Self-Stabilizing Decentralized Signal Control of Realistic, Saturated Network Traffic

Stefan Lämmer
Dirk Helbing

SFI WORKING PAPER: 2010-09-019

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author’s copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu
A coordination of vehicle flows is usually reached by a cyclical operation of traffic lights, and by synchronizing these cycles. The typical conditions, however, for which traffic lights are normally optimized for, never occur exactly. Large fluctuations in the number of vehicles arriving during one cycle time may lead to an inefficient usage of green times, which are often either too short or too long. The method we propose here allows for variable adjustments not only of the duration, but also of the order of green phases, while it reaches at least the same intersection throughput capacity as an optimized fixed-time controller. This is particularly important, when intersections are highly saturated, road networks are heterogeneous, or if public transport is to be prioritized.

We reach the stabilization of queues and red-time durations by a decentralized supervision of (potentially unstable) locally optimizing traffic light controllers. The proposed supervisory concept makes sure that all network flows get a green light regularly and long enough. In addition, it enables flexible responses to local fluctuations in the demand, and it favors a self-organized coordination of traffic flows. In a simulation of the city center of Dresden, Germany, we could compare the proposed concept with an adaptive state-of-the-art controller (which has been optimized within the same simulation suite and includes green waves). Results indicate that not only the mean value of travel times is reduced, but also in their variance, which is positive for the reliability of public transport and individual trip scheduling.

**Key words**: traffic signal control; dynamic instability; public transport prioritization; self-organization; decentralized coordination; intelligent transportation system
to respond to random or irregular variations in demand quickly enough. Due to stochastic fluctuations, even optimized green times are usually either too short (creating multiple red lights for queued vehicles that could not be served) or too long (creating unnecessary delays to vehicles of other flow directions). Traffic-responsive control strategies are essential to overcome these problems and to reduce travel times, fuel consumption, and the emission of pollutants effectively.

1.2. Traffic-responsive Strategies

Also when public transport is prioritized, it is not sufficient to operate traffic lights according to pre-specified signal plans. In order to react to prioritization requests by public transport vehicles that serve a large number of passengers, traffic control must enable a high degree of flexibility and responsiveness. According to the Transit Signal Priority Handbook (Smith et al. (2005)) and others (Skabardonis (2000)), a pre-condition for an active and well-balanced prioritization of public transport is that traffic lights facilitate variable green times and a variable selection of green phases.

In the last years, a huge variety of traffic-responsive control strategies has been proposed. One of the most widely used approaches is called rolling-horizon optimization, where a repetitive re-optimization of the signal switching sequences shall enable an adaptation to changing traffic conditions. It is applied in several traffic light controllers, e.g. OPAC (Gartner (1983)), PRODYN (Henry et al. (1983)), UTOPIA (Mauro and Taranto (1990)), CRONOS (Boillot et al. (1992)), ALLONS-D (Porche et al. (1996)), and RHODES (Mirchandani and Head (2001)). They mainly differ in how the optimization problem is solved: OPAC basically enumerates the solution space, PRODYN applies an efficient heuristic, and ALLONS-D searches a complex decision tree using back-tracking. The associated dynamic programming techniques could be further improved if combined with online learning techniques, as Cai et al. (2009) propose. The resulting signal plans are, in general, acyclic with no fixed order of green phases. Another real-time strategy is TUC (Papageorgiou et al. (2003), Aboudolas et al. (2009)), which is based on a store-and-forward modeling approach and tries to balance queue lengths by flexible adjustments of green times. Alternative approaches treat intersections as autonomous, cooperative agents in a multi-agent system (France and Ghorbani (2003)), for an overview see Bazan (2009). Other approaches are based on autonomic and organic computing (Prothmann et al. (2009)), where a coordination of the traffic lights is reached by local communication among neighboring intersections. A similar synchronization principle, in which intersections are treated as dynamically coupled oscillators, was proposed by Lämmer et al. (2006).

Furthermore, fuzzy-logic-based traffic signaling strategies have been suggested in order to cope with imprecise, uncertain, or ambiguous information (Trabia et al. (1999), Rahman and Ratrout (2009)). Genetic algorithms are used to apply multi-goal optimization, focusing on the minimization of emissions and fuel consumption (Park et al. (2009)) or on a dynamic re-optimization of green-wave-programs (Braun et al. (2008)).

Instead of evaluating complex decision trees and different switching sequences, an alternative optimization method was recently proposed, which is based on a dynamic prioritization of traffic flows (Lämmer and Helbing (2008)). The formula, according to which the priority $\pi$ of each flow is calculated, considers the queue lengths as well as the anticipated arrival times of vehicles, which are determined from the measured inflows into road sections and the outflows from them (Lämmer et al. (2007)). Assuming that there are no other than the anticipated arrivals, the ranks of the $\pi$-values define the order of switching of the traffic lights. To minimize the total delay of all vehicles at an intersection, the controller simply has to give a green light to the flow with the highest priority.

When applying this local optimizing technique, signal plans are not based on cyclic control schemes determined from average traffic conditions, but they respond instead to actual real-time detector data. This makes the traffic control more flexible with respect to local
demands and more robust to perturbations (such as variations in the arrival flows, turning rates, accidents, building sites, or special events).

1.3. The Price of Flexibility
The introduction of flexibility and traffic-responsiveness, unfortunately, contradicts strategic concepts such as the idea of a network-wide coordination. Friedrich (2007) calls this the “jungle-principle”, when each intersection in the network controls traffic as is seems best from its local perspective. From a game-theoretical point of view, selfish optimization does not necessarily establish the system optimum – it could even create a poor performance as is known from the “tragedy of the commons” (Lämmer and Helbing (2008)). If traffic lights mainly respond to local demands, each intersection just reacts to the traffic coming from the neighboring intersections, which consequently makes the dynamic inter-dependencies in the network impossible to predict. We agree with Papageorgiou et al. (2003) that “the properties of a completely decentralized operation (e.g., independent algorithm application at each intersection) are currently not fully analyzed or understood”. Generally, cyclic control schemes tend to lack flexibility, while acyclic schemes tend to lack coordination, despite best efforts. Both lead to efficiency losses, which have prevented a widespread practical application of locally optimizing strategies so far. Without their appropriate improvement, the inefficient usage of intersection capacities can cause situations, in which vehicle queues will eventually grow longer and longer as depicted in Fig.1.

1.4. Instability Problems
When locally optimizing traffic lights fail to serve each traffic flow regularly and long enough, instability problems are likely to occur. Reasons for an ineffective usage of intersection capacities are manifold. If traffic lights switch too often, for example, this may result in a lower throughput as there is no service during switching time periods, also called setup times (Duenyas and van Oyen (1996)). But even if setup times are not a problem, a locally optimizing controller could assign too long green times to highly prioritized traffic flows and, eventually, marginalize side flows by too long red times. Moreover, the unpredictability of traffic can systematically mislead short-sighted optimization algorithms, especially at higher saturation levels or if the intersections. For material flow networks, Kumar and Seidman (1990) have discovered that clearing strategies of distributed controllers may systematically reduce the system throughput if demand is high, leading to a growth of queues and service times. In the Appendix, we prove this analytically for a road network with two independently controlled intersections.

According to the definition of Perkins and Kumar (1989), a controlled queueing network is stable, if all queues lengths remain bounded for all times. Of course, all traffic jams will resolve sooner or later, e.g. at night or when drivers choose alternative routes. The notation of stability, however, allows us to distinguish between those cases, in which congestion has a tendency to grow and those cases, in which the controllers succeed to avoid growing congestion or to dissolve it after a while. Of course, no controller can prevent queues from growing if saturation is too high. It would be reasonable, however, to demand bounded queues in all situations, in which conventional fixed-time programs exist for stationary inflows. We will therefore use the following terminology: As soon as there are traffic demands, for which the vehicle queues remain bounded with a fixed-time control, but not if operated with a locally optimizing controller, the latter is “unstable” and potentially inferior.
It seems hopeless to identify all possible sources of instability in a road network of reasonable size, and there is no chance to implement specific counter-measures for all of them. Instead, we will seek for a general decentralized stabilization mechanism that maintains the performance of the road network even when each intersection is controlled locally.

1.5. Outline

The paper is structured as follows. In Sec. 2 we develop a distributed stabilizing mechanism, which assists a locally optimizing controller in providing green times frequently and long enough. This allows an underlying locally optimizing controller to handle the same amount of traffic as a designated fixed-time controller, but also facilitates flexible responses to local fluctuations in demand. The parameters of the proposed mechanism are discussed in Sec. 3. In Sec. 4, we test our stabilization method with a realistic network simulation of the city center of Dresden, Germany, and compare it to the currently implemented state-of-the-art controller. As our results indicate, locally optimizing traffic lights can, if well stabilized, lead to a superior performance for all modes of transport.

Compared to previous work, Lämmer and Helbing (2008), the current paper presents an extension of the stabilization mechanism as well as the results of a realistic network simulation including public transport prioritization and pedestrian flows.

2. Stabilization Mechanism

When it is possible to operate a network with stable fixed-time programs, why should it not be possible to operate the intersections with properly designed traffic-responsive strategies and to serve the same amount of traffic frequently enough? Moreover, why should it not be possible to utilize excess capacities for local optimization by means of green time extensions or additional green phases?

2.1. Demand-Driven Service Concept

For fixed-time controllers, a desired throughput can be achieved by adjusting cycle times and splits accordingly. Although offsets and coordination are important for service quality and waiting times as well, we will neglect these for the time being. Let us instead discuss the essentials of throughput and stabilization, and derive an alternative service concept, into which we will afterwards incorporate aspects of optimization.

When the length and frequency of green times is fixed, but the number of vehicles arriving during one cycle time is not, the efficiency of the green times varies randomly. Therefore, stochastic variations in the arrival and turning rates can decrease the efficiency of fixed-time controls considerably (van den Broek et al. (2006)). To overcome this, we propose to modify the underlying operation principle as follows (see Fig. 2): Instead of waiting for a certain point in time before switching to green, we now wait for a critical number of vehicles ready for service at maximum rate which given by the saturation flow, i.e. the outflow from congested traffic. The critical number of vehicles, our demand-driven service concept is waiting for before giving a green light, is specified by a threshold function. For example, if the critical number corresponds to the average number of vehicles arriving during one cycle time, the arriving vehicles will accumulate up to this number and initiate a green light once within a cycle time on average.

Before we discuss the properties of the flexible control principle in more detail, let us briefly explain how the proposed concept can be used in combination with locally optimizing traffic light controllers to stabilize them. By giving a green light after a critical number of vehicles is waiting for service, red lights are terminated.
when this is needed to clear the queues regularly. In other words, a green light is given whenever there is a definite demand for it. This particular property makes the proposed concept suitable as a stabilizing supervisor, which assists a (possibly unstable) locally optimizing traffic light controller. The supervisor determines what traffic flows need to get a green light before which point in time and, thereby, formulates stabilizing constraints. Stable and regular behavior is guaranteed, if the local optimization of green times is subject to these constraints. Note that the locally optimizing controller is still free to start green times earlier or to extend them if possible. As long as it assigns the green times not later and not shorter than the stabilizing supervisor would assign them, the service intervals as well as the queue lengths are kept bounded.

2.2. Supervisor

As indicated in Fig. 3, the stabilizing supervisor can be understood as a separate process that observes the current traffic state and puts constraints on the signal timing of the locally optimizing controller. Supervisory control has been established as a concept to stabilize nonlinear systems. Its purpose is to identify critical states of the system and to provide appropriate measures driving the system back into a desired state. Previous work has applied distributed supervisory principles to stabilize material flow networks (for an overview see Bramson (2008) and references therein). Note, however, that applying these principles to traffic control would not limit red light durations and service intervals, which makes their direct transfer useless. In contrast, the supervisory mechanism we propose in this paper ensures that all traffic flows are served frequently and long enough.

We formulate the supervisor rules as follows:

1. A traffic flow is identified to be critical, if its corresponding point in the state space of Fig. 2 exceeds the threshold function (to be defined in Sec. 2.3). In that case, the flow is served as soon as possible.
2. If there is more than one flow in a critical state, these flows have to be served in the order in which they became critical.

3. A traffic flow becomes uncritical again when it has been served by a green time that an associated fixed-time program would attribute to it or when its queue has been cleared before that.

4. If no flows are in a critical state, the controller performs a local optimization. A fifth rule is discussed at the end of Sec. 2.3.

These rules are independent of the underlying optimization procedure. Hence, the stabilizing supervisor can be combined with any local controller, which can be subject to the constraints defined by these rules. Depending on the particular implementation, signal timings that lead to critical states can be penalized in the goal function, for example, or just excluded from the decision tree. As the points in time, at which the vehicle queues become critical, can be anticipated relatively early, the controller can harmonically impose its constraints on the optimization procedure. To give an example, instead of serving a minor flow exactly at the same time when it becomes critical, the optimization strategy can also choose to schedule it earlier, if this helps to avoid conflicts with large vehicle platoons on other streets.

For a large range of traffic demands, we achieve a load-dependent interplay between optimization and stabilization. The difference between the intersection capacity and traffic demand is not necessarily needed to maintain stability. This so-called “excess capacity” can be freely used for green time extensions or additional switching operations by the locally optimizing controller. If an intersection becomes more saturated, however, the supervisory mechanism steers the optimization strategy tighter. In case of over-saturation, the optimizing controller has no freedom anymore, and it must serve the flows according to an associated fixed-time program.

In summary, if supervised, a locally optimizing controller serves the same amount of traffic as a fixed-time controller and can, in addition, minimize travel times and queues due to flexible adjustments to instantaneous local demands. The results in Fig. 4 indicate a significant improvement of the proposed strategy compared to fixed-time or non-stabilized controllers.

2.3. Choice of the Threshold Function

The choice of the threshold function in Figs. 2 and 5 has a major impact on the regularity of service intervals. According to the supervisor rules discussed above, the service of a traffic flow is triggered as soon as its corresponding point in the state space $(\hat{z}, \hat{n})$ exceeds the threshold function. In order to derive a suitable specification of the threshold function, we first assume that the average arrival rate of that traffic flow as well as the green times of an associated stable fixed-time program are given. We require that

(i) each traffic flow shall be served once, on average, within a desired service interval $Z$, and

(ii) each traffic flow has to be served once, at least, after a maximum red time period $Z^{\max}$.

The first criterion shall ensure a desired level of regularity, and the second one guarantee the compliance with maximum red light durations for the sake of road safety.
The supervisory mechanism needs to anticipate the service interval \( \hat{z} \) as well as the number \( \hat{n} \) of vehicles to be served within a subsequent green time. With suitable procedures, their values can be anticipated already during the preceding red time period Lämmer et al. (2007), Helbing et al. (2007). Both variables, \( \hat{z} \) and \( \hat{n} \), refer to the earliest time point \( t' \) at which the corresponding vehicle queue could be cleared. The service interval \( \hat{z} \) corresponds to the time interval since the end of the last green time until this time point \( t' \). Likewise, \( \hat{n} \) represents the number of vehicles that need to be served until the queue is cleared, considering new arrivals of vehicles during the service process. The interrelation between these two variables is that \( \hat{n} \) captures the number of vehicles that arrive during the service interval \( \hat{z} \). Also note that the green time required to resolve the queue increases with every vehicle arriving during the service process. This implies several interesting features. The arrival of a vehicle platoon, for example, lets \( \hat{n} \) abruptly jump to a higher value, since it becomes possible to serve considerably more vehicles at a maximum rate. If that jump in \( \hat{n} \) triggers the green light, the platoon is served without being stopped as it is intended by “green wave” control. Based on these anticipative characteristics, the proposed concept favors flexibly coordinated, platoon-oriented service patterns.

According to the rules formulated in Sec. 2.2, a traffic flow should be served by a green light as soon as there is a critical number \( n(\hat{z}) \) of vehicles ready to be served at maximum rate, e.g.

\[
\hat{n} \geq n(\hat{z}). \tag{1}
\]

Herein, \( n(\hat{z}) \) denotes the threshold function depicted in Fig. 5. Let us now discuss three important constraints for the choice of \( n(\hat{z}) \). First of all, it must be ensured that every traffic flow will eventually become critical as red time increases, and that it cannot become uncritical before the vehicle queue is served. Since \( \hat{z} \) and \( \hat{n} \) are both non-decreasing during red time periods, the threshold function \( n(\hat{z}) \) must be strictly decreasing in order to be crossed by the state trajectory \((\hat{z}, \hat{n})\) exactly once, which implies

\[
dn/d\hat{z} < 0. \tag{2}
\]

Second, the threshold function \( n(\hat{z}) \) shall be exceeded when there are as many vehicles ready to be served at maximum rate as are arriving on average within the desired service interval \( Z \). In the following, we identify \( Z \) with the cycle time of the associated stable fixed-time program. Then the specification

\[
n(Z) = qZ, \tag{3}
\]

where \( q \) denotes the average arrival rate, ensures that the service interval will, on average, be \( \hat{z} = Z \) as required.

Above we have discussed that it is favorable to have a variable service interval \( \hat{z} \) in order to be able to adapt to irregular inflows. For example, our mechanism delays the start of a green time period automatically, if vehicle platoons have been delayed at an upstream intersection. In order to ensure that a maximum red time period \( Z_{\text{max}} \) (with \( Z_{\text{max}} > Z \)) is never exceeded, even not in the absence or non-detection of vehicle arrivals, we finally require the threshold function to meet the condition

\[
n(Z_{\text{max}}) = 0. \tag{4}
\]

That means, \( n(\hat{z}) \) falls below zero as the service interval \( \hat{z} \) exceeds \( Z_{\text{max}} \). This makes \( Z_{\text{max}} \) the latest possible time point after which a traffic flow becomes critical and, therefore, is served. In order to fulfill road safety regulations, one can extend the previously discussed supervisory rules by requiring that the service of flows with \( \hat{z} > Z_{\text{max}} \) must be prioritized to the service of other critical flows.

3. Parameter Choice

The parameters of the proposed stabilization mechanism are the desired service interval \( Z \), the green times of a stable fixed-time program with cycle time \( Z \), as well as the maximum red time period \( Z_{\text{max}} \).

3.1. Desired Service Interval

The choice of the service interval \( Z \) determines how frequently all traffic flows are served, even in cases of over-saturation, large variations in the arrival flows or frequent prioritization requests of public transport vehicles. In such
cases, the supervisor will enforce each traffic flow to be served once, on average, within the time period $Z$. The green times and, thus, the service capacities reserved for the traffic flows correspond to those of the associated fixed-time program. Therefore, our decentralized control approach reaches the same service capacity in saturated flow conditions.

### 3.2. Green Time Durations

Modern urban planning aims at a bundling of traffic on arterial roads and to restrict traffic in residential areas, which is classically reached by a centrally imposed allocation or restriction of road capacities through green times (see Wong and Yang (1997), Yang et al. (2000)). However, the same goal can be achieved by decentralized control concepts, if the supervisory mechanism is parameterized as follows: Minor flows are given green times just long enough to satisfy the corresponding average demand, while main flows with a large importance for the overall network traffic are given all the remaining green time. An example of how this helps to prevent spill-overs on main roads and to resolve their queues more efficiently is given in Fig. 6.

Note that the difference between $Z$ and the minimum possible cycle time is the amount of time the locally optimizing controller can use for travel time optimization, e.g. by green time extensions or additional switches. If the traffic demand is not too high, this may lead to short, but frequent green phases. An illustration of how a supervised locally optimizing controller utilizes excess capacities for flexible switching sequences is shown in Fig. 7.

### 3.3. Maximum Red Time Periods

The maximum red time period $Z_{\text{max}}$ applies in cases in which neither the optimizing controller nor the $Z$-mechanism of the supervisor selects each traffic flows frequently and long enough for service. This, typically happens when arrivals are absent or if they are not detected. Then the supervisor initiates the minimum admissible green time to the neglected traffic flow after a maximum red time period of $Z_{\text{max}}$. Fur-
Figure 6 Temporal evolution of queue lengths (left) and corresponding state trajectory (right) at a two-armed intersection with constant arrival rates. Initially, both queues are critical (bold lines), but the stabilizing supervisor manages to reduce the queues until local optimization can take over (thin lines). Top: If the green times are assigned proportionally to the saturation of the flows, both queues resolve equally fast. Bottom: The concept we propose allows for a distinction between streets of different importance for the overall network flows. By assigning only the minimum stabilizing green time to the less important side street (dashed line) and all the remaining green time to the more important two-lane main street (continuous line), the queue in the side street is not resolved before the situation has relaxed on the main street. The used parameters are: $Z = 90$ s, $Z_{\text{max}} = 120$ s, saturation flow rate $S = 1/1.8$ vehicles per lane, constant arrival rate $q_1 = q_2 = 0.4$ vehicles per lane, and setup time $\tau = 5$ s.

4. Real-world Network
4.1. Self-Control
The proposed supervisory mechanism (as presented in Sec. 2) is capable of stabilizing locally optimizing traffic light controllers and to support them by an efficient throughput man-

thermore, the closer the values of $Z$ and $Z_{\text{max}}$ are, the steeper declines the threshold function. Therefore, in the limiting case $Z \to Z_{\text{max}}$, the service concept becomes analogous to a fixed-time controller, compare Figs. 2 and 5. In contrast, when choosing very large values of $Z_{\text{max}}$, the service concept becomes purely traffic responsive. In order to reach a maximum level of traffic-responsiveness, we recommend to set $Z_{\text{max}}$ to the largest acceptable value that is compatible with regulations regarding the maximum red times.

The decentralized stabilizing mechanism presented above stabilizes queue lengths and service intervals at all intersections in a network. While the intersections may still be dynamically coupled by the traffic flows between them, the supervisory control principle guarantees that it does not lead to a loss of throughput capacity (see Appendix). Instead, the dynamic coupling may now trigger coordinated service patterns. In contrast to a pre-optimized coordination based on periodic green waves, the demand-driven service concept favors a self-organized coordination based on local responses to traveling vehicle platoons. Larger platoons are mostly served without stopping, simply because giving them a green light immediately before they arrive at an intersection helps minimizing delays. An illustrative example of platoon propagation along an arterial road is depicted in Fig. 8.
Lämmert and Helbing: Self-Stabilizing Decentralized Signal Control of Realistic, Saturated Network Traffic

Figure 8 Green-bands along an arterial road with 10 irregularly spaced intersections, mutually conflicting traffic flows, and stochastic arrivals at the network boundaries. Left: The fixed-time control implements regular green waves traveling from West to East with a cycle time of 120 seconds. The opposite flow direction (for which the trajectory of a single vehicle is indicated by a black solid line) experiences large delays compared to free traffic conditions (dashed line). Right: The flexible self-control, in contrast, assigns shorter green times at a higher frequency and, thereby, serves the same amount of traffic with significantly reduced average travel times (Lämmert (2007)). The intersection controllers are loosely coupled by vehicle platoons traveling between them. This creates the emergence of coordinated service patterns.

Figure 9

4.2. Area of Investigation

The road network considered in the following consists of 13 traffic-light-controlled intersections (see Fig. 9), which are part of Dresden’s city center. This area was chosen as it experiences unsatisfactory delays and notorious coordination problems despite the use of an adaptive state-of-the-art-control. Two parallel main roads pass the train station “Dresden Mitte” on each side, which is used by more than 13,000 passengers on an average day. No less than 7 bus and tram lines cross the network every 10 minutes in opposite directions (Fig. 10), and 68 pedestrian crossings lie within that area. The intersections are spaced very irregularly. Some of them are closer than 100m. Due to the heterogeneity of transport, the irregularity of the traffic and queueing dynamics in this study have been simulated using the software package PTV Vissim. The algorithms of the self-control mechanism have been implemented in Java. The Java application periodically reads in the vehicle positions and writes back the updated states of the traffic lights using a COM interface. The following comparative analysis is based on simulation runs for a typical afternoon rush-hour. Related simulation videos are available at www.stefanlaemmer.de/selfcontrol.
network topology, and the relatively high traffic load, local traffic authorities agree that this part of Dresden is most challenging to control (see the corresponding statements in Ref. Lämmer et al. (2009)).

Today, the traffic lights in this area are operated with a centralized traffic-adaptive state-of-the-art control system of type VS-PLUS (see www.vs-plus.com). The intersections are coordinated such that green waves propagate on both main arterial roads in both directions, and also on one of the orthogonal roads (Schäferstraße/Schweriner Straße). The common cycle time is 100 s most of the day. For trams and buses, there are two time slots reserved per cycle, in which they can be scheduled upon request. Pedestrians get a green light with a delay and only on demand by pressing a button.

4.3. Results
Note that we are using the same simulation suite and the same empirically measured traffic flow data, for which the currently used state-of-the-art control has been optimized. A comparison with our proposed flexible self-control strategy is particularly interesting, as both traffic light control approaches follow different “philosophies”: The first one aims at globally coordinating network flows, while the second one aims at locally minimizing the waiting times of all modes of transport.
As Fig. 11 shows, the traffic-light-induced delay of trams and buses could be reduced by more than 50 percent. Even cars and trucks, which are currently served by green waves, can travel faster through the network, when the flexible self-control is applied. Pedestrians wait 36 percent less for the next green light.

The significant reductions in waiting times for all modes of transport are also found for different overall traffic demands, as Fig. 12 indicates. This becomes possible due to the more flexible, demand-oriented service concept, as we will explain in the following.

For example, the flexible self-control typically does not stop cars and trucks at the entrance to the arterial roads, while the current system builds up large queues that wait for the next green wave (see Fig. 13). Instead, each intersection serves incoming vehicles as soon and as often as possible, which favors the formation of smaller vehicle platoons. Sometimes, these small platoons are stopped at some downstream intersection, for example, in favor of prioritizing a tram or bus while large platoons are hardly ever stopped. Buses and trams are weighted like 15 vehicles to reach a harmonic prioritization of public transport. Gaps between the platoons are utilized by the locally optimizing controller to serve vehicle flows of side streets, left turners, or pedestrians. This principle may be called “gap” or “chance management”.

Pedestrians get a green light whenever the time gap between the crossing vehicles is large enough. In our model, pedestrian streams are treated as virtual flows, e.g. assuming a pedestrian arrival every 10 seconds. As a consequence, the “pressure” on the traffic lights is permanently increasing during the red time. The distribution of red times for pedestrians is shown in Fig. 14. For the state-of-the-art controller, the distribution shows several prominent peaks, which reflect that the program underlay certain temporal restrictions. In contrast, the proposed flexible self-control can fully adopt to fluctuations in the traffic flows. The resulting red time distribution appears more natural and has, in addition, a significantly smaller mean value.

In order to evaluate how the two control strategies can deal with randomness, we varied the random seed of the simulation. The result of 24 independent simulation runs are plotted in Fig. 15. The variance in horizontal and vertical direction indicates how sensitive each of the strategies is to random fluctuations. The variance of delays is much higher for the state-of-the-art controller (horizontal axis). This can be explained by the dominance of green waves, which leave buses and trams only two time slots per cycle to cross an intersection. If they missed a time slot, they need to wait half a minute or
Figure 13  Vehicle trajectories along Könneritzstraße southbound. With the state-of-the-art control (top), large vehicle platoons are propagated through the network by regular green waves. The price for this is the formation of long vehicle queues at the entrance to the coordinated arterial. Self-controlled traffic lights (bottom), in contrast, can serve the incoming vehicles as soon as possible. This concept favors smaller vehicle platoons, which allows one to solve conflicts with crossing trams or buses in a much more flexible way.

Figure 15  Average delay of public transport vehicles in 24 simulation runs with different random seeds. Public transport delays have a very low variance, if traffic lights are self-controlled (vertical axis), in contrast to the results for the state-of-the-art control obtained for the same random seeds (horizontal axis). That is, buses and trams are not only faster, but they are also on time more often.

5. Summary and Discussion
Due to limited prognosis horizons and dynamic feedback loops in the network, locally optimizing control strategies are facing problems to find optimal solutions for a traffic network. Therefore, we have proposed a decentralized stabilization mechanism which ensures that all traffic flows get at least the same green time as a stable fixed-time schedule would attribute to them. In contrast to a fixed-time controller, however, the green times are requested only when there is definite demand for them, the cycle time is not fixed, and the service is not necessarily periodic. The “cycle time” rather becomes a random variable, which is distributed around a desired service interval $Z$ and limited by a maximum red time period $Z_{\text{max}}$. An implementation of the proposed self-control concept is based on the following elements:

(i) Short-term anticipation of vehicle arrivals by measurements of arrival and departure flows

more reliably. Note that we do not require any sophisticated decision logic to handle all different modes of transport and to reduce the average of the travel times as well as their variance.
of the road sections (see Lämmer et al. (2007)). Where measurements are unavailable, typical flows can be assumed instead.

(ii) Local optimization of vehicle delays based on a dynamic prioritization of traffic flows (see Lämmer and Helbing (2008)).

(iii) Local stabilization based on the supervisory principle presented in Sec. 2. The method can be augmented by real-time simulations of network traffic (see Helbing et al. (2007)), and it can also operate together with existing traffic lights controlled by fixed-time or other control schemes.

The acyclic operation of the traffic lights is an essential feature of the proposed concept, which compensates for random fluctuations in the traffic flows, and which consequently reduces the mean value as well as the variance of vehicle and pedestrian queues and delay times. Other interesting features are:

(i) The traffic light switching sequences are not centrally imposed on the traffic flows, but it is the traffic flows, which determine the prioritization scheme. This is done “on the fly”.

(ii) Instead of implementing a particular event-logic to solve conflicts between different transportation modes, we just attribute higher weights to trams and buses in the goal function, and thereby reach a public transport prioritization, which harmonizes with other flows.

(iii) The service concept is demand-driven: Order, beginning, and duration of green time periods are flexibly adjusted to local traffic conditions.

(iv) Instead of serving each flow exactly once in a cycle time, self-controlled traffic lights tend to give green lights to heavier traffic flows as long as possible and more often.

(v) Growing queues are stabilized to avoid spill-overs (see Mazloumian et al. (2010)).

(vi) The short-term anticipation of gaps and platoons enables a flexible and efficient opportunity management. Pedestrians, for example, get a green light whenever there is a larger gap in the crossing vehicle stream.

(vii) Travel times are reduced on average and become more predictable.

As we could demonstrate in the simulation of a realistic, very irregular road network with 13 traffic-light-controlled intersections and many crossing public transport lines, the proposed flexible self-control mechanism provides a superior performance for all modes of transport as compared to the currently implemented state-of-the-art control.

Our results show that it is crucial to admit a heterogeneity in the service intervals, which is in sharp contrast to the classical synchronization approach, requiring homogeneous (identical) cycle times. This heterogeneity allows to resolve the conflict between two incompatible optimization goals resulting from travel time minimization in the network, namely throughput maximization and avoidance of spill-over effects. These conflicting goals are due to the fact that traffic optimization involves two different kinds of capacities, the maximum departure flow from road sections and their storage capacity for vehicles. Putting it differently, travel time minimization in the network implies a multi-goal optimization problem at the nodes, which cannot be solved in a satisfactory way by a prioritization, superposition, or weighting of several goal functions for a node (or by other classical approaches in multi-goal optimization). It rather requires a suitable switching strategy between both optimization goals, considering that the finite setup time ${\tau}$ discourages frequent switching. It is interesting that classical optimization approaches fail, when applied to the nodes of a flow network with variable arrival flows, since dynamical interaction effects between nodes are highly significant. We believe that this also applies to logistic and production networks (Helbing (2003), Sipahi et al. (2009)), and even to administrative processes. The highest possible system performance can only be reached, when each system element does not behave strictly optimally, but if there is a local variability that allows for a flexible coordination of the interacting elements (Helbing and Lämmer (2005)).

Acknowledgments

The authors are grateful for partial financial support by the German Research Foundation (DFG projects He 2789/5-1,2 and 8-1), the Volkswagen Foundation (project I/82 697), and the Gottlieb Daimler and Karl Benz Foundation (BioLogistics project).
Appendix. Analytical Example

An isolated intersection with two traffic flows and no setup times can be optimally controlled by a clearing policy, i.e. by clearing the queues in an alternating way one after another without green time extensions (see Hofri and Ross (1987), van Oyen et al. (1992), Chase and Ramadge (1992)). This very simple, purely traffic-responsive control strategy provides the intersection with maximum throughput, while is able to dynamically counterbalance inflow variations by variations of green times, i.e. longer queues are given longer green times automatically. These favorable dynamic properties, however, get lost as soon as the intersection is coupled with others in a network.

The following example refers to the simple network of two intersections and two traffic flows shown in Fig. 16. Similar network topologies have been studied by Kumar and Seidman (1990), Rybko and Stolyar (1992) in the context of manufacturing systems. In order to keep the analysis simple, we assume constant arrival rates $q_1$ and $q_2$ and also neglect the travel times between the intersections. The saturation flow rates per lane are denoted by $S$. The network flows can be served in a stable way with a fixed-time program as long as none of the intersections is saturated, i.e. as long as both conditions

$$\frac{q_1}{2S} + \frac{q_2}{S} < 1 \quad \text{and} \quad \frac{q_1}{S} + \frac{q_2}{2S} < 1 \quad (5)$$

hold. This is the case, for example, if we choose $q_1 = q_2 = 0.6S$. In order to show that the alternating clearing of queues may produce dynamic instabilities, we now analyze the evolution of the queues in the network over one switching period. The analysis is similar to the second example in Ref. Kumar and Seidman (1990) and can be followed in Fig. 17. Corresponding simulation videos are provided at www.stefanolammer.de/stability.

Initially, there are $n_1(t_0)$ vehicles waiting on the main street at the left intersection, but nowhere else. According to the switching rule under consideration, the left intersection switches to green for the queue on the two-lane main street. While the queue is resolved within the next $n_1(t_0)/(2S - q_1)$ seconds, vehicles proceed to the second intersection, where they have to merge on the single, left-turning lane. On that lane, a queue of

$$n_2(t_1) = n_1(t_0) \frac{S}{2S - q_1} \quad (6)$$

vehicles builds up until $t_1$ requiring another $n_1(t_0)/(S - q_1)$ seconds to resolve. Note that in time interval $[t_1, t_2]$ a problematic situation occurs: The left intersection has successfully cleared the queue on the main street, when there are no vehicles arriving from the right, while the intersection on the right is still busy serving vehicles turning left on 2'. In the words of Kumar and Seidman (1990), the left intersection “is being starved of input”. Due to the lack of alternatives, the left intersection continues serving the arrivals at a rate lower than the saturation flow rate. Doing so, the left intersection is continuously sending new vehicles to the right intersection and, thereby, delaying the time point at which it will eventually switch to the main street. In the meantime, there are

$$n_1(t_2) = n_1(t_0) \frac{q_2}{S - q_1} \quad (7)$$

vehicles queued up on the main street 2 and nowhere else in the network. Since this is symmetrical to the initial state, we can conclude that, after a full control cycle, there are

$$n_1(t_4) = a n_1(t_0) \quad \text{where} \quad a = \frac{q_1}{S - q_1} - \frac{q_2}{S - q_2} \quad (8)$$

vehicles waiting in queue 1. Obviously, the queues are reduced from one cycle to the next, only if $a < 1$. This condition for the stability of the local clearing policy can be rewritten as

$$\frac{q_1}{S} + \frac{q_2}{S} < 1 \quad (9)$$

It clearly covers a smaller region in the $(q_1, q_2)$-plane compared to the condition (5) for a stable fixed-time program, see Fig. 18. The example demonstrates that a controller, which is optimal for a single intersection in isolation, may fail to be stable, if the intersection is linked with others in a network.

The same clearing policy can be stabilized, however, if combined with the supervisor proposed in Sec. 2. Fig. 19 shows the resulting solution, in which each traffic stream is being served sufficiently long at least once within the desired service period of $Z = 90s$.

The above example should learn us that even if the traffic light controllers at the intersections operate independently of each other, the switching in the
Figure 17 Temporal evolution of the unstable queues at the two intersections of Fig. 16. If the queues are cleared in an alternating way one after another, oscillations start to grow. The right intersection has most of the traffic load in the interval $[t_1, t_2]$, the left one in the interval $[t_3, t_4]$. At the end of a control cycle, the queues are longer than before. The chosen parameters are: $n_1(t_0) = 10$, $S = 1/1.8$ s, and $q_1 = q_2 = 0.6 S$.

Figure 19 Temporal evolution of the stabilized queues for the Kumar-Seidman-network depicted in Fig. 17. The proposed supervisory stabilization guarantees that all traffic flows are assigned green times often and long enough. Note that the stationary solution extends over a period that is larger than $Z$, during which each traffic flow is being served twice. However, the service interval itself (sum of red and subsequent green time) never exceeds $Z$. The parameters used are: $Z = 90$ s, $Z_{\text{max}} = 120$ s, $S = 1/1.8$ s, and $q_1 = q_2 = 0.6 S$. 
network is correlated, as the intersections are coupled by the network flows. The switching sequence of one intersection directly influences the arrival patterns of vehicles at other intersections, and as soon as its controllers dynamically respond to it, feedback loops are created. Any street segment with bidirectional traffic flows can act as a feedback loop. Clearing policies are stable only in acyclic networks (Kumar and Seidman (1990), Rybko and Stolyar (1992)). Road networks, however, are by definition never acyclic, because one can travel in the network from any node to any other and back. That means, whenever traffic lights are programmed never to interrupt the clearing of queues, there can be situations in which they become unstable.

References


