

# Operation regimes and slower-is-faster effect in the control of traffic intersections

D. Helbing<sup>a</sup> and A. Mazloumian

ETH Zurich, UNO D11, Universitätsstr. 41, 8092 Zurich, Switzerland

Received 3 December 2008 / Received in final form 6 March 2009

Published online 23 June 2009 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2009

**Abstract.** The efficiency of traffic flows in urban areas is known to crucially depend on signal operation. Here, elements of signal control are discussed, based on the minimization of overall travel times or vehicle queues. Interestingly, we find different operation regimes, some of which involve a “slower-is-faster effect”, where a delayed switching reduces the average travel times. These operation regimes characterize different ways of organizing traffic flows in urban road networks. Besides the optimize-one-phase approach, we discuss the procedure and advantages of optimizing multiple phases as well. To improve the service of vehicle platoons and support the self-organization of “green waves”, it is proposed to consider the price of stopping newly arriving vehicles.

**PACS.** 89.40.Bb Land transportation – 87.19.lr Control theory and feedback – 47.85.L- Flow control

## 1 Introduction

The study of urban traffic flows has attracted the interest of physicists for quite a while (see, e.g., Refs. [1–4]). This includes the issue of traffic light control and the resulting dynamics of vehicle flows [5–11]. Theoretical investigations in this direction have primarily focussed on single intersections and grid-like street networks, e.g. adaptive control [12–14] of a single traffic light or coordination of traffic lights in Manhattan-like road networks with unidirectional roads and periodic boundary conditions. Some of the fascination for traffic light control is due to the relationship with the synchronization of oscillators [15–17] and other concepts of self-organization [18–25].

The efficiency of traffic light control is essential to avoid or at least delay the collapse of traffic flows in traffic networks, particularly in urban areas. It is also crucial for attempts to reduce the fuel consumption and CO<sub>2</sub> emissions of vehicles. Both, delay times and acceleration maneuvers (i.e. the number of stops faced by vehicles)<sup>1</sup> cause additional fuel consumption and additional CO<sub>2</sub> emissions [26]. Within the USA alone, the cost of congestion per year is estimated to be 63.1 billion US\$, related with 3.7 billion hours of delays and 8.7 billion liters of “wasted” fuel [27]. Climate change and political goals to reduce CO<sub>2</sub> emissions force us to rethink the design and operation of traffic systems, which contributes about one third to the energy consumption of industrialized countries. On free-

ways, traffic flows may eventually be improved by automated, locally coordinated driving, based on new sensor technologies and intervehicle communication [29,30].

But what are options for urban areas? There, traffic lights are used to resolve conflicts of intersecting traffic streams. In this way, they avoid accidents and improve the throughput at moderate or high traffic volumes. For a discussion of the related traffic engineering literature, including the discussion of traffic light coordination and adaptive signal control, see references [23,25] and references therein. In the following, we will focus our attention on some surprising aspects of traffic flow optimization.

### 1.1 Paradoxical behavior of transport systems

Besides Braess’ paradox (which is related to selfish routing) [31–33], the slower-is-faster effect is another counter-intuitive effect that seems to occur in many transport networks. It has been found for pedestrian crowds, where a rush of people may delay evacuation [34].

Slower-is-faster effects have fascinated scientists for a long time. Smeed [35], for example, discussed “some circumstances in which vehicles will reach their destinations earlier by starting later”, but Ben-Akiva and de Palma [36] showed that this effect disappeared under realistic assumptions. Moreover, it is known from queuing theory that idle time can decrease the work in process (i.e. basically the queue length) in cyclically operated production systems under certain circumstances, particularly when the variance in the setup times is large [37]. These circumstances, however, do not seem to be very relevant for

<sup>a</sup> e-mail: [dhelbing@ethz.ch](mailto:dhelbing@ethz.ch)

<sup>1</sup> For formulas to estimate these quantities as a function of the utilization of the service capacity of roads see reference [28].

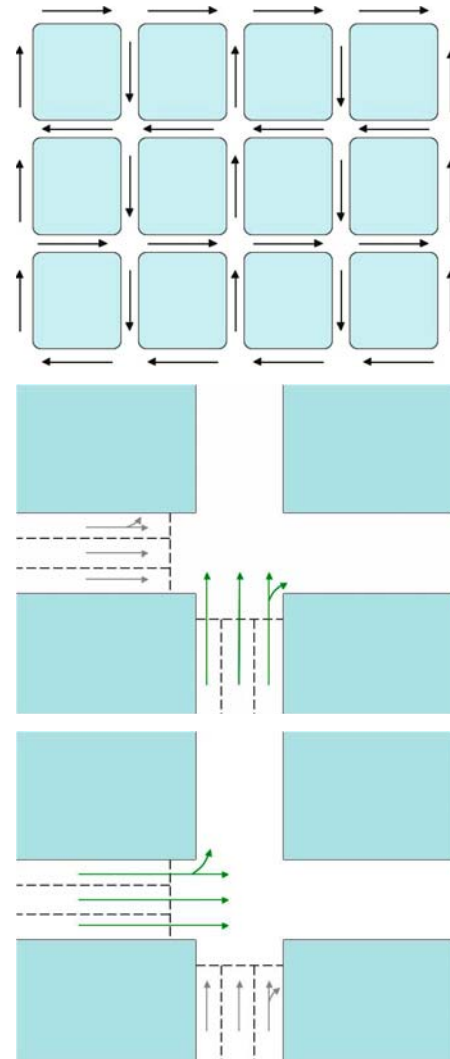
traffic light control. Nevertheless, there are many examples of slower-is-faster effects in traffic, production, and logistic systems, and it has been suggested that the phenomenon is widespread in networked systems with conflicting flows that are competing for prioritization [38,39]. While there are numerical algorithms to exploit this effect systematically to improve the performance of these systems [38], there have been only a few analytical studies of the slower-is-faster effect [40–42]. Therefore, we will put a particular focus on the study of conditions leading to this counterintuitive, but practically relevant effect.

Our paper is structured as follows: while Section 2 specifies the traffic system investigated in this paper, Section 3 discusses the throughput of intersections. Section 4 continues with the problem of minimizing travel times, while Section 4.5 discusses the minimization of queue lengths. The challenge in these sections is to come up with a concept that still leads to reasonably simple formulas, allowing one to study the behavior of the proposed signal control analytically. A successful approach in this respect is the “optimize-one-phase approach”, which seems justified by the short intervals, over which traffic flows can be anticipated reliably. Among the operation regimes resulting from the optimization process are also some with extended green times, corresponding to a “slower-is-faster effect” (see Sect. 4.4). A further improvement of signal operation is reached by applying multi-phase optimization, when flow constraints are taken into account. As Section 5 shows, this approach leads to a variety of plausible operation regimes. A summary and discussion is presented in Section 6. Complementary, Section 6.1 will discuss the “price” of stopping vehicles, which is an interesting concept to support moving vehicle platoons (and, thereby, the self-organization of “green waves”). For a more sophisticated, but analytically less accessible approach to the self-organization of coordinated traffic lights and vehicle streams in road networks see references [25,38,43].

## 2 Specification of the traffic system under consideration

In this paper, we will first focus on the study of a single traffic intersection with uniform arrival flows, before we discuss later how to extend our control concept in various ways. Furthermore, for simplicity we will concentrate on the study of a traffic light control with *two* green phases only, which is generalized in Appendix B. As the traffic organization in parts of Barcelona shows, a two-phase control is *sufficient*, in principle, to reach all points in the road network: just assume unidirectional flows in all streets with alternating directions. Then, in each phase, traffic either flows straight ahead and/or turns (right or left, depending on the driving direction in the crossing road). Hence, two intersecting unidirectional roads imply two possible traffic phases, which alternate (see Fig. 1).

While the optimization approach discussed in the following can be also applied to time-dependent arrival flows



**Fig. 1.** (Color online) Top: schematic illustration of the unidirectional street layout in the center of Barcelona. Center and bottom: illustration of the two traffic phases, during which vehicles can move straight ahead or turn (either right or left, depending on the direction of the crossing road).

$A_1$  and  $A_2$  per lane, when numerical solution methods are applied, for the sake of analytical tractability and closed formulas we will focus here on the case of constant flows over the short time periods involved in our optimization.  $I_j$  will represent the number of lanes of road section  $j$ , and it will be assumed that vehicles passing a green light can freely enter the respective downstream road section. Analogously to references [24,28,44], the departure flows  $\gamma_j(t)O_j(t)$  (as long as the traffic flows are not obstructed by the downstream traffic conditions) are given by the possible outflows  $O_j(t)$  (which vary with time  $t$ ), multiplied with the permeabilities  $\gamma_j(t)$ . The latter reflect the states of the traffic lights. During amber and red time periods, the permeabilities  $\gamma_j(t)$  are zero, as there is no outflow, while  $\gamma_j(t) = 1$  during green phases. Note that the departure flows  $\gamma_j(t)O_j(t)$  may split up into a straight and a turning flow after the traffic light, but for our further

considerations, this is not relevant. The possible outflows  $O_j(t)$  are determined by the equation

$$O_j(t) = \begin{cases} \widehat{Q}_j & \text{if } \Delta N_j(t) > 0, \\ A_j(t - \mathcal{T}_j^0) & \text{otherwise.} \end{cases} \quad (1)$$

Herein,  $\widehat{Q}_j$  is the service rate per lane during the green phase as long as there is a finite number  $\Delta N_j(t) > 0$  of delayed vehicles behind the traffic light (i.e.  $\widehat{Q}_j$  corresponds to the characteristic outflow from congested traffic).  $A_j(t)$  represents the time-dependent arrival rate of vehicles per lane and  $A_j(t - \mathcal{T}_j^0)$  the rate of vehicles arriving at the traffic light under free flow conditions, where  $\mathcal{T}_j^0$  denotes the free travel time needed to pass road section  $j$ . In case of constant arrival rates  $A_j$ , the dependence on the time point  $t$  and the delay by the free travel time  $\mathcal{T}_j^0$  can be dropped.

In the following, we will use some additional variables and parameters:  $T_j$  shall denote the minimum green time, after which the vehicle queue in road section  $j$  is fully dissolved (i.e. after which  $\Delta N_j = 0$  and  $O_j = A_j$ ). In contrast,  $\Delta T_j$  will stand for the actual green time period. Consequently,

$$\Delta t_j = \Delta T_j - T_j \quad (2)$$

(if greater than zero) represents the excess green time, during which we have a free vehicle flow with  $\gamma_j(t)O_j(t) = A_j$ .  $\tau_j$  shall be the setup time *before* the green phase  $\Delta T_j$  for road section  $j$ . For illustrative reasons, it is also called the ‘‘amber time (period)’’, although it is usually somewhat longer than that. The sum

$$T_{\text{cyc}} = \tau_1 + \Delta T_1 + \tau_2 + \Delta T_2 \quad (3)$$

is normally called the cycle time. Note, however, that we do not need to assume *periodic* operation. Within the framework of our model assumptions, we may consider *stepwise* constant flows. That is, the arrival flows may vary from one cycle (or even one green time period) to the next. Under such conditions, each green phase is adjusted to the changing traffic situation.

Finally note that we do not consider pedestrian flows in this paper. In order to take them into account, one would have to consider additional traffic phases for the service of pedestrians. Alternatively, one could select the setup times  $\tau_j$  for vehicles so large that they cover the amber time for vehicles plus a sufficient time for pedestrians to cross the road.

### 3 Consideration of traffic flows

The art of traffic control is to manipulate the permeabilities  $\gamma_j(t)$  in a way that optimizes a given goal function. In fact, when the traffic volume is high enough, an oscillatory service corresponding to the operation of a traffic light can increase the effective intersection capacity as compared to the application of a first-come-first-serve rule for arriving vehicles [23,24]: while the red and amber lights (corresponding to  $\gamma_j(t) = 0$ ) cause vehicles to queue up

and wait, this implies a high flow rate and an efficient service of vehicles when the traffic light turns green (i.e.  $\gamma_j(t) = 1$ ).

One natural concept of traffic flow optimization would be to maximize the average overall throughput. This is measured by the function

$$G_t(t) = \frac{1}{t} \sum_j \int_0^t dt' \gamma_j(t') O_j(t'). \quad (4)$$

Due to equation (1),  $G_t(t)$  depends not only on the outflows  $O_j(t)$ , but also on the inflows  $A_j(t)$  to the system. This makes  $G_t(t)$  basically dependent on the time-dependent origin-destination matrices of vehicle flows.

The numbers of vehicles accumulating during the red and amber time periods are

$$I_1 \Delta N_1^{\text{max}} = I_1 A_1 (\tau_2 + \Delta T_2 + \tau_1) \quad (5)$$

and

$$I_2 \Delta N_2^{\text{max}} = I_2 A_2 (\tau_1 + \Delta T_1 + \tau_2), \quad (6)$$

where  $\Delta N_j^{\text{max}}$  represents the maximum number of delayed vehicles per lane in road section  $j$ , if the vehicle queue in it has been fully cleared before.  $I_j$  is the number of lanes. As the service rate of queued vehicles during the green time  $\Delta T_j$  is  $\widehat{Q}_j$ , and  $A_j$  is the arrival rate of additional vehicles at the end of the queue, the minimum green time required to dissolve the queue is given by

$$T_j = \frac{\Delta N_j^{\text{max}}}{\widehat{Q}_j - A_j}. \quad (7)$$

From equations (5) to (7) we obtain

$$T_1 = \frac{A_1}{\widehat{Q}_1 - A_1} \left( \tau_2 + \Delta T_2 + \tau_1 \right). \quad (8)$$

Assuming  $\Delta T_j = T_j$  (i.e. no excess green times) and inserting equation (7) yields

$$T_1 = \frac{A_1}{\widehat{Q}_1 - A_1} \left( \tau_2 + \frac{A_2 (\tau_1 + T_1 + \tau_2)}{\widehat{Q}_2 - A_2} + \tau_1 \right) \quad (9)$$

for the clearing time  $T_1$ , or

$$T_1 = (\tau_1 + \tau_2) \frac{\frac{A_1}{\widehat{Q}_1 - A_1} \left( 1 + \frac{A_2}{\widehat{Q}_2 - A_2} \right)}{1 - \frac{A_1 A_2}{(\widehat{Q}_1 - A_1)(\widehat{Q}_2 - A_2)}}. \quad (10)$$

With the analogous formula for  $T_2$  we can determine the related cycle time, if the traffic light turns red immediately when all queued vehicles have been served. After a few intermediate mathematical steps, we finally get

$$T^{\text{cyc}} = \tau_1 + T_1 + \tau_2 + T_2 = \frac{\tau_1 + \tau_2}{1 - A_1/\widehat{Q}_1 - A_2/\widehat{Q}_2}. \quad (11)$$

Moreover, one can show [28]

$$T_j = \frac{A_j}{\widehat{Q}_j} T^{\text{cyc}}. \quad (12)$$

We can see that the cycle time and the clearing times  $T_j$  diverge in the limit

$$\frac{A_1}{\widehat{Q}_1} + \frac{A_2}{\widehat{Q}_2} \rightarrow 1. \quad (13)$$

If this expression (11) becomes negative, the vehicle queues in one or both ingoing road sections are growing larger and larger in time, as the intersection does not have enough capacity to serve both arrival flows. See reference [28] for a discussion of this case.

Note that equation (11) determines the *smallest* cycle time that allows to serve all queued vehicles within the green time periods. Let us study now the effect of *extending* the green time periods  $\Delta T_j$  beyond  $T_j$ : the average throughput of the intersection is given by the overall flow of vehicles during one cycle time  $T_{\text{cyc}} = \tau_1 + \Delta T_1 + \tau_2 + \Delta T_2$ . During that time period, a total number  $(I_1 A_1 + I_2 A_2) T_{\text{cyc}}$  of vehicles is arriving in the two considered road sections. If all arriving vehicles are served during the cycle time  $T_{\text{cyc}}$ , the average throughput is

$$G_t = \frac{(I_1 A_1 + I_2 A_2) T_{\text{cyc}}}{T_{\text{cyc}}} = I_1 A_1 + I_2 A_2. \quad (14)$$

Therefore, in the case where we do not have an accumulation of vehicles over time, which requires sufficient green times ( $\Delta T_j > T_j$ ) and a sufficient resulting service capacity

$$\frac{I_1 \widehat{Q}_1 \Delta T_1 + I_2 \widehat{Q}_2 \Delta T_2}{T_{\text{cyc}}} \geq I_1 A_1 + I_2 A_2, \quad (15)$$

the throughput is determined by the sum  $I_1 A_1 + I_2 A_2$  of the overall arrival flows. Consequently, excess green times  $\Delta t_j = \Delta T_j - T_j > 0$  do not lead to smaller or larger intersection throughputs. But under what conditions should a green phase be extended, if at all? This shall be addressed in the next sections.

## 4 Travel-time-oriented signal operation

Rather than on a consideration of the flow, we will now focus on the *cumulative waiting time*

$$F(t) = \sum_j I_j \int_0^t dt' \int_0^{t'} dt'' [A_j - \gamma_j(t'') O_j(t'')] \quad (16)$$

and minimize its average growth over a time period  $t$  to be defined later. This corresponds to a minimization of the function

$$\begin{aligned} G(t) &= \frac{1}{t} \sum_j I_j \int_0^t dt' \Delta N_j(t') \\ &= \frac{1}{t} \sum_j I_j \int_0^t dt' \int_0^{t'} dt'' [A_j - \gamma_j(t'') O_j(t'')], \end{aligned} \quad (17)$$

which quantifies the time average of the overall delay time. The term on the right-hand side describes the increase of the overall waiting time proportionally to the number  $\Delta N_j$  of delayed cars, which is given by the integral over the difference between the arrival and departure flows [24,28].

Note that the formula (17) makes an implicit simplification by assuming that delays occur only in the vehicle queues behind traffic lights, while no delays accumulate under uncongested flow conditions. This assumes a triangular flow-density diagram, which, however, seems to be sufficiently justified for urban traffic flows [24,44]. Moreover, while approaching a vehicle queue, it usually does not matter, when vehicles travel more slowly than the speed limit allows: If they would travel faster, they would be queued earlier, i.e. the delay would stay the same. In other words, most of the time it is irrelevant, whether vehicles lose their time in the vehicle queue or by decelerating before.

### 4.1 The optimize-one-phase approach

When minimizing the goal function  $G(t)$ , it is essential upto what time  $t$  we extend the integral. In principle, it is possible to integrate over a full cycle or even many cycles of traffic operation, but the resulting formulas do not provide an intuitive understanding anymore. We will, therefore, focus on the optimization of a single phase, with full amber time periods  $\tau_j$  in the beginning and  $\tau_{j+1}$  at the end. This turns out to result in explicit and plausible formulas, while some other approaches we have tried, did not result in well interpretable results. Besides this practical aspect, when analytical results shall be obtained, the specification  $t = \tau_1 + \Delta T_j + \tau_2$  chosen in the following makes sense: it “charges” the switching-related inefficiencies to the road that “wants” to be served. The switching of a traffic light should lead to a temporary *increase* in traffic performance. After completion of each green phase, the travel time optimization is repeated, so that one can compose the traffic light schedule as a sequence of optimized single phases (see Appendix B for details).

In Section 5, we will show that a multi-phase optimization yields better results, but requires a higher degree of sophistication. The treatment of situations with varying or pulsed traffic flows is even more difficult and can usually be solved only numerically. This issue is addressed in reference [25].

In our calculations, we will assume that the green time for road Section 2 lasted for a time period  $\Delta T_2$  and ended at time  $t = 0$ . That is, we have now to determine the optimal duration  $\Delta T_1$  of the green phase for road Section 1 after an intermediate amber time period  $\tau_1$ . For this, we minimize the function

$$G_1(\tau_1 + \Delta T_1 + \tau_2) = \frac{F_1(\tau_1 + \Delta T_1 + \tau_2)}{\tau_1 + \Delta T_1 + \tau_2}, \quad (18)$$

where the subscript “1” of  $G$  and  $F$  refers to road Section 1, for which the green phase is determined. Assuming a step-wise constant outflow with  $\gamma_j O_j = \widehat{Q}_j$ , if  $\Delta N_j > 0$ ,

but  $\gamma_j O_j = A_j$ , if  $\Delta N_j = 0$ , and  $\gamma_j O_j = 0$ , if  $\gamma_j = 0$ , the integral over  $t''$  results in a stepwise linear function, and the function  $F_1(t)$  is characterized by quadratic dependencies. We will distinguish two cases: (a) the green time is potentially terminated *before* all queued vehicles have been served (i.e.  $\Delta T_i \leq T_i$ ), or (b) it is potentially *extended* (i.e.  $\Delta T_i \geq T_i$ ). Let us start with the first case.

(a) No excess green time ( $\Delta T_1 \leq T_1$ ): in this case,  $A_2(t'') - \gamma_2(t'')O_2(t'') = A_2$  for  $0 \leq t'' \leq \tau_1 + \Delta T_1 + \tau_2$ , i.e. over the period  $\Delta T_1$  of the green time for road Section 1 and the amber time periods  $\tau_j$  and  $\tau_{j+1}$  before and after it. In addition,

$$A_1 - \gamma_1(t'')O_1(t'') = \begin{cases} A_1 & \text{if } 0 \leq t'' < \tau_1, \\ A_1 - \widehat{Q}_1 & \text{if } \tau_1 \leq t'' < \tau_1 + \Delta T_1, \\ A_1 & \text{otherwise.} \end{cases} \quad (19)$$

Using the abbreviation

$$\Delta N_1^{\max} = \Delta N_1(\tau_1) = \Delta N_1(0) + A_1 \tau_1, \quad (20)$$

we get

$$\begin{aligned} F_1^a(\tau_1 + \Delta T_1 + \tau_2) &= I_1 \left\{ \Delta N_1(0) \tau_1 + A_1 \frac{\tau_1^2}{2} + \Delta N_1^{\max} \Delta T_1 \right. \\ &\quad \left. - (\widehat{Q}_1 - A_1) \frac{\Delta T_1^2}{2} + [\Delta N_1^{\max} - (\widehat{Q}_1 - A_1) \Delta T_1] \tau_2 + A_1 \frac{\tau_2^2}{2} \right\} \\ &\quad + I_2 \left[ \Delta N_2(0) (\tau_1 + \Delta T_1 + \tau_2) + A_2 \frac{(\tau_1 + \Delta T_1 + \tau_2)^2}{2} \right] \\ &= I_1 \left[ \Delta N_1(0) (\tau_1 + \Delta T_1 + \tau_2) + \frac{A_1}{2} (\tau_1 + \Delta T_1 + \tau_2)^2 \right. \\ &\quad \left. - \frac{\widehat{Q}_1}{2} \Delta T_1 (\Delta T_1 + 2\tau_2) \right] + I_2 [\Delta N_2(0) (\tau_1 + \Delta T_1 + \tau_2) \\ &\quad + A_2 \frac{(\tau_1 + \Delta T_1 + \tau_2)^2}{2}], \quad (21) \end{aligned}$$

where the superscript ‘‘a’’ refers to case (a). Dividing the above function by  $(\tau_1 + \Delta T_1 + \tau_2)$  and making the plausible assumption  $\tau_1 = \tau_2$  of equal amber time periods for simplicity, we gain

$$\begin{aligned} G_1^a(\tau_1 + \Delta T_1 + \tau_2) &= I_1 \left[ \Delta N_1(0) + \widehat{Q}_1 \tau_2 \right. \\ &\quad \left. - (\widehat{Q}_1 - A_1) \frac{\tau_1 + \Delta T_1 + \tau_2}{2} \right] \\ &\quad + I_2 \left[ \Delta N_2(0) + A_2 \frac{\tau_1 + \Delta T_1 + \tau_2}{2} \right]. \quad (22) \end{aligned}$$

If  $I_1(\widehat{Q}_1 - A_1) < I_2 A_2$ , i.e. when the number of queued vehicles in road Section 2 grows faster than it can be reduced in road Section 1, the minimum of this function is reached for  $\Delta T_1 = 0$ , corresponding to a situation where it is not favorable to turn green for section  $j = 1$ . For

$$I_1(\widehat{Q}_1 - A_1) > I_2 A_2, \quad (23)$$

the value of  $G_1^a$  goes down with growing values of  $\Delta T_1$ , and the minimum is reached for a value  $\Delta T_1 \geq T_1$ .

(b) Potential green time extension ( $\Delta T_1 \geq T_1$ ): let us assume that we (possibly) have an excess green time, i.e.  $\Delta t_i = \Delta T_i - T_i \geq 0$ . In this case,

$$A_1 - \gamma_1(t'')O_1(t'') = \begin{cases} A_1 & \text{if } 0 \leq t'' < \tau_1, \\ A_1 - \widehat{Q}_1 & \text{if } \tau_1 \leq t'' < \tau_1 + T_1, \\ A_1 & \text{if } t'' \geq \tau_1 + \Delta T_1, \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

Considering that now,  $\Delta N_1(t') = 0$  for  $\tau_1 + T_1 \leq t' < \tau_1 + \Delta T_1$ , and introducing the *clearing time*

$$T_1 = \frac{\Delta N_1^{\max}}{\widehat{Q}_1 - A_1} = \frac{\Delta N_1(0) + A_1 \tau_1}{\widehat{Q}_1 - A_1}, \quad (25)$$

we obtain

$$\begin{aligned} F_1^b(\tau_1 + \Delta T_1 + \tau_2) &= I_1 \left[ \Delta N_1(0) \tau_1 + A_1 \frac{\tau_1^2}{2} \right. \\ &\quad \left. + \Delta N_1^{\max} T_1 - (\widehat{Q}_1 - A_1) \frac{T_1^2}{2} + A_1 \frac{\tau_2^2}{2} \right] \\ &\quad + I_2 \left[ \Delta N_2(0) (\tau_1 + \Delta T_1 + \tau_2) + A_2 \frac{(\tau_1 + \Delta T_1 + \tau_2)^2}{2} \right] \\ &= I_1 \left[ \Delta N_1^{\max} \tau_1 + \frac{A_1}{2} (\tau_2^2 - \tau_1^2) + \frac{(\Delta N_1^{\max})^2}{2(\widehat{Q}_1 - A_1)} \right] \\ &\quad + I_2 \left[ \Delta N_2(0) (\tau_1 + \Delta T_1 + \tau_2) + A_2 \frac{(\tau_1 + \Delta T_1 + \tau_2)^2}{2} \right]. \quad (26) \end{aligned}$$

Assuming again  $\tau_1 = \tau_2$  for simplicity, introducing the abbreviation

$$E_1 = \Delta N_1^{\max} \tau_1 + \frac{(\Delta N_1^{\max})^2}{2(\widehat{Q}_1 - A_1)}, \quad (27)$$

and dividing equation (26) by  $(\tau_1 + \Delta T_1 + \tau_2)$  yields

$$\begin{aligned} G_1^b(\tau_1 + \Delta T_1 + \tau_2) &= \frac{I_1 E_1}{\tau_1 + \Delta T_1 + \tau_2} \\ &\quad + I_2 \left[ \Delta N_2(0) + A_2 \frac{\tau_1 + \Delta T_1 + \tau_2}{2} \right]. \quad (28) \end{aligned}$$

This expression shall be minimized under the constraint  $\Delta T_1 \geq T_1$ . In order to determine the minimum, we set the derivative with respect to  $\Delta T_1$  to zero and get

$$0 = \frac{dG_1^b(\tau_1 + \Delta T_1 + \tau_2)}{d\Delta T_1} = -\frac{I_1 E_1}{(\tau_1 + \Delta T_1 + \tau_2)^2} + \frac{I_2 A_2}{2}. \quad (29)$$

The minimum is located at

$$(\tau_1 + \Delta T_1 + \tau_2)^2 = \frac{2I_1 E_1}{I_2 A_2}, \quad (30)$$



if  $\Delta T_1 \geq T_1$ . Considering equation (25),  $\Delta T_1 \geq T_1$  implies

$$(\tau_1 + \Delta T_1 + \tau_2)^2 \geq \left( \tau_1 + \frac{\Delta N_1^{\max}}{\widehat{Q}_1 - A_1} + \tau_2 \right)^2. \quad (31)$$

With equation (30) this leads to the condition

$$\frac{(\Delta N_1^{\max})^2}{\widehat{Q}_1 - A_1} \left( \frac{I_1}{I_2 A_2} - \frac{1}{\widehat{Q}_1 - A_1} \right) + 2\Delta N_1^{\max} \left( \frac{I_1 \tau_1}{I_2 A_2} - \frac{\tau_1 + \tau_2}{\widehat{Q}_1 - A_1} \right) \geq (\tau_1 + \tau_2)^2. \quad (32)$$

If inequality (32) is not fulfilled, we must have  $\Delta T_1 < T_1$ .

For completeness, we note that

$$G_1^a(\tau_1 + T_1 + \tau_2) = G_1^b(\tau_1 + T_1 + \tau_2), \quad (33)$$

i.e. the goal function  $G_1$  is continuous in  $\Delta T_1 = T_1$ , while it must not be smooth. Moreover,  $\Delta N_1(0) = A_1(\tau_2 + \Delta T_2)$  and  $\Delta N_2(0) = 0$ , if the vehicle queues have been fully cleared before the traffic light is switched. The case where the queue is not fully dissolved is treated in reference [28].

## 4.2 Transformation to dimensionless variables and parameters

For an analysis of the system behavior, it is useful to transform variables and parameters to dimensionless units. Such dimensionless units are, for example, the capacity utilizations

$$u_i = \frac{A_i}{\widehat{Q}_i} \quad (34)$$

of the road sections  $i$  and the relative size

$$\kappa = \frac{I_1 A_1}{I_2 A_2} = \frac{I_1 u_1 \widehat{Q}_1}{I_2 u_2 \widehat{Q}_2} = \frac{u_1}{u_2} K \quad (35)$$

of the arrival flows, where

$$K = \frac{I_1 \widehat{Q}_1}{I_2 \widehat{Q}_2}. \quad (36)$$

Furthermore, we may scale the green times  $\Delta T_i$  by the sum of amber time periods  $\tau_1 + \tau_2$ , which defines the dimensionless green times

$$\sigma_i = \frac{\Delta T_i}{\tau_1 + \tau_2} \quad (37)$$

and the dimensionless clearing times

$$\hat{\sigma}_j = \frac{T_j}{\tau_1 + \tau_2} = \frac{\Delta N_1^{\max}}{(1 - u_1) \widehat{Q}_i (\tau_1 + \tau_2)}. \quad (38)$$

In order to express the previous relationships exclusively by these quantities, we must consider that a number

$A_1(\tau_2 + \Delta T_2 + \tau_1)$  of vehicles per lane accumulates during the time period  $(\tau_2 + \Delta T_2 + \tau_1)$ , in which the vehicle flow on road Section 1 is not served. With equation (20) this implies

$$\Delta N_1^{\max} = \Delta N_1(0) + A_1 \tau_1 = A_1(\tau_2 + \Delta T_2 + \tau_1), \quad (39)$$

if the vehicle queue in road Section 1 has been fully cleared during the previous green time. Then, we have

$$\frac{\Delta N_1^{\max}}{\tau_1 + \tau_2} = A_1(1 + \sigma_2), \quad (40)$$

and from equations (27) and (22) we get

$$\frac{2E_1}{(\tau_1 + \tau_2)^2} = A_1(1 + \sigma_2) \frac{2\tau_1}{\tau_1 + \tau_2} + \frac{(A_1)^2(1 + \sigma_2)^2}{\widehat{Q}_1 - A_1}. \quad (41)$$

With  $A_1 = u_1 \widehat{Q}_1$  and  $\tau_1 = \tau_2$ , equation (30) belonging to the case of extended green time for road Section 1 can be written as

$$(1 + \sigma_1)^2 = [1 + \tilde{\sigma}_1(\sigma_2)]^2 = \kappa \left[ (1 + \sigma_2) + \frac{u_1}{1 - u_1} (1 + \sigma_2)^2 \right]. \quad (42)$$

The solution of this equation defines the relationship  $\tilde{\sigma}_1(\sigma_2)$  for the optimal scaled green time period  $\sigma_1$  as a function of  $\sigma_2$ , if the green time for road Section 1 is extended. Moreover, in dimensionless variables, the condition (32) for green time extension becomes

$$\frac{u_1}{1 - u_1} (1 + \sigma_2)^2 \left( \kappa - \frac{u_1}{1 - u_1} \right) + (1 + \sigma_2) \left( \kappa - \frac{2u_1}{1 - u_1} \right) \geq 1 \quad (43)$$

or

$$\left[ \frac{u_1(1 + \sigma_2)^2}{1 - u_1} + (1 + \sigma_2) \right] \left( \kappa - \frac{u_1}{1 - u_1} \right) \geq \frac{1 + u_1 \sigma_2}{1 - u_1}. \quad (44)$$

However, we can check for green time extension also in a different way, since the extension condition  $\Delta T_1 > T_1$  can be written as  $\hat{\sigma}_1 > \hat{\sigma}_1$ . Using equations (25), (38) and (39), the dimensionless green time  $\sigma_1$  for the case of no green time extension may be presented as

$$\sigma_1 = \hat{\sigma}_1(\sigma_2) = \frac{A_1(1 + \sigma_2)}{\widehat{Q}_1 - A_1} = \frac{u_1(1 + \sigma_2)}{1 - u_1} \quad (45)$$

or

$$1 + \sigma_1 = \frac{1 + u_1 \sigma_2}{1 - u_1}. \quad (46)$$

Moreover, from  $\sigma_1 = \hat{\sigma}_1(\sigma_2)$  follows

$$\frac{\sigma_1}{1 + \sigma_1 + \sigma_2} = \frac{u_1(1 + \sigma_2)}{(1 - u_1) \left[ 1 + \frac{u_1}{1 - u_1} (1 + \sigma_2) + \sigma_2 \right]} = u_1. \quad (47)$$

That is, in the case where road Section 1 is completely cleared, but there is no green time extension, the green time fraction

$$\frac{\Delta T_1}{T_{\text{cyc}}} = \frac{\sigma_1}{1 + \sigma_1 + \sigma_2} \quad (48)$$

agrees with the utilization  $u_1$ . Moreover, one can show

$$\frac{\partial}{\partial \sigma_1} \left( \frac{\sigma_1}{1 + \sigma_1 + \sigma_2} \right) = \frac{1 + \sigma_2}{(1 + \sigma_1 + \sigma_2)^2} > 0. \quad (49)$$

Therefore,  $\sigma_1 > \hat{\sigma}_1$  implies a green time fraction greater than  $u_1$ , and we have excess green time for road Section 1, if

$$\frac{\tilde{\sigma}_1(\sigma_2)}{1 + \tilde{\sigma}_1(\sigma_2) + \sigma_2} > u_1. \quad (50)$$

An analogous condition must be fulfilled, if excess green times on road Section 2 shall be optimal. It reads

$$\frac{\tilde{\sigma}_2(\sigma_1)}{1 + \sigma_1 + \tilde{\sigma}_2(\sigma_1)} > u_2, \quad (51)$$

where

$$[1 + \tilde{\sigma}_2(\sigma_1)]^2 = \frac{1}{\kappa} \left[ (1 + \sigma_1) + \frac{u_2}{1 - u_2} (1 + \sigma_1)^2 \right], \quad (52)$$

which has been gained by interchanging indices 1 and 2 and replacing  $\kappa$  by  $1/\kappa$  in equation (42).

### 4.3 Control strategies and slower-is-faster effect

Based on the results of Section 4.1 and the scaled formulas of Section 4.2, we can now formulate control strategies for a single traffic light within the optimize-one-phase approach:

- (i) Terminate the green light for road Section 1 immediately, corresponding to  $\sigma_1 = 0$ , if condition (23) is violated, i.e. if

$$1 - u_1 \leq \frac{u_2}{K} \quad (53)$$

is fulfilled. To obtain the dimensionless form of this inequality, we have considered  $A_j = u_j I_j \hat{Q}_j$  and equation (36). In case (i), travel time optimization for one phase advises against turning green for road Section 1. Of course, in reality, drivers cannot be stopped forever. Either, one would have to give them a short green phase after a maximum tolerable time period, or at least one would have to allow vehicles to turn on red, i.e. to merge the crossing flow, whenever there is a large enough gap between two successive vehicles. Alternatively, one may apply an optimize-multiple-phases approach, see Section 5. It implies a service of side roads even when the intersection capacity is insufficient to satisfy all inflows completely.

- (ii) Terminate the green phase for road Section 1, when the vehicle queue is completely resolved, if conditions (53) and (44) are violated. In this case, the scaled green time  $\sigma_1$  is given by equation (45).
- (iii) Extend the green times for road Section 1 in accordance with formula (42), if the condition (44) is fulfilled. The recommended delay in the switching time constitutes a *slower-is-faster effect*. In this situation, it takes some additional time to accumulate enough vehicles on road Section 2 to guarantee an efficient service in view of the inefficiencies caused by the setup times  $\tau_j$ .

In Figure 2, operation regime (i) is indicated in white and operation regime (iii) in red, while operation regime (ii) is shown in green, if road Section 2 is served, otherwise in orange.

### 4.4 Operation regimes for periodic operation

In the previous section, we have determined the optimal green time period  $\sigma_1$  for road Section 1, assuming that the last green time period  $\sigma_2$  for road Section 2 and  $N_1(0)$  were given. Of course,  $\sigma_1$  will then determine  $\sigma_2$ , etc. If the utilizations  $u_j$  are constant and not too high, the sequence of green phases converges towards a *periodic* signal operation (see Fig. 3). It will be studied in the following. While the formulas for the determination of  $\sigma_1$  were derived in Section 4.2, the corresponding formulas for  $\sigma_2$  can be obtained by interchanging the indices 1 and 2 and replacing  $\kappa$  by  $1/\kappa$  in all formulas. In principle, there could be the following cases, if we restrict ourselves to reasonable solutions with  $\sigma_j \geq 0$ :

- (0) According to travel time minimization, one or both road sections should not be served, if (53) is fulfilled for one or both of the road sections. This case occurs if

$$1 - u_1 - \frac{u_2}{K} \leq 0 \quad \text{or} \quad 1 - K u_1 - u_2 \leq 0 \quad (54)$$

(see the area above the white solid line in the right illustration of Fig. 2). According to this, service should focus on the main flow, while crossing flows should be suppressed, thereby enforcing a re-routing of traffic streams when this would be favorable to minimize travel times. Of course, in such situations vehicles should still be allowed to turn on red and to merge the crossing flow, when vehicle gaps are large enough.

- (1) Both green time periods are terminated as soon as the respective vehicle queues are fully dissolved. In this case, we should have the relationships  $\sigma_1 = \hat{\sigma}_1(\sigma_2)$  and  $\sigma_2 = \hat{\sigma}_2(\sigma_1)$ , where  $\hat{\sigma}_j$  is defined in equation (45). After a few steps, the condition  $\sigma_1 = \hat{\sigma}_1(\hat{\sigma}_2(\sigma_1))$  implies

$$\sigma_j = \hat{\sigma}_j = \frac{u_j}{1 - u_1 - u_2} \quad (55)$$

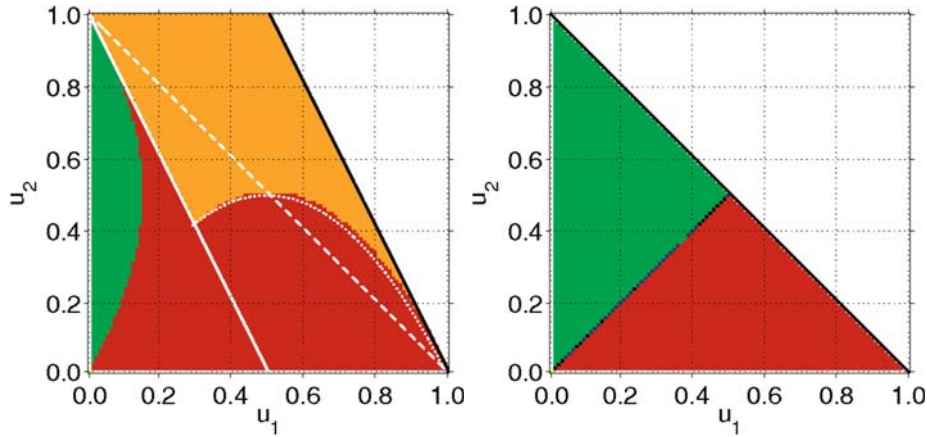
and

$$\frac{\sigma_j}{1 + \sigma_1 + \sigma_2} = u_j. \quad (56)$$

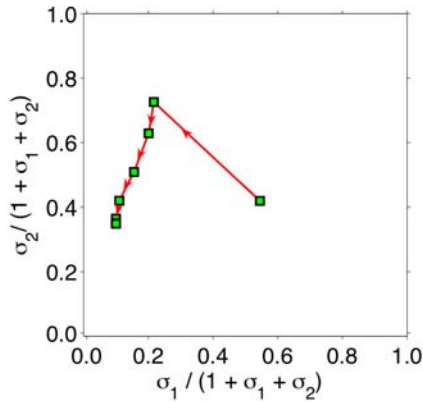
According to equation (56), the green time fraction of each road section in case (1) should be proportional to the respective utilization  $u_j$  of the flow capacity.

- (2) Road Section 2 gets an excess green time, while the green phase of road Section 1 ends after the dissolution of the vehicle queue (see green area in Fig. 2). In this case we should have  $\sigma_1 = \hat{\sigma}_1(\tilde{\sigma}_2(\sigma_1))$ , where  $(1 + \tilde{\sigma}_2)$  is defined by formula (52). This gives

$$\sigma_1 = \frac{u_1}{1 - u_1} \sqrt{\frac{1}{\kappa} \left( (1 + \sigma_1) + \frac{u_2}{1 - u_2} (1 + \sigma_1)^2 \right)}, \quad (57)$$



**Fig. 2.** (Color online) Operation regimes of (periodic) signal control for  $K = 1$  (left) and  $K = 3$  (right) as a function of the utilizations  $u_j$  of both roads  $j$  according to the one-phase travel time optimization approach. For each combination of  $u_1$  and  $u_2$ , the operation regime has been determined after convergence of the signal control procedure described in Appendix B. The separating lines are in good agreement with our analytical calculations. For example, the solid falling lines are given by equation (54), while the dotted parabolic line in the right illustration corresponds to  $u_2 = Ku_1(1 - u_1)$  and results by equalizing equation (42) with the square of equation (46), assuming  $\sigma_2 = 0$  (i.e. no service of road Sect. 2). The different operation regimes are characterized as follows: in the green triangular or parabolic area to the left of both illustrations, where the utilization  $u_1$  of road Section 1 is sufficiently small, the service of road Section 2 is extended. In the adjacent red area below the white area (left) or the solid line (right), road Section 2 is just cleared, while above the separating line  $u_2 = 1 - Ku_1$ , road Section 2 is not served at all. Road Section 1, in contrast, gets just enough green time to clear the vehicle queue in the green area (and the orange area towards the top of the right illustration), while it gets extended green time in the red area towards the bottom, where the utilization  $u_2$  of road Section 2 is sufficiently small. In the white area given by  $u_2 > K(1 - u_1)$ , road Section 1 gets no green time anymore. Between the dashed and the solid white lines, road Section 2 is not served, although there would be enough capacity to satisfy the vehicle flows in both roads. Improved operation regimes are presented in Figure 4.



**Fig. 3.** (Color online) Green time fraction  $\sigma_2/(1 + \sigma_1 + \sigma_2)$  for road Section 2 vs. green time fraction  $\sigma_1/(1 + \sigma_1 + \sigma_2)$  for road Section 1, if we apply the signal control algorithm described in Appendix B to a randomly chosen initial queue length  $\Delta N_1(0)$  in road Section 1 and  $K = 2$  (i.e. road Sect. 1 has 2 times as many lanes as road Sect. 2). One can clearly see that the green time fractions quickly converge towards values that do not change anymore over time. The solution corresponds to periodic signal operation.

which eventually leads to a quadratic equation for  $\sigma_1$ , namely

$$[u_1 u_2^2 - K(1 - u_1)^2(1 - u_2)]\sigma_1^2 + u_1 u_2(1 + u_2)\sigma_1 + u_1 u_2 = 0. \quad (58)$$

To determine  $\sigma_2$ , we can either use the relationship  $\sigma_2 = \tilde{\sigma}_2(\sigma_1)$  or invert the formula  $\sigma_1 = \hat{\sigma}_1(\sigma_2)$ . Doing the latter, equation (45) gives

$$\sigma_2 = \frac{1 - u_1}{u_1}\sigma_1 - 1. \quad (59)$$

According to equations (47) and (51), the occurrence of case (2) requires that the resulting solution satisfies

$$\frac{\sigma_1}{1 + \sigma_1 + \sigma_2} = u_1 \quad \text{and} \quad \frac{\sigma_2}{1 + \sigma_1 + \sigma_2} > u_2. \quad (60)$$

- (3) Road Section 1 gets an excess green time, while the green phase of road Section 2 ends after the dissolution of the vehicle queue (see red area in Fig. 2). The formulas for this case are obtained from the ones of case (2) by interchanging the indices 1 and 2 and replacing  $\kappa$  by  $1/\kappa$ .
- (4) Both road sections get excess green time periods. This case would correspond to  $\sigma_1 = \tilde{\sigma}_1(\tilde{\sigma}_2(\sigma_1))$ , and the solutions should fulfil

$$\frac{\sigma_1}{1 + \sigma_1 + \sigma_2} > u_1 \quad \text{and} \quad \frac{\sigma_2}{1 + \sigma_1 + \sigma_2} > u_2. \quad (61)$$

According to numerical results (see Fig. 2), cases (0), (2), and (3) do all exist, while the conditions for cases (1) and (4) are not fulfilled. Note, however, that small vehicle flows should better be treated as discrete or pulsed rather than continuous flows, in order to reflect the arrival of



single vehicles (see Ref. [45] for their possible treatment within a continuous flow framework). In other words, for rare vehicle arrivals, we either have  $u_1 > 0$  and  $u_2 = 0$ , or we have  $u_2 > 0$  and  $u_1 = 0$ . Hence, the case of small utilizations  $u_j$  will effectively imply green time extensions for both road sections due to the discreteness of the flow, and it allows single vehicles to pass the traffic light without previously stopping at the red light.

Summarizing the above, one-phase optimization provides extra green times for road sections, as long as both of them are fully served. While in one road section, this slower-is-faster effect allows some vehicles to pass the traffic light without stopping, in the other road section it causes the formation of a longer vehicle queue, which supports an efficient service of a substantial number of vehicles after the traffic light turns green. In this connection, it is useful to remember that switching is costly due to the amber times, which are “lost” service times.

#### 4.5 Minimization of vehicle queues

We have seen that travel time minimization implies the possibility of case (0), where one of the road sections (the side road) is not being served. This case should not occur as long as the intersection capacity is not fully used. According to equations (13) and (34), the intersection capacity is sufficient, if

$$u_1 + u_2 \leq 1 \quad (62)$$

As the inequalities (54) and (62) do not agree, conditions may occur, where the vehicle queue in one road section (a side road) continuously increases, even though the intersection capacity would allow to serve both flows (see the orange and red areas above the dashed white line in Fig. 2). This can result in an “unstable” signal control scheme, which causes undesired spillover effects and calls for a suitable stabilization strategy [25]. As we will see in the following, this problem can be overcome by minimizing vehicle queues rather than travel times.

Conditions (54) and (62) agree, if  $K = 1$ , particularly when  $I_1 = I_2$  and  $\hat{Q}_1 = \hat{Q}_2$ . Therefore, let us assume this case in the following, corresponding to

$$\kappa = \frac{I_1 A_1}{I_2 A_2} = \frac{I_1 u_1 \hat{Q}_1}{I_2 u_2 \hat{Q}_2} = \frac{u_1}{u_2}. \quad (63)$$

$\hat{Q}_1 = \hat{Q}_2$  holds, when the street sections downstream of the intersection do not impose a bottleneck. Furthermore,  $I_1 = I_2 = 1$  corresponds to a minimization of the average *queue length* rather than the average delay time. Such a minimization of the queue length makes a lot of sense and means that the optimization is made from the perspective of the traffic network rather than from the perspective of the driver. This minimizes spillover effects and, at the same time, keeps travel times low.

#### 4.6 Complexity of traffic light control

It is interesting that already a single intersection with constant arrival flows shows a large variety of operation regimes. In order to get an idea of the complexity of optimal traffic light control in general, let us ask about the dimension of the phase space. For such an analysis, it is common to transform all parameters to dimensionless form, as above. In this way, all formulas are expressed in terms relative flows such as

$$\kappa = \frac{I_1 A_1}{I_2 A_2}, \quad u_1 = \frac{A_1}{Q_1}, \quad u_2 = \frac{A_2}{Q_2}. \quad (64)$$

Parameters like

$$\frac{I_2(\hat{Q}_2 - A_2)}{I_1 A_1} \quad \text{and} \quad \frac{I_1(\hat{Q}_1 - A_1)}{I_2 A_2} \quad (65)$$

can be expressed through the previous set of parameters. A single intersection with 2 phases only is characterized by the 2 parameters  $u_1$  and  $u_2$ , if queue minimization is performed, and one additional parameter  $\kappa$ , if travel time is minimized. Therefore, the optimal operation of  $n$  intersections depends on  $2^n$  (or even  $3^n$ ) parameters. In view of this, it is obvious that the optimal coordination of traffic lights in an urban road network constitutes a hard computational problem [46].

The consideration of non-uniform arrival flows further complicates matters. If the traffic flows are not constant, but characterized by vehicle platoons, the *phase* of traffic light control can be significant for intersection capacity [28]. Therefore, the mutual coordination of neighboring traffic lights has a significant impact [28]. This issue is, for example, addressed in references [23,25].

### 5 Optimize-multiple-phases approach

Under certain circumstances, it may be reasonable to *interrupt* the service of a vehicle queue to clear the way for a large flow of newly arriving vehicles in the other road section. Such an interruption may be interpreted as another slower-is-faster effect, occurring in situations where the interruption-induced delay of vehicles in one road section is overcompensated for by the avoidance of delay times in the other road section. Such effects involving several green phases can clearly not be studied within the optimization of a single phase. One would rather need an approach that optimizes two or more phases simultaneously.

In the optimize-two-phases approach, it appears logical to optimize the goal function

$$G_{12}(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2) = \frac{F_{12}(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2)}{\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2}, \quad (66)$$

which considers the waiting times in the successive green phase  $\Delta T_2$  as well. The average delay time  $G_{12}(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2)$  is minimized by variation of *both* green time periods,  $\Delta T_1$  and  $\Delta T_2$ . The optimal green times are characterized by vanishing partial derivatives  $\partial G_{12}/\partial \Delta T_j$ . Therefore, we must find those values  $\Delta T_1$  and  $\Delta T_2$  which fulfil

$$\frac{\partial G_{12}}{\partial \Delta T_j} = \frac{\frac{\partial F_{12}}{\partial \Delta T_j}(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2) - F_{12}}{(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2)^2} = 0. \quad (67)$$

This implies the balancing principle

$$\frac{\partial F_{12}(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2)}{\partial \Delta T_1} = \frac{\partial F_{12}(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2)}{\partial \Delta T_2} \quad (68)$$

which is known from other optimization problems as well, e.g. in economics [47]. Condition (68) allows one to express the green time  $\Delta T_2$  as a function of the green time  $\Delta T_1$ . Both values can then be fixed by finding minima of  $G_w(\tau_1 + \Delta T_1 + \tau_2 + \Delta T_2(\Delta T_1))$ . When this optimization procedure is applied after completion of each phase, it is expected to be adaptive to changing traffic conditions. However, a weakness of the above approach is its neglect of the flows in the optimization procedure. Therefore, the resulting intersection throughput may be poor, and flows would not necessarily be served, when the intersection capacity would allow for this. Therefore, we will now modify the multiple-phase optimization in a suitable way, focusing on the two-phase case.

### 5.1 Combined flow-and-delay time optimization

The new element of the following approach is the introduction of flow constraints into the formulation of the delay time minimization. For this, let us start with the formula for the average delay time  $T_j^{\text{av}}$  in road section  $j$  derived in reference [28]. It reads

$$T_j^{\text{av}} = \frac{(1 - f_j)^2 T_{\text{cyc}}}{(1 - u_j) 2} \quad (69)$$

with

$$T_{\text{cyc}} = \tau_1 + \Delta T_1 + \tau_2 + \Delta T_2 = (\tau_1 + \tau_2)(1 + \sigma_1 + \sigma_2) \quad (70)$$

and

$$1 - f_j = \frac{T_{\text{cyc}} - \Delta T_j}{T_{\text{cyc}}} = \frac{(1 + \sigma_1 + \sigma_2) - \sigma_j}{1 + \sigma_1 + \sigma_2}. \quad (71)$$

As the number of vehicles arriving on road section  $j$  during the time period  $T_{\text{cyc}}$  is given by  $I_j A_j T_{\text{cyc}} = I_j u_j \hat{Q}_j T_{\text{cyc}}$ , the scaled overall delay time of vehicles during the two green phases  $\Delta T_1, \Delta T_2$  and amber time periods  $\tau_1, \tau_2$  covered by the cycle time  $T_{\text{cyc}}(\Delta T_1, \Delta T_2)$  is given by

$$G = \frac{\sum_{j=1}^2 T_j^{\text{av}} I_j u_j \hat{Q}_j T_{\text{cyc}}}{T_{\text{cyc}}} = \sum_{j=1}^2 \frac{[(1 + \sigma_1 + \sigma_2) - \sigma_j]^2}{2(1 - u_j)(1 + \sigma_1 + \sigma_2)} I_j u_j \hat{Q}_j (\tau_1 + \tau_2). \quad (72)$$

Let us now set  $\theta_j = \theta_j(\sigma_1, \sigma_2) = 0$ , if  $\sigma_j \leq \hat{\sigma}_j$  (corresponding to  $\sigma_j/(1 + \sigma_1 + \sigma_2) \leq u_j$ ), and  $\theta_j = 1$  otherwise. The dimensionless clearing time

$$\hat{\sigma}_j = \frac{u_j}{(1 - u_j)}(1 + \sigma_1 + \sigma_2 - \sigma_j) \quad (73)$$

was defined in equation (45). With this, we will minimize the scaled overall delay time (72) in the spirit of the optimize-two-cycles approach, but under the constraint that the average outflow

$$\begin{aligned} \bar{O} &= \sum_{j=1}^2 \frac{I_j \hat{Q}_j \{ \Delta T_j (1 - \theta_j) + [T_j + u_j (\Delta T_j - T_j)] \theta_j \}}{T_{\text{cyc}}} \\ &= \sum_{j=1}^2 \frac{I_j \hat{Q}_j \{ \sigma_j (1 - \theta_j) + [(1 - u_j) \hat{\sigma}_j + u_j \sigma_j] \theta_j \}}{1 + \sigma_1 + \sigma_2} \end{aligned} \quad (74)$$

reaches the maximum throughput

$$\hat{O}(u_1, u_2) = \min \left( G_t(u_1, u_2), O_{\text{max}}(u_1, u_2) \right). \quad (75)$$

The maximum throughput corresponds to the overall flow  $G_t(u_1, u_2) = I_1 A_1 + I_2 A_2 = u_1 I_1 \hat{Q}_1 + u_2 I_2 \hat{Q}_2$ , as long as the capacity constraint (62) is fulfilled. Otherwise, if the sum of arrival flows exceeds the intersection capacity, the maximum throughput is given by<sup>2</sup>

$$\begin{aligned} O_{\text{max}}(u_1, u_2) &= \max_{\substack{x_j \leq u_j \\ x_1 + x_2 = 1}} (x_1 I_1 \hat{Q}_1 + x_2 I_2 \hat{Q}_2) \\ &= \max_{1 - u_2 \leq x_1 \leq u_1} I_2 \hat{Q}_2 [K x_1 + (1 - x_1)] \\ &= \begin{cases} I_2 \hat{Q}_2 [(K - 1)u_1 + 1] & \text{if } K \geq 1 \\ I_2 \hat{Q}_2 [1 - (1 - K)(1 - u_2)] & \text{if } K < 1. \end{cases} \end{aligned} \quad (76)$$

Demanding the flow constraint

$$\bar{O}(\sigma_1(u_1, u_2), \sigma_2(u_1, u_2)) = \hat{O}(u_1, u_2) \quad (77)$$

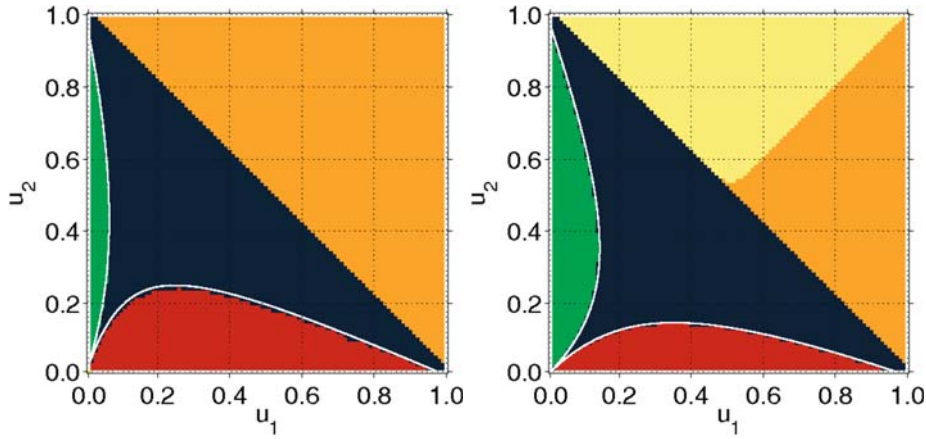
and considering equation (73), we can derive

$$\hat{O} = \sum_j I_j \hat{Q}_j \left[ \frac{\sigma_j (1 - \theta_j)}{1 + \sigma_1 + \sigma_2} + u_j \theta_j \right]. \quad (78)$$

This implies a linear relationship between  $\sigma_1$  and  $\sigma_2$ . If the denominator is non-zero, we have:

$$\begin{aligned} \sigma_1(\sigma_2) &= \frac{\theta_1 u_1 I_1 \hat{Q}_1 + \theta_2 u_2 I_2 \hat{Q}_2 - \hat{O}}{\hat{O} - (1 - \theta_1) I_1 \hat{Q}_1 - \theta_1 u_1 I_1 \hat{Q}_1 - \theta_2 u_2 I_2 \hat{Q}_2} \\ &\quad + \frac{(1 - \theta_2) I_2 \hat{Q}_2 + \theta_1 u_1 I_1 \hat{Q}_1 + \theta_2 u_2 I_2 \hat{Q}_2 - \hat{O}}{\hat{O} - (1 - \theta_1) I_1 \hat{Q}_1 - \theta_1 u_1 I_1 \hat{Q}_1 - \theta_2 u_2 I_2 \hat{Q}_2} \sigma_2. \end{aligned} \quad (79)$$

<sup>2</sup> If the cycle time  $T_{\text{cyc}}$  is limited to a certain maximum value  $T_{\text{cyc}}^{\text{max}}$ , one must replace the constraint  $x_1 + x_2 \leq 1$  by  $x_1 + x_2 \leq 1 - (\tau_1 + \tau_2)/T_{\text{cyc}}^{\text{max}}$  and  $1 - u_2$  by  $1 - u_2 - (\tau_1 + \tau_2)/T_{\text{cyc}}^{\text{max}}$ .



**Fig. 4.** (Color online) Operation regimes of periodic signal control as a function of the utilizations  $u_j$  of both road sections according to the two-phase optimization approach, assuming  $K = 1$ , corresponding to equal roads (left), and  $K = 3$ , corresponding to a three-lane road 1 and a one-lane road 2 (right). For most combinations of utilizations (if  $u_1$  is not too different from  $u_2$ ), the green phases are terminated as soon as the corresponding road sections are cleared (see the blue area below the falling diagonal line). However, extended green times for road Section 1 result (see the red area along the  $u_1$  axis), if the utilization of road Section 2 is small. In contrast, if the utilization of road Section 1 is small, extended green times should be given to road Section 2 (see the green area along the  $u_2$  axis). The white separating lines between these areas correspond to equations (94), (95) fit the numerical results well. Above the line  $u_2 = 1 - u_1$ , the intersection capacity is insufficient to serve the vehicle flows in *both* road sections. In this area, the two-phase optimization gives solutions where road Section 1 is fully cleared, but road Section 2 is served in part (orange area towards the right), or vice versa (yellow area towards the top in the left figure).

By demanding the flow constraint, we can guarantee that all arriving vehicles are served as long as the intersection capacity is sufficient, while we will otherwise use the maximum possible intersection capacity. As a consequence, operation regime (0) of the one-phase optimization, which neglected the service of at least one road section, cannot occur within this framework. Instead, it is replaced by an operation regime, in which the vehicle queue in one road section is fully cleared, while the vehicle queue in the other road section is served in part<sup>3</sup>. Of course, this will happen only, if the intersection capacity is insufficient to serve both flows completely (i.e. in the case  $1 - u_1 - u_2 < 0$ ). If  $K > 1$  (i.e. the main flow is on road Sect. 1), we have

$$\frac{\sigma_1}{1 + \sigma_1 + \sigma_2} = u_1 \quad \text{and} \quad \frac{\sigma_2}{1 + \sigma_1 + \sigma_2} = (1 - u_1). \quad (80)$$

If  $K < 1$ , the indices 1 and 2 must be interchanged.

Operation regime (1) is still defined as in Section 4.4 and characterized by

$$\sigma_j = \frac{u_j}{1 - u_1 - u_2}, \quad \frac{\sigma_j}{1 + \sigma_1 + \sigma_2} = u_j. \quad (81)$$

<sup>3</sup> In this case, we do not expect a periodic signal control anymore, as the growing vehicle queue in one of the road sections, see reference [28], has to be considered in the signal optimization procedure. Our formulas for one-phase optimization can handle this case due to the dependence on  $\Delta N_j(0)$ . In the two-phase optimization procedure, we would have to add  $\sum_j I_j \Delta N_j(0)$  to formula (72), where  $\Delta N_j(0) = A_j T_{\text{cyc}}^k - \hat{Q}_j \Delta T_j^k$  denotes the number of vehicles that was not served during the  $k$ th cycle  $T_{\text{cyc}}^k = \tau_1 + \Delta T_1^k + \tau_2 + \Delta T_2^k$ . This gives an additional term  $\sum_j u_j I_j \hat{Q}_j (\tau_1 + \tau_2) \sum_k (1 + \sigma_1^k + \sigma_2^k - \sigma_j^k / u_j)$  in equation (72).

In contrast to the one-phase optimization approach, this “normal case” of signal operation occurs in a large parameter area of the two-phase optimization approach (see blue area in Fig. 4). It implies that both green times are long enough to dissolve the vehicle queues, but not longer.

The case, where both green phases are extended, is again no optimal solution. We will, therefore, finally focus on case (2), where the vehicle queue in road Section 1 is just cleared ( $\theta_1 = 0$ ), while road Section 2 gets an excess green time ( $\theta_2 = 1$ ). With  $\hat{O} = G_t = u_1 I_1 \hat{Q}_1 + u_2 I_2 \hat{Q}_2$ , equation (79) yields the simple constraint

$$\sigma_1(\sigma_2) = \frac{u_1}{1 - u_1}(1 + \sigma_2), \quad (82)$$

which corresponds to equation (45). It implies

$$\frac{d\sigma_1}{d\sigma_2} = \frac{u_1}{1 - u_1}, \quad 1 + \sigma_1 = \frac{1 + u_1 \sigma_2}{1 - u_1}, \quad (83)$$

and

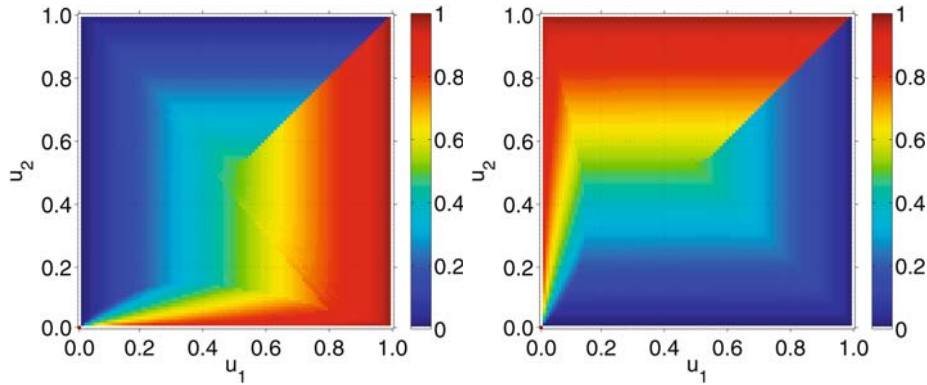
$$1 + \sigma_1 + \sigma_2 = \frac{1 + \sigma_2}{1 - u_1}, \quad \frac{\sigma_1}{1 + \sigma_1 + \sigma_2} = u_1. \quad (84)$$

We will now determine the minimum of the goal function  $G$  by setting the derivative  $\partial G / \partial \sigma_1$  to zero, considering

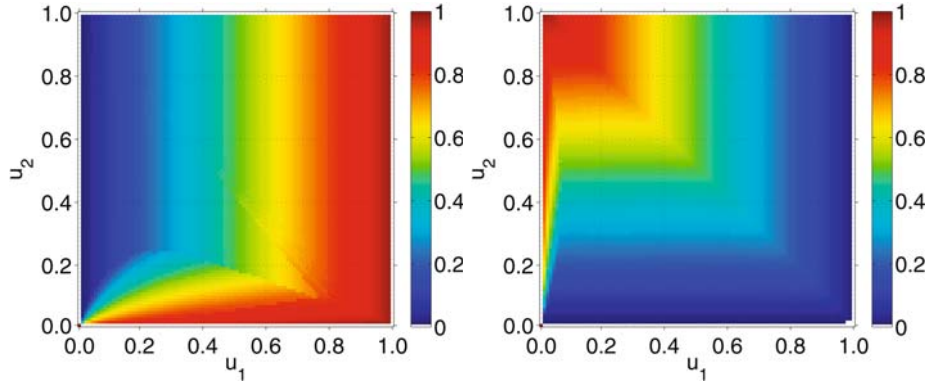
$$\frac{d\hat{\sigma}_2(\sigma_1)}{d\sigma_1} = \frac{u_2}{1 - u_2}. \quad (85)$$

Multiplying the result with  $2(1 - u_1)^3(1 - u_2)(1 + \sigma_1 + \sigma_2)^2 / (I_2 \hat{Q}_2)$ , we find the following relationship:

$$2u_1 u_2 (1 + \sigma_2)(1 + u_1 \sigma_2) + 2K u_1 (1 - u_1)(1 - u_2)(1 + \sigma_2)^2 = u_2 (1 + u_1 \sigma_2)^2 + K u_1 (1 - u_1)(1 - u_2)(1 + \sigma_2)^2, \quad (86)$$



**Fig. 5.** (Color online) Optimal green time fractions  $\Delta T_j/T_{\text{cyc}} = \sigma_j/(1 + \sigma_1 + \sigma_2)$  for road section  $j = 1$  (left) and road section  $j = 2$  (right) as a function of the utilizations  $u_j$  of both roads  $j$ , assuming periodic signal operation according to the two-phase optimization approach with  $K = 1$ . For combinations  $(u_1, u_2)$  with several solutions (with extended green time and without), we display the solution which minimizes the goal function (72). The results are qualitatively similar to the ones belonging to the one-phase optimization approach displayed in Figure 7, but we find periodic solutions above the capacity line  $u_2 = 1 - u_1$ , where one road section (the one with the greater utilization) is fully cleared, while the other one is served in part.



**Fig. 6.** (Color online) Same as Figure 5, but for  $K = 3$ , corresponding to a three-lane road Section 1 (arterial road) and a one-lane road Section 2 (crossing side road).

which finally leads to

$$\begin{aligned} (1 + \sigma_2)^2 &= \frac{u_2(1 - u_1)^2}{u_1^2 u_2 + K u_1(1 - u_1)(1 - u_2)} \\ &= \frac{(1 - u_1)^2}{u_1^2 + \kappa(1 - u_1)(1 - u_2)}. \end{aligned} \quad (87)$$

According to equation (60), for an extended green time on road Section 2, the condition

$$\frac{\sigma_2}{1 + \sigma_1 + \sigma_2} > u_2 \quad (88)$$

must again be fulfilled. If the solution  $\sigma_2(u_1, u_2)$  of equation (87) satisfies this requirement, it can be inserted into equation (82) to determine the scaled green time period  $\sigma_1(u_1, u_2)$  as a function of the capacity utilizations  $u_1$  and  $u_2$  within the framework of the optimize-two-phases approach. The corresponding results are displayed in Figures 4 to 6. A generalization to signal controls with more than two phases is straightforward.

Finally, let us calculate the separating line between case (1) and case (2). Inserting equation (82) into (72),

we can express the goal function  $G$  as a function  $H$  of a single variable  $\sigma_2$ :

$$H(\sigma_2) = G(\hat{\sigma}_1(\sigma_2), \sigma_2). \quad (89)$$

As equation (82) holds for both cases, an exact clearing of road Section 2 or an excess green time for it, the functional dependence of goal function (89) on  $\sigma_2$  must be the same for both cases. Now, on the one hand, we may apply equation (81) for the case without excess green time, which yields

$$1 + \sigma_2 = \frac{1 - u_1}{1 - u_1 - u_2} \quad \text{and} \quad (1 + \sigma_2)^2 = \frac{(1 - u_1)^2}{(1 - u_1 - u_2)^2}. \quad (90)$$

On the other hand, in the case of excess green time, we may use equation (87). The goal function must be the same along the separating line between both cases, which requires

$$\frac{(1 - u_1)^2}{(1 - u_1 - u_2)^2} = \frac{(1 - u_1)^2}{u_1^2 + \kappa(1 - u_1)(1 - u_2)}. \quad (91)$$

This implies

$$\kappa(1 - u_1)(1 - u_2) = (1 - u_1 - u_2)^2 - u_1^2 \quad (92)$$



or

$$\kappa(1 - u_1)(1 - u_2) = (1 - 2u_1 - u_2)(1 - u_2). \quad (93)$$

The finally resulting equation for the separating line between the regimes with and without excess green time is given by

$$\frac{1}{\kappa} = \frac{u_2}{Ku_1} = \frac{1 - u_1}{1 - 2u_1 - u_2}. \quad (94)$$

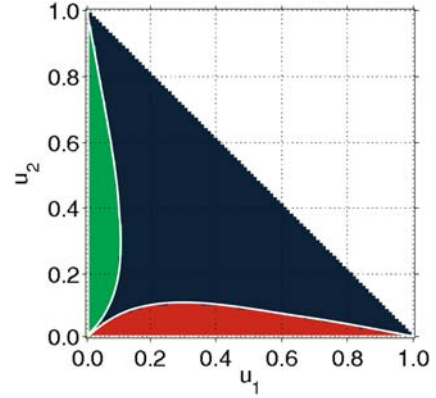
As Figure 4 shows, this analytical result fits the result of our numerical optimization very well. The separating line between case (1) and case (3) is derived analogously. It may also be obtained by interchanging the subscripts 1 and 2 and substituting  $\kappa$  by  $1/\kappa$ , yielding

$$\kappa = K \frac{u_1}{u_2} = \frac{1 - u_2}{1 - 2u_2 - u_1}. \quad (95)$$

## 6 Summary, discussion, and outlook

We have studied the control of traffic flows at a single intersection. Such studies have been performed before, but we have focussed here on some particular features:

- For the sake of a better understanding, we were interested in deriving analytical formulas, even though this required some simplifications.
- A one-phase minimization of the overall travel times in all road sections tended to give excess green times to the main flow, i.e. to the road section with the larger number of lanes or, if the number of lanes is the same ( $K = 1$ ), to the road section with the larger utilization (see Fig. 2). The excess green time can lead to situations where one of the vehicle flows is not served, although there would be enough service capacity for all flows.
- A minimization of vehicle queues rather than travel times simplifies the relationships through the special settings  $\hat{Q}_j = \hat{Q}$  and  $I_j = 1$ , resulting in  $K = 1$ . Moreover, these settings guarantee that the case of no service only occurs, if the intersection capacity is exceeded.
- An optimize-multiple-phases approach considering flow constraints gives the best results among the optimization methods considered. It makes sure that both roads are served even when the intersection capacity is exceeded.
- For all considered optimization approaches, we have derived different operation regimes of traffic signals control: one of them is characterized by ending a green time period upon service of the last vehicle in the queue, which implies that all vehicles are stopped once by a traffic signal. However, we have also found conditions under which it is advised to delay switching for one of the road sections (“slower-is-faster effect”), which allows some vehicles to pass the signal without stopping.
- Compared to the one-phase optimization, a two-phase optimization tends to give much less excess green



**Fig. 7.** (Color online) Operation regimes of periodic signal control as a function of the utilizations  $u_j$  of both roads, if one specifies the clearing times and excess green times according to equations (45) and (42) of the one-phase optimization, but selects the solution that minimizes the overall delay time (72) over two successive phases. For most combinations of utilizations (if  $u_1$  is not too different from  $u_2$ ), the green phases are terminated as soon as the corresponding road sections are cleared (see the blue area below the falling diagonal line). However, extended green times for road Section 1 result (see the red area along the  $u_1$  axis), if the utilization of road Section 2 is small. In contrast, if the utilization of road Section 1 is small, extended green times should be given to road Section 2 (see green area along the  $u_2$  axis) [48]. Above the line  $u_2 = 1 - u_1$ , the intersection capacity is insufficient to serve the vehicle flows in *both* road sections.

times, in particular if the utilizations of the road sections are comparable. We hypothesize that this is an effect of the short-sightedness of the one-phase optimization: it does not take into account future delay times caused by current excess green times. This hypothesis is confirmed by Figure 7 (which is to be contrasted with the left illustration in Fig. 2). It specifies the green time durations according to equations (42) and (45) of the one-phase optimization, but selects the solution that minimizes the scaled overall delay time (72) over two phases.

- Although the multi-phase optimization approach provides extended green times in a considerably smaller area of the parameter space spanned by the utilizations  $u_j$ , the slower-is-faster effect still persists when signal settings are optimized over a full cycle time (as we effectively did with the periodic two-phase optimization approach). The slower-is-faster effect basically occurs when the utilization of a road section is so small that it requires some extra time to collect enough vehicles for an efficient service during the green phase, considering the efficiency losses by switching traffic lights during the amber phases.
- In complementary appendices, we discuss traffic controls with more than two phases and an exponentially weighted goal function for short-term traffic optimization. Furthermore, we propose how to take into account the effect of stopping newly arriving vehicles and how to assess its impact as compared to queues of



waiting vehicles. As stopping vehicles causes additional delay times, it becomes often favorable to implement excess green times (i.e. to apply the slower-is-faster effect”).

In summary, our approach successfully delivers analytical insights into various operation regimes of traffic signal control, including the occurring slower-is-faster effects. Moreover, as the two-phase optimization approach takes care of side roads and minor flows, it has similar effects as the stabilization rule that was introduced in reference [25] to compensate for unstable service strategies. This stabilization rule tries to avoid spillover effects via an earlier green time by the next traffic light downstream.

Note that spillover effects imply growing delay times even in road sections which have a green light. Therefore, if the utilization is greater than the intersection capacity, travel time minimization may additionally demand to interrupt the green times of the next traffic lights upstream (in favor of a road section that could be successfully left by vehicles when a green light would be given to them). This effectively requires to generalize the traffic light control principle discussed before towards a consideration of the traffic conditions in upstream and downstream road sections. Such a control is considerably more complicated and will be addressed in future publications, based on formulas and principles developed in references [28,38].

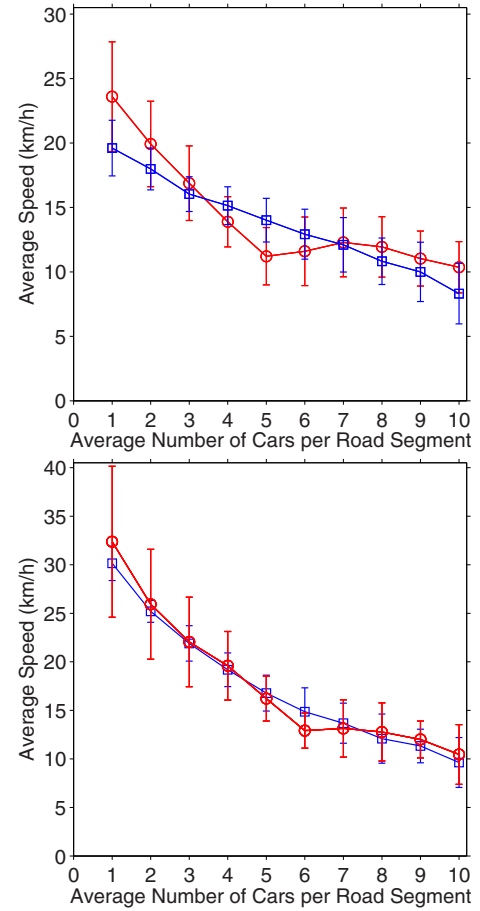
### 6.1 Self-organized traffic light control

Our restriction to analytical calculations implied certain simplifications such as the assumption of two traffic phases, the assumption of constant arrival flows, and no obstructions of the outflow. However, these restrictions can be easily overcome by straight-forward generalizations (see Appendices). The assumption of constant arrival flows, for example, is not needed. Assuming a short-term prediction based on upstream flow measurements [43], the expected delay times or queue lengths can be determined via the integral (17). The optimal solution must then be numerically determined, which poses no particular problems. Although the behavior may become somewhat more complicated and the boundaries of the operation regimes may be shifted, we expect that the above mentioned signal operation modes and the control parameters  $u_1 = A_1/\widehat{Q}_1$ ,  $u_2 = A_2/\widehat{Q}_2$ , and  $\kappa = I_1 A_1/(I_2 A_2)$  still remain relevant.

In the following, we show that the optimize-one-phase approach works surprisingly well, when it is applied to signal-controlled networks with their typical, pulsed vehicle flows. Rather than performing strict travel time optimization, however, we use a simplified approach that determines exponential averages  $A'_j(t)$  of the arrival flows  $A_j(t)$  according to

$$A'_j(t) = \alpha_j A_j(t) + (1 - \alpha_j) A'_j(t - 1), \quad (96)$$

and inserts these values into the formulas for the control strategies that were derived for constant arrival flows. The



**Fig. 8.** (Color online) Comparison of the average velocity resulting for an optimized fixed cycle time control (red circles) with a self-organized control based on the optimize-one-phase approach (blue squares) for a speed limit  $V_j^0 = 50$  km/h (top) and  $V_j^0 = 70$  km/h (bottom). Both control approaches perform similarly well. Error bars represent standard deviations. Details of the simulation scenarios are given in the main text.

averaging parameters  $\alpha_j$  are specified such that the average vehicle speed over 30 min is maximized.

Figure 8 shows simulation results for a Barcelona kind of road network (see Fig. 1) with 72 links, the lengths of which are uniformly distributed between 100 and 200 m. For simplicity, the turning fractions have been set to 1/2 for all intersections, the setup times  $\tau_j$  to  $\tau = 5$  s. Traffic flows were simulated in accordance with the section-based traffic model [24,44]. The parameters determining the assumed triangular flow-density relationship on the road sections are the safe time headway  $T = 1.8$  s, the maximum density  $\rho_{\max} = 140$  vehicles per kilometer, and the speed limit  $V_j^0$ , which is either set to 50 or to 70 km/h. As one can see, the average speed for the self-organized traffic light control performs similarly well as a fixed cycle strategy, where the cycle time is adjusted to the traffic volume. Specifically, the green times  $\Delta T_j$  are linearly increased from 15 s for an average number of 1 car per road section upto 60 s for an average number of 10 cars per road segment. The offsets of the green phases are optimized by

means of Particle Swarm Optimization (PSO) [49]. This serves to minimize the stopping of moving vehicle platoons.

A more detailed, numerical comparison of fixed cycle control schemes with self-organized traffic light controls for urban road networks will be presented in forthcoming publications. Note that, in reference [25], a somewhat more sophisticated self-control principle has been studied, which involves a short-term anticipation based on measurements of the arrival flows  $A_j$ . This self-control performs particularly well in cases of heterogeneous road networks and stochastically varying arrival flows, and it can create coordinated flow patterns similar to “green waves” (where vehicle platoons are not stopped at every traffic light).

*Author contributions:* DH set up the model, performed the analytical calculations, and wrote the manuscript. AM produced the numerical results and figures, and derived the separating lines given by equations (94) and (95).

The authors would like to thank for partial support by the ETH Competence Center “Coping with Crises in Complex Socio-Economic Systems” (CCSS) through ETH Research Grant CH1-01 08-2, the VW Foundation Project I/82 697 on “Complex Self-Organizing Networks of Interacting Machines: Principles of Design, Control, and Functional Optimization”, the Daimler-Benz Foundation Project 25-01.1/07 on “BioLogistics”, and the “Cooperative Center for Communication Networks Data Analysis”, a NAP project sponsored by the Hungarian National Office of Research and Technology under Grant No. KCKHA005.

## Appendix A: Considering the price of stopping vehicles

The previous considerations have only taken into account delays by vehicles in a vehicle queue. However, it would also make sense to consider the price of stopping vehicles. In particular, it must be possible that a large flow of moving vehicles in one road section is prioritized to a short queue of standing vehicles in the other road section. But how can we assess the relative disadvantage of stopping newly arriving vehicles as compared to stopping the service of a vehicle queue at the intersection? If the arrival flow is not large enough, it would certainly be better to continue serving the standing vehicle queue in the other road until it is fully dissolved.

We pursue the following approach: while the flow model used before implicitly assumes instantaneous vehicle accelerations and decelerations, we will now consider that, in reality, a finite vehicle acceleration  $a$  causes additional delays of  $V_j^0/(2a)$ , where  $V_j^0$  denotes the free speed or speed limit. Furthermore, the reaction time  $T_r$  must be taken into account as well. This leads to an additional delay of

$$T_j' = T_r + \frac{V_j^0}{2a} \quad (97)$$

for each vehicle that leaves a queue.  $T_r$  is of the order of the safe time gap  $T$ . Note that delays  $V_j^0/(2b)$  due to a

finite deceleration  $b$  do not additionally contribute to the delay times, as it does not matter whether delayed vehicles spend their time decelerating or stopped<sup>4</sup>.

Furthermore, we must determine the rate at which such additional delays are produced. This is given by the rate at which freely moving vehicles join the end of a traffic jam, i.e. by

$$\rho_{\text{jam}}|C_j| = \frac{\rho_{\text{jam}}}{\rho_{\text{jam}}/A_j - 1/V_j^0} \geq A_j, \quad (98)$$

where  $\rho_{\text{jam}}$  denotes the density of vehicles per lane in a standing queue. The propagation speed

$$C_j = \frac{A_j - 0}{A_j/V_j^0 - \rho_{\text{jam}}} \quad (99)$$

of the upstream front of the queue corresponds to the propagation speed of shock fronts, see references [24,44,50]. Depending on the values of  $C_j$  (or  $A_j$ ) and  $T_j'$ , newly arriving vehicles can have an impact  $T_j' C_j \rho_{\text{jam}}$  equivalent to about  $\Delta N_j = 10$  queued vehicles.

Summarizing the above considerations, we suggest to replace the goal function  $G_1(t)$  by the generalized formula

$$\widehat{G}_1(t) = \frac{1}{t} \sum_j I_j \int_0^t dt' \left[ \Delta N_j(t') + T_j' |C_j| \rho_{\text{jam}} \Theta(\Delta N_j > 0) \right], \quad (100)$$

where  $\Theta(\Delta N_j > 0) = 1$ , if  $\Delta N_j > 0$ , and  $\Theta(\Delta N_j > 0) = 0$  otherwise. In case (a) with  $\Delta T_i \leq T_i$ , we find

$$\begin{aligned} \widehat{F}_1^a(\tau_1 + \Delta T_i + \tau_2) &= F_1^a(\tau_1 + \Delta T_i + \tau_2) \\ &+ I_1 T_1' |C_1| \rho_{\text{jam}} (\tau_1 + \Delta T_1 + \tau_2) \\ &+ I_2 T_2' |C_2| \rho_{\text{jam}} (\tau_1 + \Delta T_1 + \tau_2). \end{aligned} \quad (101)$$

This implies

$$\begin{aligned} \widehat{G}_1^a(\tau_1 + \Delta T_i + \tau_2) &= G_1^a(\tau_1 + \Delta T_i + \tau_2) \\ &+ I_1 T_1' |C_1| \rho_{\text{jam}} + I_2 T_2' |C_2| \rho_{\text{jam}} \end{aligned} \quad (102)$$

with  $G_1^a(\tau_1 + \Delta T_i + \tau_2)$  according to equation (22). Therefore, the partial derivative of  $\widehat{G}_1^a(\tau_1 + \Delta T_i + \tau_2)$  with respect to  $\Delta T_1$  remains unchanged, and we find the same optimal green time period  $\Delta T_1 = 0$  or  $\Delta T_1 \geq T_1$ . However, in case (b) with  $\Delta T_1 \geq T_1$ , we obtain

$$\begin{aligned} \widehat{F}_1^b(\tau_1 + \Delta T_i + \tau_2) &= F_1^b(\tau_1 + \Delta T_i + \tau_2) \\ &+ I_1 T_1' |C_1| \rho_{\text{jam}} (\tau_1 + T_1 + \tau_2) \\ &+ I_2 T_2' |C_2| \rho_{\text{jam}} (\tau_1 + \Delta T_1 + \tau_2), \end{aligned} \quad (103)$$

<sup>4</sup> The finite deceleration only matters slightly, when the exact moment must be determined when a road section becomes fully congested.

which implies

$$\begin{aligned} \widehat{G}_1^b(\tau_1 + \Delta T_i + \tau_2) &= G_1^b(\tau_1 + \Delta T_i + \tau_2) \\ &+ I_1 T_1' |C_1| \rho_{\text{jam}} + I_2 T_2' C_2 \rho_{\text{jam}} - I_1 T_1' |C_1| \rho_{\text{jam}} \frac{\Delta T_1 - T_1}{\tau_1 + \Delta T_1 + \tau_2} \end{aligned} \quad (104)$$

with  $G_1^b(\tau_1 + \Delta T_i + \tau_2)$  according to equation (28). In cases where an excess green time is favorable, the corresponding formula for the green time duration becomes

$$(\tau_1 + \Delta T_1 + \tau_2)^2 = \frac{2I_1}{I_2 A_2} [E_1 + T_1' |C_1| \rho_{\text{jam}} (\tau_1 + T_1 + \tau_2)], \quad (105)$$

i.e. the optimal green times tend to be longer. In order to support excess green times, the condition  $(\tau_1 + \Delta T_1 + \tau_2)^2 \geq (\tau_1 + T_1 + \tau_2)^2$  must again be fulfilled, which requires

$$\begin{aligned} \frac{(\Delta N_1^{\text{max}})^2}{\widehat{Q}_1 - A_1} \left( \frac{I_1}{I_2 A_2} - \frac{1}{\widehat{Q}_1 - A_1} \right) \\ + 2\Delta N_1^{\text{max}} \left( \frac{I_1 \tau_1}{I_2 A_2} - \frac{\tau_1 + \tau_2}{\widehat{Q}_1 - A_1} \right) \geq (\tau_1 + \tau_2)^2 \\ - \frac{2I_1 T_1' |C_1| \rho_{\text{max}}}{I_2 A_2} \left( \tau_1 + \frac{\Delta N_1^{\text{max}}}{\widehat{Q}_1 - A_1} + \tau_2 \right). \end{aligned} \quad (106)$$

Comparing this with formula (32), we can see that the threshold for the implementation of excess green times  $\Delta T_j > T_j$  is reduced. Therefore, excess green times will be implemented more frequently, as this reduces the number of stopped vehicles.

## Appendix B: More than two traffic phases

The above formulas for the optimize-one-phase approach can be easily generalized to multiple traffic phases of more complicated intersections as in the case of Barcelona's center (see Fig. 1). For

$$\Delta T_i \leq T_i = \frac{\Delta N_i^{\text{max}}}{\widehat{Q}_i - A_i} \quad (107)$$

with

$$\Delta N_i^{\text{max}} = \Delta N_i(0) + A_i \tau_i, \quad (108)$$

for example, we can derive from equation (102)

$$\begin{aligned} \widehat{G}_i^a(\tau_i + \Delta T_i + \tau_{i+1}) &= I_i \left[ \Delta N_i(0) + \widehat{Q}_i \tau_{i+1} \right. \\ &- \left. (\widehat{Q}_i - A_i) \frac{\tau_i + \Delta T_i + \tau_{i+1}}{2} \right] + \sum_{j(\neq i)} I_j \left[ \Delta N_j(0) \right. \\ &+ \left. A_j \frac{\tau_i + \Delta T_i + \tau_{i+1}}{2} \right] + \sum_j I_j T_j' |C_j| \rho_{\text{jam}}. \end{aligned} \quad (109)$$

In contrast, for  $\Delta T_i \geq T_i$  and with

$$E_i = \Delta N_i^{\text{max}} \tau_i + \frac{(\Delta N_i^{\text{max}})^2}{2(\widehat{Q}_i - A_i)}, \quad (110)$$

from equations (104) and (28) we obtain

$$\begin{aligned} \widehat{G}_i^b(\tau_i + \Delta T_i + \tau_{i+1}) &= \frac{I_i E_i}{\tau_i + \Delta T_i + \tau_{i+1}} \\ &+ \sum_{j(\neq i)} I_j \left[ \Delta N_j(0) + A_j \frac{\tau_i + \Delta T_i + \tau_{i+1}}{2} \right] \\ &+ \sum_j I_j T_j' |C_j| \rho_{\text{jam}} - I_i T_i' |C_i| \rho_{\text{jam}} \frac{\Delta T_i - T_i}{\tau_i + \Delta T_i + \tau_{i+1}}. \end{aligned} \quad (111)$$

The minimum of this function is reached for

$$(\tau_i + \Delta T_i + \tau_{i+1})^2 = \frac{I_i E_i + I_i T_i' |C_i| \rho_{\text{jam}} (\tau_i + T_i + \tau_{i+1})}{\sum_{j(\neq i)} I_j A_j / 2}. \quad (112)$$

The occurrence of excess green time requires  $(\tau_i + \Delta T_i + \tau_{i+1})^2 \geq (\tau_i + T_i + \tau_{i+1})^2$ , i.e.

$$\begin{aligned} \frac{(\Delta N_i^{\text{max}})^2}{\widehat{Q}_i - A_i} \left( \frac{I_i}{\sum_{j(\neq i)} I_j A_j} - \frac{1}{\widehat{Q}_i - A_i} \right) \\ + 2\Delta N_i^{\text{max}} \left( \frac{I_i \tau_1}{\sum_{j(\neq i)} I_j A_j} - \frac{\tau_i + \tau_{i+1}}{\widehat{Q}_i - A_i} \right) \geq (\tau_i + \tau_{i+1})^2 \\ - \frac{2I_i T_i' |C_i| \rho_{\text{jam}}}{\sum_{j(\neq i)} I_j A_j} \left( \tau_i + \frac{\Delta N_i^{\text{max}}}{\widehat{Q}_i - A_i} + \tau_{i+1} \right). \end{aligned} \quad (113)$$

It can be seen that the existence of more traffic phases is unfavorable for providing excess green times. For their existence, a small number of phases is preferable.

## Procedure of traffic signal control

Based on the above formulas, the next green phase  $i$  is determined as follows:

1. Set the time  $t$  to zero, after the last green phase  $i'$  has been completed.
2. Apply the required service time (amber time) of duration  $\tau_{i'+1}$  and set  $\tau_j = \tau_{i'+1}$  for all road sections  $j$ . Then, calculate  $\Delta N_j^{\text{max}}$  and  $E_j$  for all  $j$  with formulas (108) and (110).
3. During the service time, determine the green times  $\Delta T_j$  and  $T_j$  with and without green time extension, for each road section  $j$  with formulas (112) and (107).
4. If  $\Delta T_j > T_j$  and  $\widehat{G}_j^b(\tau_j + \Delta T_j + \tau_{j+1}) < \widehat{G}_j^a(\tau_j + T_j + \tau_{j+1})$ , see equations (111) and (109), consider the implementation of the extended green time  $\Delta T_j$  and set  $\widehat{G}_j = \widehat{G}_j^b(\tau_j + \Delta T_j + \tau_{j+1})$ . Otherwise consider the implementation of the clearing time  $T_j$  and set  $\widehat{G}_j = \widehat{G}_j^a(\tau_j + T_j + \tau_{j+1})$ , but if  $\widehat{G}_j^a(\tau_j + \tau_{j+1}) < \widehat{G}_j$ , set  $\Delta T_j = 0$  and  $\widehat{G}_j = \widehat{G}_j^a(\tau_j + \tau_{j+1})$ .
5. Among all road sections  $j'$  different from the previously selected one  $i'$ , choose that one  $i$  for service, for which the expected average travel time  $\widehat{G}_i$  is smallest (i.e.  $\widehat{G}_i = \min_{j(\neq i')} \widehat{G}_j$ ). Implement the selected green phase  $\Delta T_i$ .

6. Update the length of the vehicle queue in road section  $i$  according to

$$\Delta N_i(\tau_i + \Delta T_i) = 0 \quad (114)$$

and the queue lengths in all other road sections  $j \neq i$  according to

$$\Delta N_j(\tau_i + \Delta T_i) = \Delta N_j(0) + A_j(\tau_i + \Delta T_i). \quad (115)$$

If road section was not served ( $\Delta T_i = 0$ ), update the vehicle queues in *all* road sections  $j$  (including  $i$ ) according to equation (115).

7. At the end of the corresponding green time duration  $\Delta T_i$ , set  $i' = i$  and continue with step 1.

The optimize-multiple-phases approach can be generalized in a similar way. Then, among all solutions satisfying preset flow constraints, that multi-phase solution is chosen, which minimizes the goal function and does not start with a service of the previously served road section. In order to flexibly adjust to varying traffic conditions, one may repeat the optimization after completion of one phase rather than after completion of all the phases considered in the multi-phase optimization.

### Appendix C: Limited forecast time horizon

While traffic light optimization is an NP-hard problem [46], we have simplified it here considerably by restricting ourselves to local optimization and to limited time horizons. Both simplifications may imply a potentially reduced traffic performance in the urban street network, but this loss of performance is small if traffic lights adjust to arriving vehicle platoons [25]. The reliable look-ahead times are anyway very limited for fundamental reasons (see the Appendix in Ref. [25]). Therefore, one can restrict traffic light optimization to time periods  $1/\lambda$ , over which the traffic forecast can be done with sufficient accuracy. When traffic lights are switched frequently, the value of  $1/\lambda$  of the forecast time horizon will go down.

Note that an optimization based on unreliable long-term forecasts will yield bad results. Therefore, it is not only *justified*, but also *successful* to replace the optimization of one or several full cycles by the optimization of, say, two phases. Alternatively, one may minimize the exponentially weighted travel times, i.e. minimize the function

$$\tilde{G} = \sum_j \lambda I_j \int_0^{\infty} dt e^{-\lambda t} \left[ \Delta N_j(t) + T_j' |C_j| \rho_{\text{jam}} \Theta(\Delta N_j > 0) \right] \quad (116)$$

by variation of the duration and sequence of green phases. While this approach is less suited for an analytical optimization, it reminds of formulations of discounted functions in economics [47]. Goal function (116) can be optimized *numerically*, limiting the evaluation of the integral to the range  $t < 3/\gamma$ .

### References

1. O. Biham, A.A. Middleton, D. Levine, Phys. Rev. A **46**, R6124 (1992)
2. T. Nagatani, Phys. Rev. E **48**, 3290 (1993)
3. J. Esser, M. Schreckenberg, Int. J. Mod. Phys. B **8**, 1025 (1997)
4. P.M. Simon, K. Nagel, Phys. Rev. E **58**, 1286 (1998)
5. D. Chowdhury, A. Schadschneider, Phys. Rev. E **59**, R1311 (1999)
6. E. Brockfeld, R. Barlovic, A. Schadschneider, M. Schreckenberg, Phys. Rev. E **64**, 056132 (2001)
7. M. Fouladvand, M. Nematollahi, Eur. Phys. J. B **22**, 395 (2001)
8. M. Sasaki, T. Nagatani, Phys. Stat. Mech. Appl. **325**, 531 (2003)
9. B.A. Toledo, V. Munoz, J. Rogan, C. Tenreiro, J.A. Valdivia, Phys. Rev. E **70**, 016107 (2004)
10. T. Nagatani, Phys. Stat. Mech. Appl. **368**, 560 (2006)
11. M. Garavello, B. Piccoli, Discrete Contin. Dyn. Syst. B **5**, 599 (2005)
12. R. Barlovic, T. Huisinga, A. Schadschneider, M. Schreckenberg, Adaptive traffic light control in the ChSch model for city traffic, in *Traffic and Granular Flow'03*, edited by P.H.L. Bovy, S.P. Hoogendoorn, M. Schreckenberg, D.E. Wolf (Springer, Berlin, 2004)
13. M.E. Fouladvand, Z. Sadjadi, M.R. Shaebani, J. Phys. A **37**, 561 (2004)
14. M.E. Fouladvand, M.R. Shaebani, Z. Sadjadi, J. Phys. Society Jpn **73**, 3209 (2004)
15. B. Faieta, B.A. Huberman, *Firefly: A synchronization strategy for urban traffic control*, Internal Report No. SSL-42 (Xerox PARC, Palo Alto, CA, 1993)
16. D.-W. Huang, W.-N. Huang, Phys. Rev. E **67**, 056124 (2003)
17. S. Lämmer, H. Kori, K. Peters, D. Helbing, Physica A **363**, 39 (2006)
18. T. Nakatsuji, S. Seki, T. Kaku, Transport. Res. Rec. **1324**, 137 (1995)
19. K. Sekiyama, J. Nakanishi, I. Takagawa, T. Higashi, T. Fukuda, IEEE Int. Conf. Syst. Man. Cybern. **4**, 2481 (2001)
20. D. Helbing, R. Jiang, M. Treiber, Phys. Rev. E **72**, 046130 (2005); R. Jiang, D. Helbing, P.K. Shukla, Q.-S. Wu, Physica A **368**, 567 (2006)
21. C. Gershenson, Complex Syst. **16**, 29 (2005)
22. S.-B. Cools, C. Gershenson, B. D'Hooghe, *Self-organizing traffic lights: A realistic simulation*, *Advances in Applied Self-Organizing Systems*, edited by M. Prokopenko (Springer, New York, 2007)
23. D. Helbing, S. Lämmer, J.-P. Lebacque, Self-organized control of irregular or perturbed network traffic, in *Optimal Control and Dynamic Games*, edited by C. Deissenberg, R.F. Hartl (Springer, Dordrecht, 2005), pp. 239–274
24. D. Helbing, J. Siegmeier, S. Lämmer, NHM **2**, 193 (2007)
25. S. Lämmer, D. Helbing, J. Stat. Mech. (JSTAT) P04019 (2008)
26. S. Lämmer, *Reglerentwurf zur dezentralen Online-Steuerung von Lichtsignalanlagen in Straßennetzwerken*, *Controller design for a decentralized control of traffic lights in urban road networks*, Ph.D. thesis, Dresden University of Technology, 2007

27. D. Schrank, T. Lomax, *The 2005 Urban Mobility Report* (Texas Transportation Institute, 2005)
28. D. Helbing, Eur. Phys. J. B (2009), DOI: 10.1140/epjb/e2009-00093-7
29. M. Schönhof, M. Treiber, A. Kesting, D. Helbing, Transp. Res. Rec. **1999**, 3 (2007)
30. A. Kesting, M. Treiber, M. Schönhof, D. Helbing, Transp. Res. Rec. **2000**, 16 (2007)
31. D. Braess, Unternehmensforsch. **12**, 258 (1968)
32. R. Steinberg, R.E. Stone, Transp. Sci. **22**, 231 (1988)
33. T. Roughgarden, *Selfish Routing and the Price of Anarchy*. (MIT Press, 2005)
34. D. Helbing, I. Farkas, T. Vicsek, Nature **407**, 487 (2000)
35. R.J. Smeed, Transp. Sci. **1**, 308 (1967)
36. M. Ben-Akiva, A. de Palma, Transp. Sci. **20**, 52 (1986)
37. R.B. Cooper, S.-C. Niu, M.M. Srinivasan, Manag. Sci. **44**, 1079 (2000)
38. D. Helbing, S. Lämmer, Verfahren zur Koordination konkurrierender Prozesse oder zur Steuerung des Transports von mobilen Einheiten innerhalb eines Netzwerkes. Method for coordination of concurrent processes for control of the transport of mobile units within a network. Patent WO/2006/122528 (2006)
39. D. Helbing, T. Seidel, S. Lämmer, K. Peters Self-organization principles in supply networks and production systems, in *Econophysics and Sociophysics – Trends and Perspectives*, edited by B.K. Chakrabarti, A. Chakraborti, A. Chatterjee (Wiley, Weinheim, 2006), pp. 535–558
40. D. Helbing, R. Jiang, M. Treiber, Phys. Rev. E **72**, 046130 (2005)
41. D. Helbing, A. Johansson, J. Mathiesen, M.H. Jensen, A. Hansen, Phys. Rev. Lett. **97**, 168001 (2006)
42. H.-U. Stark, C.J. Tessone, F. Schweitzer, Adv. Complex Syst. **11**, 551 (2008)
43. S. Lämmer, R. Donner, D. Helbing, Eur. Phys. J. B **63**, 341 (2008)
44. D. Helbing, J. Phys. Math. Gen. **36**, L593 (2003)
45. D. Helbing, Production, supply, and traffic systems: A unified description, in *Traffic and Granular Flow '03* edited by S.P. Hoogendoorn, S. Luding, P.H.L. Bovy, M. Schreckenberg, D.E. Wolf (Springer, Berlin, 2005), pp. 173–188
46. C.H. Papadimitriou, J.N. Tsitsiklis, Math. Oper. Res. **24**, 293 (1999)
47. G. Feichtinger, R.F. Hartl, *Optimale Kontrolle ökonomischer Prozesse. Optimal Control of Economic Processes* (de Gruyter, Berlin, 1986)
48. C. Chase, P.J. Ramadge, IEEE Trans. Autom. Contr. **37**, 491 (1992)
49. J. Kennedy, R. Eberhart, Particle swarm optimization, in: *Proceedings of the IEEE International Conference on Neural Networks* (Perth, WA, Australia, 1995), Vol. 4, pp. 1942–1948
50. G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974)