will compare the computational results of the cyclically time-expanded network model with simulation results obtained with MATSim.

MATSim is a multi-agent traffic simulation. It can simulate user equilibrium traffic flows with an iterative approach consisting of simulation and replanning rounds. Therefore, each agent has its own set of plans. A single plan consists of activities, routes, and departure times. Furthermore, a plan executed in a simulation round is evaluated and gets an individual score. This score includes travel times and other parameters with arbitrary weighting. New plans can also be generated via shortest paths calculations in the network with the current travel times. In the replanning phase, the plans for execution in the next simulation round are chosen. A certain percentage of the agents simply chooses the plans with the best score so far, whereas the other agents try a randomly chosen alternative plan. After several rounds of simulation and replanning, this process converges to a stable state. Since each agent has its own set of plans with individual scores, this state can be accepted as a user equilibrium.

We consider the Cottbus network. Out of the 250 commodities, we chose 54 commodities with the highest demands which add up to about 50 % of the total demand. At first, we optimized traffic and signals with our approach. Afterwards, we used the obtained signal settings and the same origin-destination pairs to compute user equilibrium assignments with the multi-agent simulation MATSim. Furthermore, we fixed traffic signal offsets to random values to compare assignments only. Hence, the base for the results presented in Table 1 differs in two aspects. Firstly, we compare calculated results to measured results. Secondly, we compare system optima to user equilibria.

scenario	cyclically expanded network	MATSim simulation	Gap (%)
opt	1,103,380	1,140,077	3.33
best random	1,156,888	1,173,303	1.41
med random	1,197,666	1,210,419	1.06
avg random	1,198,688	1,236,851	3.18
worst random	1,285,766	1,291,993	1.00

Table 1: Comparison of system optimal solutions of the cyclically time-expanded network model and user equilibrium solutions of MATSim. All values represent total travel times in seconds for the morning peak from 6:30 to 9:30 am. Besides the optimized traffic signals (opt), we also compare optimized assignments for 100 random traffic signal settings.

First of all, the optimized solution and the random signal settings appear in the same

order for both calculation and simulation. But surprisingly, the difference between the travel times in the cyclically time-expanded network and in MATSim is really small, i.e., one to four percent. Since no calibration was applied and some simplifying assumptions were made in the developing of the cyclically time-expanded model, this suggests that the gap between system optimum and user equilibrium is almost negligible in this scenario.

Furthermore, the optimized signal settings of the system optimal assignment also improve the value of the user equilibrium by three percent compared to the best random solution. Thinking in absolute values, optimized offsets can already save about 10 hours of delay in every peak compared to a good random traffic signal setting in this scenario. Extrapolating this result to full demand, two peaks per day, and 250 working days per year, this adds up to about 10.000 hours per year. Even for a small city like Cottbus, where traffic congestion is not a serious problem, this is a worthwhile economic gain.

Additionally, the consecutive approach of first computing an assignment and afterwards optimizing traffic signals which was used as our benchmark described in the previous subsection yields a total travel time of 1,172,967 seconds. Thus, the consecutive approach is worse than the best random guess.

6 Conclusions

In this paper, we focused on travel times in our cyclically time-expanded network. As a main insight, we have shown that even a completely linear model is capable of creating complex travel time functions. This is achieved at the cost of a much higher dimension, i.e., much more variables are used to describe traffic flow on a single link. Simulation results show that the cyclically time-expanded model is realistic and can capture much more properties of inner-city traffic than static link performance functions can do.

Even better, the cyclically time-expanded model can be used to optimize traffic assignment and traffic signal settings simultaneously. This yields very efficient solutions compared to iterative approaches. Furthermore, the model allows strict mathematical programming. Thus, we can use solvers like CPLEX, which provide a guarantee on the optimality of a solution. This is a considerable advantage compared to other approaches like TRANSYT and others, which use genetic programming or similar heuristics. These tools produce very good solutions, but they do not provide an estimate of the gap towards the optimal solution.

After this proof of concept, our further research focuses on the applicability of our model. Thus, we will try to develop branch&bound strategies or to adapt cutting plane techniques to accelerate the optimization process. There are some other shortcomings in the modeling. For example, we cannot guarantee the First In First Out property in every situation. Besides this rather technical problems, studying user equilibria in our model is also interesting from a theoretical as well as from a practical point of view. One main question is the influence of traffic signal optimization on user equilibria. In a network without signals, system optimum and user equilibrium may look quite different. Is our method of simultaneously optimizing signals and assignment capable of pushing the equilibrium flow towards the system optimum flow by privileging the routes of the

system optimum with progressive signal settings and penalizing the bad links of the user equilibrium with high waiting times?

Acknowledgments

The authors thank Gregor Wünsch and Christian Liebchen who were involved in the first ideas of the cyclically time-expanded model. We also thank Kai Nagel and his group at TU Berlin as well as Klaus Nökel (ptv AG) for providing various insights into traffic simulation concepts and the corresponding simulation tools. This work was supported by the German Research Foundation (DFG, grant number KO 2256/2-1).

References

- [1] R. K. Ahuja, T. L. Magnanti and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, 1993.
- [2] R. E. Allsop and J. A. Charlesworth. Traffic in a signal-controlled road network: An example of different signal timings inducing different routeings. *Traffic Eng. Control*, 18(5):262–265, 1977.
- [3] M. G. H. Bell and H. Ceylan. Traffic signal timing optimization based on genetic algorithm approach, including drivers’ routing. *Transportation Research Part B*, 38:329–342, 2004.
- [4] M. van den Berg, B.D. Schutter, J. Hellendoorn, and A. Hegyi. Influencing route choice in traffic networks: a model predictive control approach based on mixed-integer linear programming. In *17th IEEE International Conference on Control Applications*, pages 299–304, 2008.
- [5] D. Braess. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12:258–268, 1968.
- [6] S. Chiou. Joint optimization for area traffic control and network flow. *Computers and Operations Research*, 32:2821–2841, 2005.
- [7] Y. Chiu, J. Bottom, M. Mahut, A. Paz, R. Balakrishna, T. Waller, and J. Hicks. A primer for dynamic traffic assignment. Technical report, ADB30 Transportation Network Modeling Committee of the Transportation Research Board, 2010.
- [8] J. R. Correa, A. S. Schulz and N. E. Stier-Moses. Selfish Routing in Capacitated Networks. *Mathematics of Operations Research*, 29(4):961–976, 2004.
- [9] J. de Dios Ortúzar and L. G. Willumsen. *Modelling Transport*. Wiley, Chichester, 2006.
- [10] T. J. Dickson. A note on traffic assignment and signal timings in a signal-controlled road network. *Transportation Research Part B*, 15B(4):267–271, 1981.

- [11] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.
- [12] S. Fortune, John Hopcroft, and J. Wyllie. The directed subgraph homeomorphism problem. *Theoretical Computer Science*, 10:111–121, 1980.
- [13] B. Friedrich. Adaptive signal control: an overview. In *Proceedings of the 9th Meeting of the Euro Working Group Transportation*, Bari, 2002.
- [14] M. R. Garey and D. S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, 1979.
- [15] N. H. Gartner, J. D. C. Little, and H. Gabbay. Optimization of traffic signal settings by mixed-integer linear programming, Part I: The network coordination problem. *Transportation Sciences*, 9:321–343, 1966.
- [16] N. H. Gartner, C. J. Messer, and A. K. Rathi, editors. *Revised Monograph on Traffic Flow Theory: A State-of-the-Art Report*. Turner-Fairbank Highway Research Center, Federal Highway Administration, U.S. Dept. of Transportation, 2005. Available at <http://www.tfhrc.gov/its/tft/tft.htm>.
- [17] Olaf Jahn, Rolf H. Möhring, Andreas S. Schulz, and Nicolás E. Stier-Moses. System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations Research*, 53(4):600–616, 2005.
- [18] E. Köhler, R. H. Möhring, K. Nökel, and G. Wünsch. Optimization of signalized traffic networks. In W. Jäger and H.-J. Krebs, editors, *Mathematics—Key Technology for the Future*, pages 179–188. Springer, 2008.
- [19] E. Köhler and M. Skutella. Flows over time with load-dependent transit times. *SIAM Journal on Optimization*, 15(4):1185–1202, 2005.
- [20] E. Köhler and M. Strehler. Traffic signal optimization using cyclically expanded networks. In Thomas Erlebach and Marco Lübbecke, editors, *Proceedings of the 10th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, OpenAccess Series in Informatics (OASiCS), 2010.
- [21] E. Köhler and M. Strehler. Combining Static and Dynamic Models for Traffic Signal Optimization: Inherent Load-dependent Travel Times in a Cyclically Time-expanded Network Model. Traffic signal optimization using cyclically expanded networks. In Proceedings of {EWGT2012} - 15th Meeting of the {EURO} Working Group on Transportation, Procedia - Social and Behavioral Sciences, volume 54, pages 1125–1134, 2012.
- [22] MATSim home page. <http://www.matsim.org/>. Last accessed August 4th, 2015.
- [23] K. Nagel and G. Flötteröd. Agent-based traffic assignment: Going from trips to behavioral travelers. In *Proceedings of the 12th Conference of the International Association for Travel Behaviour Research (IATBR)*, Jaipur, India, 2009.

- [24] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang. Review of road traffic control strategies. In *Proceedings of the IEEE*, volume 91 (12), pages 2043–2067, 2003.
- [25] D. I. Robertson. TRANSYT method for area traffic control. *Traffic Engineering & Control*, 10:276 – 281, 1969.
- [26] T. Roughgarden. *Selfish Routing and the Price of Anarchy*. MIT Press, Cambridge, 2005.
- [27] P. Serafini and W. Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581, 1989.
- [28] Y. Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice Hall, 1984.
- [29] M. Skutella. An introduction to network flows over time. In W. Cook, L. Lovász, and J. Vygen, editors, *Research Trends in Combinatorial Optimization*, pages 451–482. Springer Berlin Heidelberg, 2009.
- [30] M. Smith. Bilevel Optimisation of Prices and Signals in Transportation Models In , and M. Smith, editors, *Mathematical and Computational Models for Congestion Charging*, pages 159–200, Springer Science+Business Media, 2006.
- [31] M. Smith. Traffic signal control and route choice: A new assignment and control model which designs signal timings *Transportation Research Part C*, 2015.
- [32] M. Strehler. *Signalized Flows – optimizing traffic signals and guideposts and related network flow problems*. PhD thesis, Brandenburgische Technische Universität, Cottbus, Germany, 2012. <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:kobv:co1-opus-23637>.
- [33] W. Y. Szeto and H. K. Lo. Dynamic traffic assignment: Properties and extensions. *Transportmetrica*, 2(1):31–52, 2006.
- [34] F. Teklu, A. Sumalee, and D. Watling. A genetic algorithm approach for optimizing traffic control signals considering routing. *Computer-Aided Civil and Infrastructure Engineering*, 22:31–43, 2007.
- [35] VISSIM page. <http://vision-traffic.ptvgroup.com/en-uk/products/ptv-vissim/>. Last accessed August 4th, 2015.
- [36] G. Wunsch. *Coordination of Traffic Signals in Networks*. PhD thesis, Technische Universität Berlin, 2008.
- [37] H. Yang and Q. Meng. Departure time, route choice and congestion toll in a queuing network with elastic demand. *Transportation Research Part B*, 32(4):247–260, Elsevier, 1998.