

# Empirical Features of Congested Traffic States and Their Implications for Traffic Modeling

Martin Schönhof, Dirk Helbing

Institute for Economics and Traffic, Dresden University of Technology, Andreas-Schubert Str. 23, 01062 Dresden, Germany  
{martin@vwi.tu-dresden.de, helbing@vwi.tu-dresden.de}

We address the controversial issue of traffic flow modeling, whether first-order, second-order, or other traffic models are best supported by empirical facts and theoretical considerations. This is done by a critical discussion of the pros and cons of the different theoretical approaches and by the analysis of a large set of empirical data with new evaluation techniques. Specifically, we investigate characteristic properties of the congested traffic states on a 30-km-long stretch of the German freeway A5 near Frankfurt/Main. Among the approximately 245 breakdowns of traffic flow at several different bottlenecks in 165 days, we have identified five different kinds of spatiotemporal congestion patterns and their combinations. Based on an “adaptive smoothing method” for the visualization of detector data, we also discuss particular features of breakdowns, such as signs of unstable traffic flow and the “boomerang effect,” which often seems to be caused by overtaking maneuvers of trucks. Controversial issues such as “synchronized flow” or stop-and-go waves are addressed as well. Our empirical results are compared with the implications of different theoretical concepts such as first-order traffic models and the phase diagram of congested traffic states predicted by some second-order models and the nonlocal, gas-kinetic based traffic model (GKT model). For a correct understanding of empirical observations such as the “general pattern,” it is important to consider particularities such as the fact that off-ramps can act as bottlenecks, when activated by downstream on-ramp bottlenecks. As sequences of off- and on-ramps generate different congestion patterns than single on-ramps, they must be treated as interconnected bottlenecks. Furthermore, our empirical results question Kerner’s three-phase theory.

**Key words:** traffic flow theories; characterization of empirical traffic states; traffic dynamics; fundamental diagram; phase diagram; three-phase traffic theory

**History:** Received: May 2004; revision received: November 2006; accepted: December 2006.

## 1. Introduction

In the past years, the field of traffic modeling has been characterized by a controversy about the right understanding of one of its most fundamental subjects: traffic jams. One can distinguish at least three mutually inconsistent standpoints:

1. The Lighthill-Whitham-Richards model (LWR model) is sufficient to understand the observed traffic phenomena (Daganzo 1995b, 2002b; Laval and Daganzo 2004; Daganzo 1995a, 1994; Cassidy and Mauch 2001; Lebacque 1995; Lebacque and Khoshyaran 2005).

2. So-called second-order models would be needed to describe the observations (Nagel and Nelson 2005; Helbing 2001b).

3. All previous approaches assuming a fundamental diagram, including the LWR and second-order models, are wrong (Kerner 1999, 2004).

This controversy may be considered so fundamental that it questions traffic flow modeling as a scientific discipline. Some confusion has been caused by the evergrowing number of special terms invented for certain observations, so that it is often unclear

whether certain terms are supposed to describe the same observation or simulated traffic state, or not. Additional confusion has been caused by questioning certain theories for the wrong reasons. We will discuss these issues in §§2.2, 3.2, and 7.1. A third source of confusion is the interpretation of empirical measurement. This will be addressed in §5.1.

Our presentation is structured as follows: §§2 and 3 give a critical discussion of first- and second-order traffic flow models to point out their essential features and identify the unresolved questions. While this part has the features of a survey, it is needed to resolve some misunderstandings in the literature and remind us of relevant arguments that may have been forgotten. Section 2.5 and the sections from 4.2 on will then present our own new results, both theoretical and empirical.

In the following, we give a more detailed overview: §2 will shortly introduce the Lighthill-Whitham-Richard (LWR) first-order traffic model and its properties. Based on this, we will discuss what criticisms have been raised against it and how seriously they must be taken. In particular, we will give empir-

ical evidence of growing perturbations pointing to an instability of traffic flow (see §2.5). As these are inconsistent with the LWR model, §3 will introduce second-order traffic models that imply unstable traffic flows. However, second-order models have been blamed for theoretical inconsistencies. Therefore, §3.2 will discuss whether and how these criticisms can be overcome. This finally leads us to the nonlocal, gas-kinetic-based traffic model (GKT model) in §4, which is theoretically consistent. It predicts different kinds of congested traffic states, which will be introduced in §4.1. These states are predicted to occur under certain conditions concerning the freeway flow and bottleneck strengths, which can be illustrated by means of a phase diagram (see §4.2). In contrast to previous publications, we will distinguish situations of small and large perturbations, which imply different phase diagrams due to the predicted multistable nature of traffic flows. This has great significance for the comparison with empirical data of congested traffic states, which is the subject of §6. Before that, §5 motivates the choice of our measurement site and describes its features. It also introduces an adaptive smoothing method, which is suited to visualize spatiotemporal traffic patterns in a three-dimensional way and to distinguish small but significant perturbations from random fluctuations (see §5.3). This method helps to classify different traffic states (see §6), to understand spatial or temporal transitions between them (see §§6.7, 6.8), and to determine an empirical phase diagram (see §6.6). Finally, §7 presents our summary and conclusions, including a discussion of Kerner's criticism of traffic models (see §7.1) and his attempts to classify and explain empirical observations from the same measurement site by the three-phase theory and its extensions (see §7.2).

## 2. The Lighthill-Whitham-Richards Model and Its Limits

### 2.1. Form and Properties of the Lighthill-Whitham-Richards Model (LWR Model)

One of the oldest and most-cited models of traffic flow is the fluid dynamic model by Lighthill and Whitham (1955) and Richards (1956). It is based on the continuity equation

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial Q(x, t)}{\partial x} = v_+(x, t) - v_-(x, t), \quad (1)$$

where  $\rho(x, t)$  denotes the vehicle density and  $Q(x, t)$  the traffic flow as a function of location  $x$  and time  $t$ . This equation is a direct mathematical implication of the conservation of the number of vehicles, i.e., the fact that no vehicle is produced or lost along the freeway (apart from cases of accidents, maybe). Therefore,

it is one of the most fundamental laws of traffic flows and applies without any doubt (with the restriction that traffic flows are not continuous, but composed of discrete vehicles).

For the time being, we will set the source terms  $v_{\pm}(x, t)$  due to ramp flows equal to zero (where the plus sign corresponds to on-ramps and the minus sign to off-ramps). Lighthill, Whitham, and Richard have closed the above continuity equation assuming a direct flow-density relationship of the kind

$$Q(x, t) = Q(\rho(x, t)), \quad (2)$$

which is called the “fundamental diagram.” This requires an adiabatic (i.e. effectively instantaneous) adaptation of the traffic flow and average vehicle speed  $V = Q/\rho$  to the density  $\rho$ . Because the resulting “fluid-dynamic” traffic model does not contain a second-order derivative (such as a diffusion term  $D\partial^2\rho/\partial x^2$ ), it is called a first-order model.

One can rewrite the LWR model (1), (2) as a nonlinear wave equation

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{dQ(\rho)}{d\rho} \frac{\partial \rho(x, t)}{\partial x} = 0. \quad (3)$$

According to this, the density profile propagates with the speed  $c(\rho) = dQ(\rho)/d\rho$ , i.e., a formal solution of Equation (3) is given by

$$\rho(x, t) = \rho\left(x - \int_0^t dt' c(\rho(y_x(t'), t')), 0\right), \quad (4)$$

where

$$y_x(t') = x - \int_{t'}^t dt'' v(t'') \quad \text{with } \frac{dy_x(t')}{dt'} = v(t') = c(\rho(y_x(t'), t')) \quad (5)$$

is the location with vehicle density  $\rho$  as a function of time  $t'$ . As this defines an implicit equation for the density  $\rho(x, t)$ , it is hard to solve explicitly (if  $c$  is not a constant). However, we can see from Equation (4) that the initial density profile  $\rho(x, 0)$  does not change its amplitude in the course of time.<sup>1</sup> The shape of the density profile can change if the propagation velocity  $c(\rho)$  is density dependent. For example, it is possible that initially smooth density profiles will eventually form shock fronts. These are predicted to propagate with the velocity

$$C(\rho_+, \rho_-) = \frac{Q(\rho_+) - Q(\rho_-)}{\rho_+ - \rho_-}, \quad (6)$$

<sup>1</sup> According to Kerner, Klenov, and Konhäuser (1997), however, the interaction of shock waves can lead to a reduction of the amplitude. The opposite—i.e., an increase of the amplitude—has, to our knowledge, not been reported.

where  $\rho_+$  denotes the density immediately upstream of the shock front and  $\rho_-$  the density immediately downstream of it (Whitham 1974). Because this equation is a consequence of vehicle conservation, it is, by the way, also expected to be approximately valid for congestion fronts in second-order models.

## 2.2. Criticisms of the Lighthill-Whitham-Richard Model

From a theoretical point of view, the LWR model should be able to describe the propagation of upstream and downstream shock fronts along homogeneous road sections and the formation of vehicle queues upstream of bottlenecks. Empirical studies suggest that formula (6) is indeed suitable to describe the propagation of upstream shock fronts of spatially extended congestion patterns (Muñoz and Daganzo 2003). Even Kerner uses this formula in his patented traffic state estimation tool ASDA/FOTO (Kerner et al 2004; Kerner, Kirschfink, and Rehborn 1998; Kerner and Rehborn 2000; Kerner, Aleksic, and Denneler 2001; Kerner 2001), although he has strongly criticized the LWR model for not describing two subsequent wide moving jams well (Kerner, Klenov, and Konhäuser 1997). Kerner and Konhäuser (1994a), stated that, according to the LWR model, when assuming a parabolic fundamental diagram, the downstream shock fronts would eventually change their shape and become smoother and smoother. In contrast, the experimental data support a stable velocity profile over time, moving upstream with constant velocity (Kerner and Rehborn 1996a).

Although Lin and Lo (2003) have pointed out that two wide moving clusters propagating at the same speed can be simulated with the LWR model, if the fundamental diagram  $Q(\rho)$  is linear in the congested (high-density) region, Kerner's criticism has not been convincingly addressed. Lin and Lo have *not* shown that, without being sensitive to the assumed form of the fundamental diagram and to initial and boundary conditions, wide moving jams have invariant, continuous wave profiles and characteristic constants as results of self-organization. According to Kerner (1998b), these characteristic constants include the resolution speed  $c = c_0$  of downstream congestion fronts (while formula (6) allows for different speeds  $C$ ), the density  $\rho_{\text{jam}}$  inside of jams, and the outflow  $Q_{\text{out}}$  from jams. Kerner and Rehborn (1996a) have given some empirical support of these features. It is claimed that the propagation speed of wide jams would not even be influenced by passing congested flow (Kerner 1998b), ramps, or intersections (Kerner and Rehborn 1996a; Kerner 2000a, b).

Recently, Nagel and Nelson (2005) summarized the criticisms of the LWR model, stating that

1. it would not be able to describe unstable flow,

2. it would not describe spontaneous breakdowns of traffic flow, and
3. it would not explain the two-capacity phenomenon ("capacity drop").

While Kerner and Rehborn (1996b, 1998) find quite remarkable capacity drops by about one-third of the maximum measured flow values (one-minute aggregates), other authors question the measurement method (Daganzo, Cassidy, and Bertini 1999; Persaud and Hurdle 1991; Banks 1991a, b) or find considerably smaller capacity drops in data of U.S. freeways (Persaud, Yagar, and Brownlee 1998; Westland 1998; Cassidy and Bertini 1999; Banks 1991b). Therefore, the capacity drop seems not to be a finally accepted fact, although the concept is very old (Edie and Foote 1960). From our point of view, the capacity drop is not the most important issue. It depends on what reduction in the time gaps drivers find acceptable under free (i.e., easily predictable) traffic conditions compared with the reaction time to unexpected events (Treiber, Kesting, and Helbing 2006).

Concerning spontaneous breakdowns, some evaluations of vehicle trajectories based on aerial photographs suggest the existence of *phantom traffic jams* (Treiterer and Myers 1974; Treiterer and Taylor 1966), i.e., the spontaneous formation of traffic jams with no obvious reason, such as an accident or a bottleneck. This interpretation has been questioned by Daganzo (2002a), who states that the breakdown of free traffic "can be traced back to a lane change in front of a highly compressed set of cars" and suggests that there actually is a reason for jam formation. However, its origin can be a rather small disturbance. Muñoz and Daganzo (2002a, p. 134) also proposes that the "spontaneous" formation of a traffic jam reported by Kerner and Rehborn (1997) may actually have been caused by disturbances by vehicles trying to leave the freeway over a nearby off-ramp.

The existence of unstable flow is a controversial issue as well (see next paragraph). However, it seems to be commonly accepted that the LWR model, without particular assumptions or extensions, cannot describe growing amplitudes in the density, speed, or flow profiles, the *spontaneous emergence* of stop-and-go waves, or other kinds of oscillating traffic patterns.

## 2.3. Stop-and-Go Waves (Also: Start-Stop Waves)

Stop-and-go waves have been empirically studied by many authors, including Edie and Baverez (1965), Mika, Kreer, and Yuan (1969), and Koshi, Iwasaki, and Ohkura (1983). The latter have found that the parts of the velocity profile that belong to the fluent stages of stop-and-go waves do not significantly depend on the flow (regarding their height and length), while their oscillation frequency does. Correspondingly, there is no *characteristic* frequency of stop-and-go traffic. The

average duration of one wave period is normally between 4 and 20 minutes for wide traffic jams; see, e.g., Mika, Kreer, and Yuan (1969), Kühne (1987), Helbing (1997d), Helbing (1997a), Helbing (1997c); and the average wave length between 2.5 and 5 km, see, e.g., Kerner (1998a). Stop-and-go waves propagate against the direction of the vehicle flow (Edie and Baverz 1965; Mika, Kreer, and Yuan 1969) with a velocity  $c = c_0 := -15 \pm 5$  km/h (see, e.g., Mika, Kreer, and Yuan 1969; Kerner and Rehborn 1996a; Kerner 2000a, b; Cassidy and Mauch 2001), and without spreading (Cassidy and Windover 1995; Windover 1998; Muñoz and Daganzo 1999).

#### 2.4. Possible Origins of Oscillations

The existence of stop-and-go waves seems to be an accepted fact in the meantime. For example, Daganzo (2002a) notes in one of his studies that “large oscillations in flow, speed and cumulative count increase in amplitude across the detectors spanning a long freeway queue and its intervening on-ramps” (p. 133). Moreover, Cassidy and Bertini (1999) state that “the discharge flows in active bottlenecks exhibit near-stationary patterns that (slowly) alternate about a constant rate [...] The onset of upstream queueing was always accompanied by an especially low discharge flow followed by a recovery rate and these are the effects of driver behavior we do not yet understand” (p. 40).

Even today, there is still no agreement about the mechanism generating stop-and-go waves. Laval and Daganzo (2006) suggests explaining emergent oscillations by lane-changing maneuvers, particularly at locations close to on-ramps or off-ramps. In fact, it has recently been shown that coordination problems at bottleneck areas can trigger intermittent outflows, which can be explained with a slightly modified continuity equation, i.e., a first-order model (Helbing et al. 2006). However, how should small oscillations grow larger until they finally form wide-moving jams or stop-and-go waves? Daganzo (2002b) and Mauch and Cassidy (2002) suggest that small oscillations may be increased in amplitude due to a “pumping effect” at ramps, i.e., a positive feedback mechanism.

#### 2.5. Growing Perturbations and the Boomerang Effect

The open question is: Do oscillations grow only when ramps or changes in the number of lanes can support a pumping effect? If not, this would seriously question the LWR model, while it would support models predicting unstable traffic flows. While the experiments by Sugiyama et al. (2005) studying circular traffic flows at high vehicle densities support the idea of unstable traffic flow, it must be considered that drivers may have unintentionally been influenced by the experimental setup.

Therefore, let us discuss some of our empirical results to motivate the following sections of our paper. A Fourier analysis (not shown here) has been carried out to make sure that the oscillations on the investigated German freeway A5 were not induced by pulsating inflows, because they could be produced by traffic lights on the way to the on-ramp. Nevertheless, we found several examples of the growth of downstream moving, small perturbations along a three-lane freeway section without on- and off-ramps; and of the growth of upstream moving, large perturbations (see Figures 1 and 2). These examples show that neither the proposed pumping effect nor the LWR equation are satisfactory explanations.

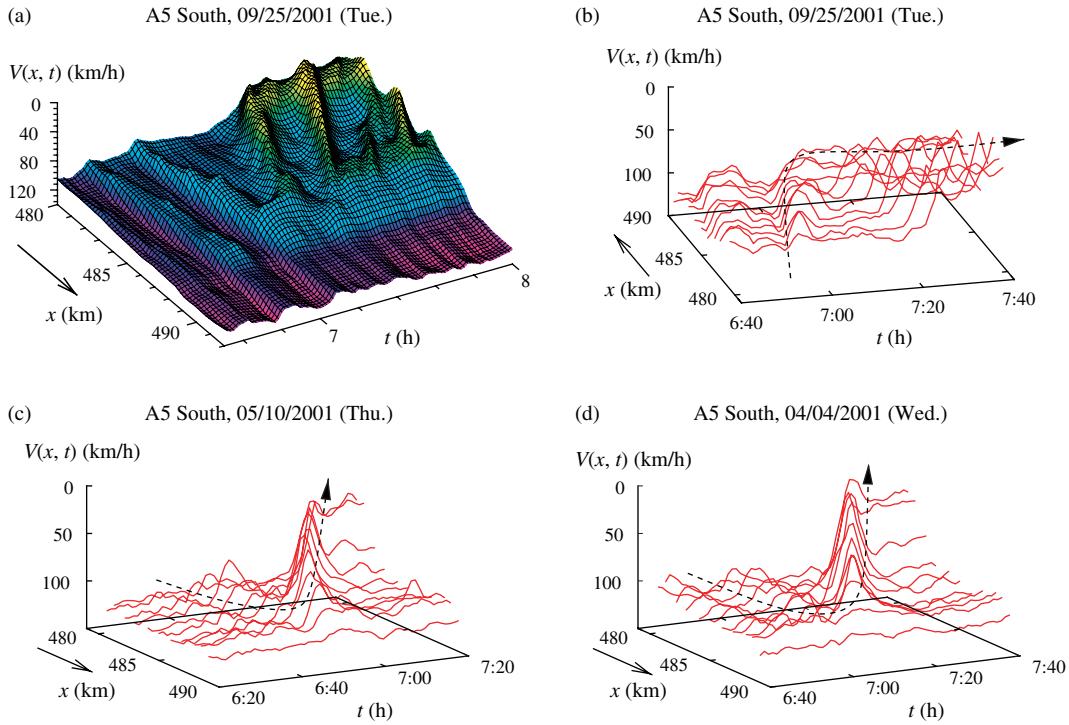
We call the mechanism of generating large moving clusters shown in Figures 1 and 2 the boomerang effect (Helbing 2001a; Helbing et al. 2003). The boomerang effect clearly supports traffic models with an unstable density range. It can be seen in traffic simulations with second-order models Kerner and Konhäuser (1993) and explained as follows: On the one hand, a small perturbation corresponds to a cluster of vehicles, which moves forward in downstream direction together with the vehicles. On the other hand, a large perturbation like a moving localized traffic jam propagates upstream. While inside of the traffic jam vehicles are standing, vehicles are leaving it at its downstream front, and new vehicles are joining the jam at its upstream front. Altogether, this implies an upstream motion of the jam. When a small perturbation grows in size due to unstable traffic flow, the propagation direction is eventually turning around, which is reflected by the term “boomerang effect.”

Although a breakdown of traffic flow is not always related to the boomerang effect, it occurs rather frequently (see also Figures 9b, 13b, 20b, and 21a). At least 18 out of 245 congestion patterns were triggered in this way. However, the effect is hard to recognize without a suitable method of data preprocessing such as the adaptive smoothing method (see §5.3). Nevertheless, it is unlikely that our observation of the boomerang effect is an artifact of our method of data representation: It can also sometimes be seen in classical contour plots without particular preprocessing (Helbing et al. 2003). The same applies to some three-dimensional plots without any spatial smoothing, when the data are suitably smoothed in time (see Figure 1), or to the raw time series (see Figure 2).

### 3. Second-Order Models and Their Criticism

#### 3.1. Mathematical Form of Second-Order Models

Having presented empirical results questioning the LWR model, let us now discuss second-order traffic



**Figure 1 Illustration of Growing Perturbations on a Homogeneous Freeway Section Without On- and Off-Ramps Between Kilometers 481.3 and 488 (See §5.3 for Details of the Data Evaluation and Visualization)**

*Notes.* According to the LWR model, the amplitude of the spatiotemporal velocity profile should not grow along a freeway section without ramps. However, in (a), we see a small perturbation propagating downstream and growing slowly. After reaching a sufficiently large size and crossing a bottleneck, traffic breaks down, the perturbation changes its direction (“boomerang effect”) and travels upstreams with a significant growth in amplitude. In order to exclude that the growth could be an interpolation effect due to the application of the adaptive smoothing method (see §5.3), (b) shows the same data set, but without any spatial smoothing. In contrast to (a), the time-dependent velocity data at the different measurement cross-sections are shown from behind. (c), (d) Examples for the growth of medium-sized perturbations propagating upstream in the same representation as (b), but shown from the front (looking against the driving direction). Arrows are guides for the eyes, indicating the growth of the velocity amplitudes.

models as a possible alternative. In contrast to Equation (2), which implies an immediate adaptation of the vehicle velocity  $V(\rho(x, t)) = Q(\rho(x, t))/\rho(x, t)$  to a changing density  $\rho(x, t)$ , most second-order models assume the general fluid-dynamic relationship  $Q(x, t) = \rho(x, t)V(x, t)$  with an additional partial differential equation for the spatiotemporal change of the average velocity  $V(x, t)$  of vehicles. This can often be written in the form

$$\frac{\partial V}{\partial t} + \underbrace{V \frac{\partial V}{\partial x}}_{\text{transport term}} = -\underbrace{\frac{1}{\rho} \frac{\partial P}{\partial x}}_{\text{pressure term}} + \underbrace{\nu(\rho) \frac{\partial^2 V}{\partial x^2}}_{\text{viscosity term}} + \underbrace{\frac{1}{\tau} [V_e(\rho) - V]}_{\text{relaxation term}}. \quad (7)$$

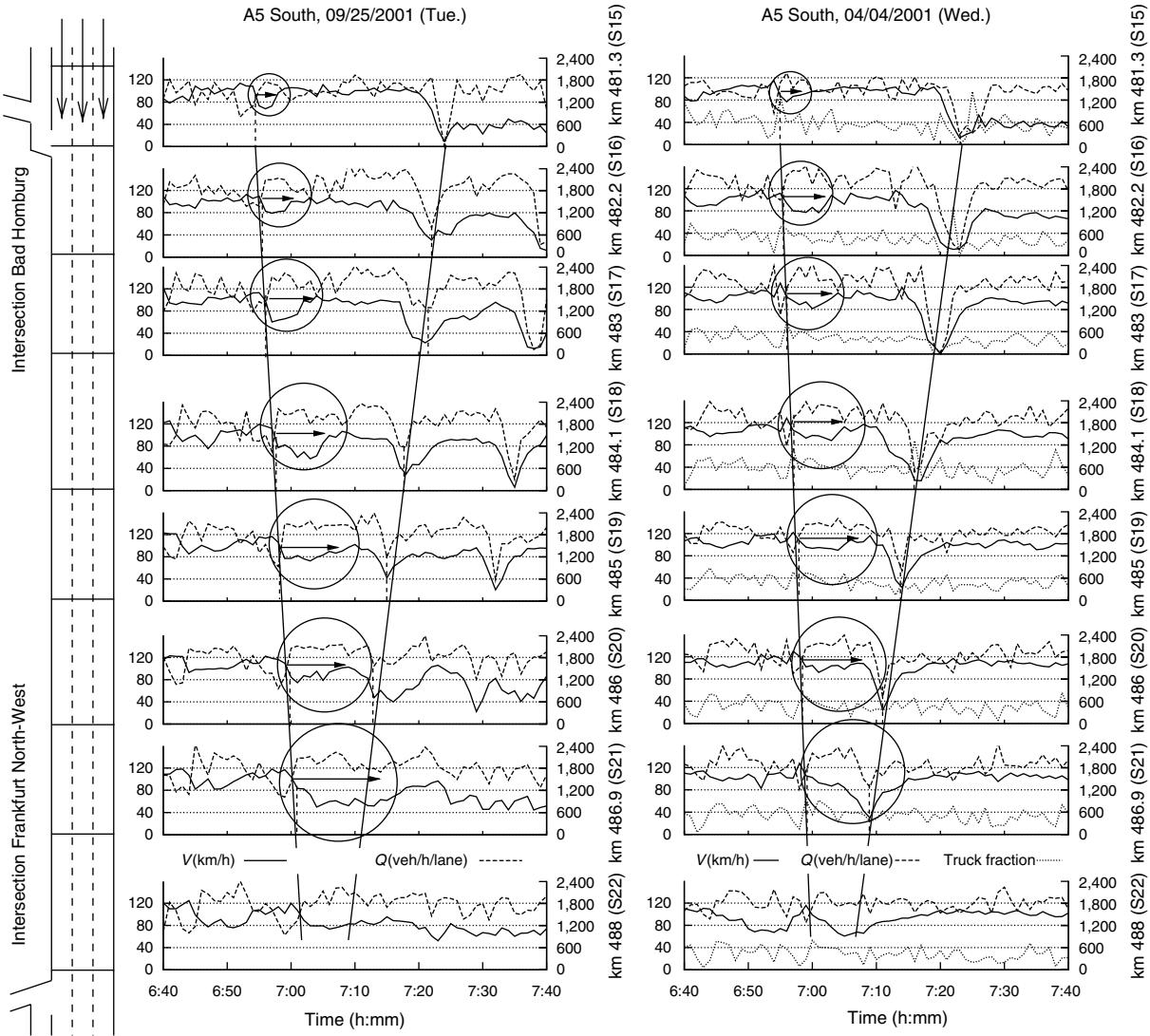
This equation was inspired by the Navier-Stokes equations for viscous fluids. Together with Equation (1), we may call this equation the Payne-Whitham-Kühne-Kerner-Konhäuser-Lee model (Payne 1971, 1979; Whitham 1974; Kühne 1984; Kerner and

Konhäuser 1993; Lee, Lee, and Kim 1999). These equations predict linear instability under the condition

$$\rho \left| \frac{dV_e}{d\rho} \right| > \sqrt{\frac{dP}{d\rho}} [1 + \tau \nu(\rho) k^2], \quad (8)$$

where the wave number  $k$  is inversely proportional to the wave length (Helbing 1997d). The variable  $P$  is called the traffic pressure,  $\tau$  the relaxation or adaptation time, and  $V_e$  the density-dependent “optimal” velocity. Furthermore,  $\nu$  was chosen constant or set to  $\nu(\rho) = \eta/\rho$ , where  $\eta \geq 0$  denotes a viscosity coefficient.

Notice that the Lighthill-Whitham model results in the limit  $\tau \rightarrow 0$  of instantaneous velocity adjustments. However, Whitham himself proposed a second-order extension of the LWR model in 1974. The models by Payne (1971) and Papageorgiou (1983) are obtained for  $P(\rho) = [V_0 - V_e(\rho)]/(2\tau)$ , with the “free” or “desired” average velocity  $V_0 = V_e(0)$ . For  $dP/d\rho = -\rho/[2\tau(\rho + \kappa)]dV_e/d\rho$ , one ends up with Cremer’s model (Cremer 1979). In the model of Phillips (1979b, a), there is  $\nu = 0$  and  $P = \rho\theta$ , where  $\theta$  denotes the velocity variance. The models of Kühne (1984,



**Figure 2 Further Analysis of the Boomerang Effect Depicted in Figure 1 (For Details of the German Autobahn A5, See §5.2)**

**Notes.** According to the LWR model, the amplitude of the spatiotemporal velocity profile should not grow along a freeway section without ramps. However, in both data sets, we see a small perturbation, which eventually increases: First, while propagating downstream, the *extension* of the localized perturbation grows, i.e., the time intervals, in which flows are high and velocities are reduced (cf. the arrows and circles). This is because more and more vehicles join the slower vehicle cluster from behind, which may be caused by a moving bottleneck due to overtaking trucks, as suggested by the peaks in the truck fraction at the beginning of the vehicle cluster (right). From one cross-section to the next downstream one, traffic needs more time to reach a “normal” state again. Finally, approximately one kilometer before the off-ramp at “Intersection Frankfurt North-West,” which induces many lane-changing maneuvers because of the high off-ramp flow (left: 1,980 veh/h/lane between 7:05 A.M. and 7:15 A.M.; right: 1,767 veh/h/lane between 7:05 A.M. and 7:15 A.M.), the velocity drops to values around 60 km/h, and a traffic jam travelling against the driving direction appears. Afterward, the reduction in the average speed and flow increases with time, thereby causing a growth in the amplitude of the perturbation. To calculate the velocities of the stop waves, we determined the location-specific time when the velocity exceeded the value 60 km/h after the stop wave had passed. Applying linear regression, these times, together with the detector locations, implied the mean velocity of the stop wave. On 09/25/01 (left) it was  $-16.8 \pm 1.4$  km/h, whereas on 04/04/01 (right) it was  $-18.0 \pm 0.8$  km/h. The velocities of the downstream propagating perturbations were  $(66 \pm 6)$  km/h (left) and  $(76 \pm 6)$  km/h (right). To obtain these velocities, we used the times when the mean velocity at the cross-sections fell below 110 km/h.

1987) and Kerner and Konhäuser (1993) result for  $P = \rho\theta_0$ , where  $\theta_0$  is a positive constant. In comparison with a similar model by Whitham (1974), they assume a finite viscosity to smooth out shock fronts. This is desirable from an empirical and numerical point of view, but implies theoretical inconsistencies (see the next paragraph).

### 3.2. How to Survive the Requiem for Second-Order Traffic Models

Second-order models have been seriously criticized by Daganzo (1995b):

1. The second-order models would not be well derived from microscopic models, and there were problems with terminating series expansions.

2. The gas-kinetic derivations would imply that the speed was a property of the street rather than the drivers, and different personalities of drivers, e.g., aggressive and timid ones, could not be distinguished.

3. The applied numerical solution schemes were often improper, with many “engineering fixes” and too many parameters to fit to empirical data. As a consequence, these models would have low predictive power.

4. There were inconsistencies regarding the vehicle velocities. For example, in the case of a discontinuous density change at the end of a traffic jam, vehicles should be moving backwards due to the symmetric viscosity term.

5. It would be a problem that one of the characteristic speeds of the second-order model (describing the propagation of perturbations) was faster than  $V$ , the speed of vehicles.

At the time of publication of the “requiem,” most of these criticisms were justified. The many different specifications of the pressure and relaxation term in the literature indicate that there was considerable confusion about the meaning and derivation of these terms. In physics, the velocity equation can be derived from a gas-kinetic description. Such an approach has been followed in traffic modeling as well, starting with Prigogine (1961), Paveri-Fontana (1975), and Phillips (1977). Based on today’s knowledge, the conclusions from successful efforts to link microscopic and macroscopic traffic models (Helbing et al. 2002) are as follows:

1. The viscosity term cannot be justified for one-dimensional flows as it is in physics. It either drops out completely or results from a gradient expansion, but then it looks differently. The pressure has the form  $P = \rho\theta$ , where  $\theta$  represents the velocity variance of vehicles. Of course, the variance must vanish whenever vehicles are stopped, i.e.,  $\theta = 0$  if  $V = 0$ . The effect of the pressure term can be imagined as follows: Starting with a cluster of vehicles on a multilane freeway, the cluster will eventually widen due to the faster vehicles leaving the slower ones behind. Note that problems of series expansion can be avoided by direct derivation of macroscopic models from microscopic ones (Helbing et al. 2002, pp. 543 ff.).

2. Gas-kinetic models can be used to derive multiclass multilane models (Helbing 1997d, b; Klar and Wegener 1999a, b; Hoogendoorn and Bovy 1998) in order to model passing interactions and different behaviors of driver-vehicle units. These allow us to distinguish different personalities of drivers (e.g., aggressive and timid ones) as demanded. These multiclass models also fulfill the requirement that the vehicle velocities should not be a property of the road, but of the drivers. The same already applies to some earlier models (Paveri-Fontana 1975; Helbing 1996; Helbing and Greiner 1997; Wagner et al. 1996).

3. Today, proper numerical integration schemes are known (Kronjäger and Konhäuser 1997; Helbing and Treiber 1999), and high values of anticipation coefficients can be avoided by nonlocal interaction terms (Helbing and Treiber 1999).

4. The phenomenological viscosity term must be dropped, but the nonlocal interaction terms result in reasonable, anisotropic smoothing effects (Helbing and Treiber 1999).

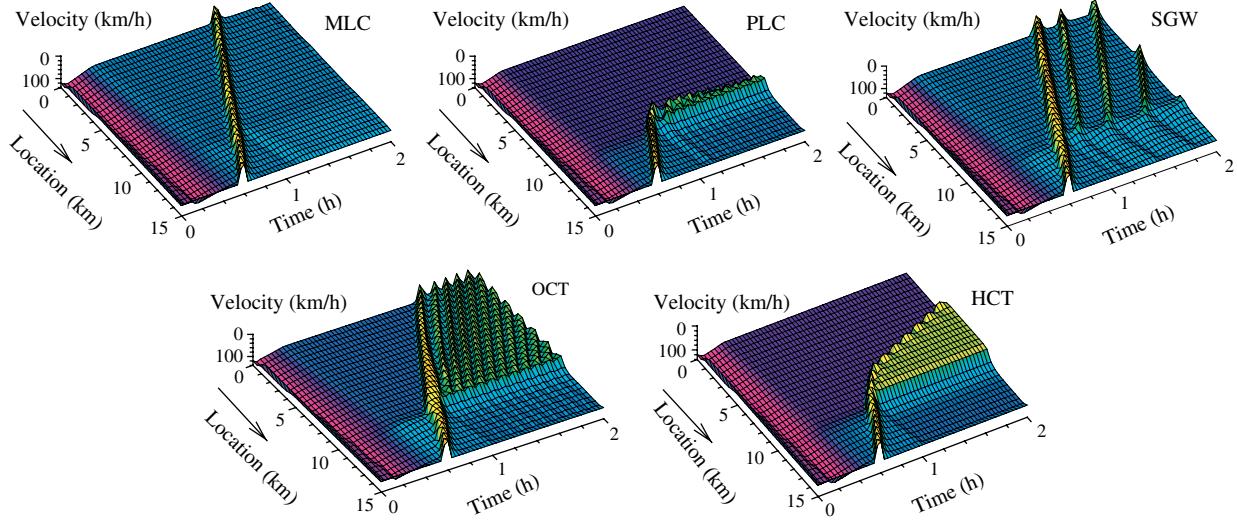
5. The faster characteristic speed is actually *not* a problem if  $\theta$  is correctly interpreted as velocity variance. Then, the two characteristic speeds are  $c_{+-} \approx V \pm \sqrt{\theta}$ , i.e., the faster characteristic speed is just one standard deviation faster than the average velocity. Clearly, information can be propagated by the faster vehicles at this speed if we assume a finite velocity variance. Otherwise, if  $\theta = 0$ , the pressure term disappears, implying  $c_+ = c_- = V$ . Moreover, perturbations moving with the faster speed  $c_+$  decay very quickly (Whitham 1974, pp. 339 ff.; Yi et al. 2003). Therefore, they are not expected to occur in normal traffic situations.

#### 4. The Nonlocal, Gas-Kinetic-Based Traffic Model (GKT Model)

To overcome Daganzo’s criticisms, it is necessary to derive a macroscopic traffic model from an empirically supported microscopic traffic model (not vice versa). The main problem to be avoided is the possible occurrence of negative speeds, not the existence of a characteristic speed  $c_+$  faster than the average velocity  $V$  (as long as it does not exceed the speed of the *fastest* car). However, negative speeds can be easily prevented by interpreting  $\theta$  as the velocity variance of cars, which must vanish when vehicles are stopped. This constraint can, for example, be fulfilled by setting  $\theta(\rho, V) = \alpha(\rho)V^2$ , with an empirically determined function  $\alpha(\rho) > 0$  (Helbing 1997d, 1998; Treiber, Hennecke, and Helbing 1999).

One critical question, however, remains: As the pressure term  $-(1/\rho)\partial(\rho\theta)/\partial\rho$  can become positive, would that mean that cars would potentially accelerate into high-density areas? This is *not* the case, if the relaxation term  $(V_e - V)/\tau$  compensates for negative  $\partial(\rho\theta)/\partial\rho$ , as in the nonlocal, gas-kinetic-based traffic (GKT) model (Helbing 1997d, 1998). The non-locality of this model reflects the interaction with upstream vehicles in a certain distance and causes a forward smoothing with no reaction to backward conditions, in contrast to a viscosity term (Helbing and Treiber 1999; Helbing et al. 2002).<sup>2</sup> In summary,

<sup>2</sup>This also implies a higher numerical efficiency despite its more complicated mathematical form. Naturally, the nonlocal GKT model takes into account effects of space requirements by vehicle



**Figure 3** Simulation of Traffic on a Freeway with an On-Ramp at Location  $x = 10$  km, Assuming an Initial Disturbance Travelling Upstream and Using the Nonlocal, Gas-Kinetic-Based Traffic (GKT) Model

*Notes.* Depending on the respective traffic flows on the ramp and on the freeway, five different kinds of congested traffic states emerge: a moving localized cluster (MLC), a pinned localized cluster (PLC), stop-and-go waves (SGW), oscillating congested traffic (OCT), or homogeneous congested traffic (HCT). During the first minutes of the simulation, the flows on the freeway and the on-ramp were increased from low values to their final values.

the GKT model is theoretically consistent and numerically efficient (Helbing and Treiber 1999). But is it also realistic? In order to compare it with empirical facts (see §6), we will now discuss its implications regarding the predicted congestion states and their respective preconditions.

#### 4.1. Congested Traffic States

For our classification of the traffic states predicted by the nonlocal GKT model, it is not necessary to present its mathematical form here (for details, see Treiber, Hennecke, and Helbing 1999; Helbing 2001b; Helbing et al. 2001, 2002). In this model, congested traffic is typically triggered by bottlenecks (i.e., spatial inhomogeneities along the freeway). As an example, we will discuss the situation of an on-ramp entering an  $n$ -lane freeway at location  $x = 5$  kilometer. If the ramp flows  $Q_{\text{ramp}}$  are equally distributed over a ramp length  $L$ , we have an inflow of

$$\nu_+ = \frac{Q_{\text{ramp}}}{nL}, \quad (9)$$

which must be inserted into Equation (1). This inflow consumes some of the capacity of the freeway and causes a bottleneck strength, the size of which is given by the entering ramp flow, divided by the number of freeway lanes:

$$\Delta Q = \frac{Q_{\text{ramp}}}{n}. \quad (10)$$

lengths and safe time clearances. Correlations between the velocities of successive cars can be easily treated as well (Shvetsov and Helbing 1999).

When systematically varying the ramp flow  $Q_{\text{ramp}}$  and the freeway flow  $Q_{\text{up}}$  upstream of the bottleneck, we find free flow or different kinds of congested flow (Helbing, Hennecke, and Treiber 1999). Typical representatives are shown in Figure 3.

In contrast to Kerner's classification, we do not distinguish the different traffic states based on measurements at discrete cross-sections of the freeway. Instead, we study the average vehicle speed  $V_t(x)$  along a freeway section as a function of time or the speeds  $v(x(t))$  along trajectories of representative vehicles given by the relationship  $dx(t)/dt = v(x(t))$ . These quantities are determined based on the three-dimensional representation of the average velocity  $V(x, t)$  as a function of location  $x$  and time  $t$ , which is obtained from cross-sectional measurements by means of an adaptive smoothing method (see §5.3).

If  $V_t(x)$  or  $v(x(t))$  are above a certain threshold  $V_{\text{tr}}$ , where  $x$  is varied within a homogeneous freeway section upstream of a bottleneck (an inhomogeneity), we call the traffic state *free traffic* (FT)—otherwise, we have congested traffic. If these speeds fall below  $V_{\text{tr}}$  only over a short freeway subsection, and the length of this section is approximately stable or stabilizes over time, we talk about *localized clusters* (LC), otherwise, of *spatially extended congestion states*. According to simulations, there are two forms of localized clusters: *Pinned localized clusters* (PLC) stay at a fixed location over a longer period of time, while *moving localized clusters* (MLC), corresponding to Kerner's wide-moving jams (Kerner and Rehborn 1996a), propagate upstream with the characteristic speed  $c_0$ . *Stop-and-go waves* (SGW) may be interpreted as a sequence of

several moving localized clusters. Alternatively, they may be viewed as special case of *oscillating congested traffic* (OCT), but with free traffic flow of order  $Q_{\text{out}}$  between the upstream propagating jams. Generally, however, OCT states are just characterized by oscillating speeds in the congested range. If the speeds are congested over a spatially extended area, but not oscillating, we call this *homogeneous congested traffic* (HCT). This is typically related to low velocities.

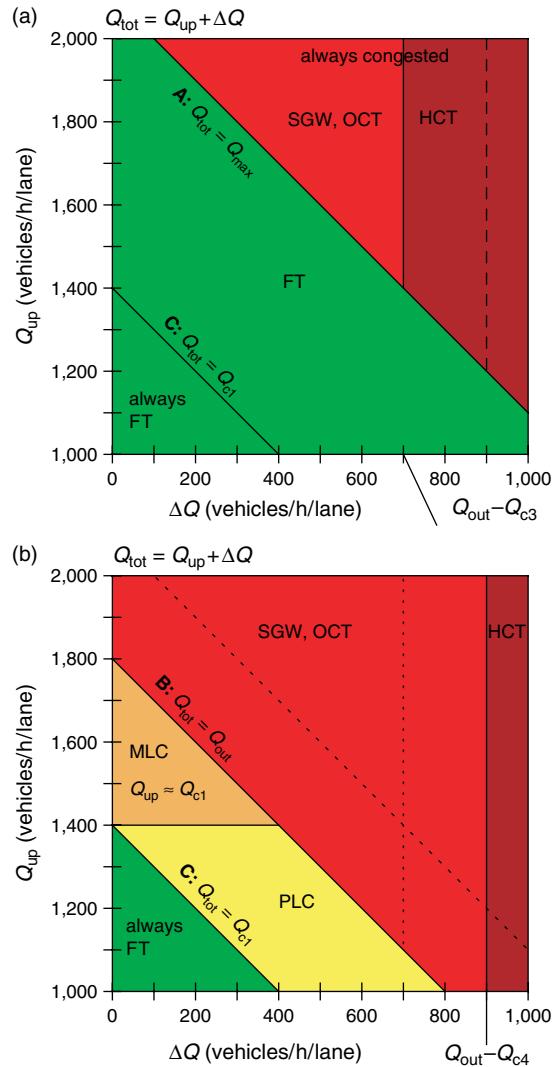
In summary, besides free traffic, the GKT model predicts five different spatiotemporal patterns of congested traffic states at a simple on-ramp bottleneck: PLC, MLC, SGW, OCT, and HCT. Similar traffic states have been identified for second-order models such as the Kerner-Konhäuser model (Lee, Lee, and Kim 1999) and for flow-conserving bottlenecks in a car-following model (Treiber, Hennecke, and Helbing 2000; Helbing et al. 2002). Note that in empirical data, the SGW state may be hard to distinguish from OCT states due to their similar appearance. Moreover, models assuming adaptive driver behavior or heterogeneous traffic (i.e., different driver-vehicle classes) may predict variants of the states depicted in Figure 3. However, the traffic patterns are expected to be similar for all models with the same stability diagram (see next section). Finally, in cases of several bottlenecks along the freeway, it should be possible that several different congestion patterns are combined with each other, giving rise to a large number of possible congestion patterns with a complex appearance (see §§6.7 and 6.9).

#### 4.2. Phase Diagram of Traffic States

The preconditions for the possible occurrence of the different traffic states can be illustrated by a phase diagram (see Figure 4). Each area of the phase diagram represents the parameter combinations for which certain kinds of traffic states can exist. The borderlines between different areas (the so-called phase boundaries) can be theoretically understood, based on the instability diagram and the dynamic capacity  $Q_{\text{out}}$ , i.e., the characteristic outflow from congested traffic (Kerner and Rehborn 1996a). As a consequence, second-order models with a similar instability diagram are expected to imply similar traffic states and phase diagrams as does the GKT model (see, e.g., Lee, Lee, and Kim 1999).

For the theoretical derivation of the phase boundaries, one must know that the nonlocal GKT model (like the Kerner-Konhäuser model from 1993) predicts stable traffic flow, when the velocity changes little with the density. More specifically, there is stable traffic below some critical density  $\rho_{c1}$  and above some critical density  $\rho_{c4}$  (see Figure 5).

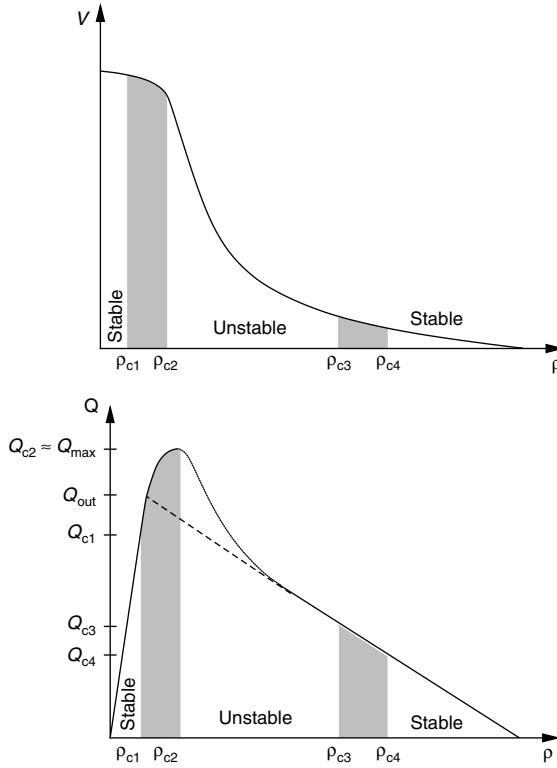
For medium traffic densities between two critical densities  $\rho_{c2}$  and  $\rho_{c3}$ , traffic flow is linearly unstable,



**Figure 4** Schematic Representation of Phase Diagrams of Traffic States Assuming (a) Negligible and (b) Large Perturbations of the Traffic Flow

*Notes.* The different areas indicate for which combinations of the upstream freeway flow  $Q_{\text{up}}$  and the bottleneck strength  $\Delta Q$  certain traffic states (see Figure 3) are predicted to exist. Superimposing both diagrams shows that many areas are multistable (see Figure 15a). That is, depending on the initial and boundary conditions, one may find one out of several possible states. Note that the value of  $Q_{\text{max}}$  depends on the speed limit  $V_0$ . SGW states are a special case of OCT states, in which the traffic between the congested areas flows freely. The existence of PLC and MLC states for a certain range of flows indicates that their actual outflow can assume a range of different values between  $Q_{c1}$  and  $Q_{c4}$ .

i.e., even small perturbations can grow and cause a breakdown of traffic flow. In the intermediate density ranges  $\rho_{c1} \leq \rho < \rho_{c2}$  and  $\rho_{c3} < \rho \leq \rho_{c4}$ , one finds metastable traffic, i.e., sufficiently small perturbations will fade away, while large enough ones will grow and cause a breakdown of traffic flow. The value of  $Q_{\text{out}}$  falls into the metastable regime between  $\rho_{c1}$  and  $\rho_{c2}$ . It depends on the maximum density and the aver-



**Figure 5 Schematic Illustration of Velocity  $V$  and Flow  $Q$  as a Function of the Vehicle Density  $\rho$**

*Notes.* Grey regions indicate density ranges of metastable traffic flow (cf. text), whereas  $Q_{ck} = Q(\rho_{ck})$  denotes the flows belonging to the instability thresholds  $\rho_{ck}$ .

age time gap, i.e., on the weather conditions, the truck fraction, and lane.<sup>3</sup>

With this background information, we can now formulate conditions for the existence of the different traffic states. Doing so, it is important not to restrict ourselves to the situations with large initial perturbations, which have been studied before (Helbing, Hennecke, and Treiber 1999). This will highlight the multistable nature of traffic flows. That is, even for the same average ramp and freeway flows, different traffic states can ultimately result if the initial and boundary conditions—e.g., the perturbation sizes—differ. This will be important for the comparison with empirical data in §6. Figure 4 displays the phase diagram resulting for small perturbations separately from the one for large perturbations in order to resolve a misunderstanding by Kerner, according to which the GKT model would predict a transition from free flow to “synchronized” congested flows only *after* the occurrence of moving jams (Kerner 2005). For

<sup>3</sup> According to Kerner and Rehborn (1996a), the typical outflow  $Q_{out}$  from congested traffic on the German autobahn A5 near Frankfurt is 1,100 veh/h for the right lane, 1,600 veh/h for the middle lane, and 1,800 veh/h for the left lane if the vehicles inside the congestion (nearly) get to a standstill.

small perturbations, this is actually not the case (see Figure 4a).

Remembering that the bottleneck strength  $\Delta Q = Q_{rmp}/(nL)$  relates to the effective ramp flow and  $Q_{up}$  denotes the upstream traffic flow on the freeway, the total freeway capacity required downstream of the ramp amounts to

$$Q_{tot} = Q_{up} + \Delta Q. \quad (11)$$

Therefore, we *always* expect to find spatially extended, growing congestion states, if the maximum freeway flow  $Q_{max} = \max_{\rho} Q(\rho)$  is exceeded, i.e.,

$$Q_{tot} = Q_{up} + \Delta Q > Q_{max}, \quad (12)$$

which corresponds to the region above line A in Figure 4.<sup>4</sup> Free traffic flow *can* in principle exist under the condition  $Q_{up} + \Delta Q < Q_{max}$ , the region below line A in Figure 4a. This region can be further subdivided: While above the threshold  $Q(\rho_{c1})$  of metastable traffic—i.e., for

$$Q_{tot} = Q_{up} + \Delta Q > Q(\rho_{c1}), \quad (13)$$

corresponding to the region above line C in Figure 4, large enough perturbations can also trigger congested traffic states, no congested state can occur for

$$Q_{tot} = Q_{up} + \Delta Q < Q(\rho_{c1}). \quad (14)$$

Under this condition, i.e., below line C, one should *always* find free flow.

Assuming model parameters for which the maximum flow  $Q_{max}$  is realized for a density approximately equal to  $\rho_{c2}$ , traffic flows between the two diagonal lines A and C, i.e.,  $Q_{up} + \Delta Q = Q(\rho_{c1})$  and  $Q_{up} + \Delta Q = Q_{max}$  in the  $\Delta Q$ -over- $Q_{up}$  phase diagram can be either congested or free, depending on the initial and boundary conditions.<sup>5</sup> Particularly, while free flow may persist over long time periods, large perturbations may trigger congested traffic states.

Spatially extended congestion can only emerge for

$$Q_{up} > Q_{out} - \Delta Q, \quad (15)$$

i.e., if the required capacity  $Q_{tot} = Q_{up} + \Delta Q$  is greater than the available dynamic capacity in the case of congested traffic, which is given the outflow from congested traffic  $Q_{out}$ . This line B in Figure 4, in principle, does not have to be parallel to lines A and C, because  $Q_{out}$  may depend on the bottleneck

<sup>4</sup> Note that  $Q_{max}$  is a function of the speed limit  $V_0$ .

<sup>5</sup> If  $Q_{max}$  significantly differs from  $Q_{c2}$ , the situation becomes more complicated.

strength  $\Delta Q$  (Treiber, Hennecke, and Helbing 2000; Helbing 2001b). For

$$Q(\rho_{c1}) \leq Q_{\text{tot}} = Q_{\text{up}} + \Delta Q < Q_{\text{out}}, \quad (16)$$

i.e., in the region below line B, congested traffic states cannot grow because the freeway capacity for congested traffic is higher than the totally required capacity  $Q_{\text{tot}}$ . This implies that any resulting congestion state should be a localized cluster, i.e., constraint to a short freeway section. The existence of *moving localized clusters* (MLC) requires that the traffic flow in the freeway section upstream of the ramp is metastable, i.e.,

$$Q_{\text{up}} \geq Q(\rho_{c1}). \quad (17)$$

Otherwise, if  $Q_{\text{up}} < Q(\rho_{c1})$ , no MLC can survive, and only free flow can exist in the upstream area. Rather than propagating through this area, the localized congestion must stay at the location of the ramp, which results in a *pinned localized cluster* (PLC).<sup>6</sup>

Let us now discuss the different possible forms of spatially extended congested traffic: The traffic flow  $Q_{\text{cong}}$  resulting in the congested area is given by the outflow  $Q_{\text{out}}$  from congested traffic minus the inflow (or bottleneck strength)  $\Delta Q$ , i.e.,

$$Q_{\text{cong}} = Q_{\text{out}} - \Delta Q. \quad (18)$$

One can distinguish the following cases: *Homogeneous congested traffic* (HCT) will always occur if the density  $\rho_{\text{cong}}$  associated with the congested flow  $Q_{\text{cong}} = Q(\rho_{\text{cong}})$  lies in the stable range:

$$Q_{\text{cong}} \leq Q(\rho_{c4}), \quad \text{i.e.,} \quad \Delta Q \geq Q_{\text{out}} - Q(\rho_{c4}). \quad (19)$$

Moreover, it can occur in the metastable range

$$Q_{\text{cong}} \leq Q(\rho_{c3}), \quad \text{i.e.,} \quad \Delta Q \geq Q_{\text{out}} - Q(\rho_{c3}), \quad (20)$$

if perturbations in the traffic flow are small enough.

If the perturbations are supercritical, however, extended *oscillating congested traffic* (OCT) can emerge whenever

$$\Delta Q \leq Q_{\text{out}} - Q(\rho_{c4}) \quad \text{and} \quad Q_{\text{up}} > Q_{\text{out}} - \Delta Q \quad (21)$$

(see Figure 4b). In the linearly unstable range

$$\Delta Q \leq Q_{\text{out}} - Q(\rho_{c3}) \quad \text{and} \quad Q_{\text{up}} > Q_{\text{out}} - \Delta Q, \quad (22)$$

<sup>6</sup>Note that pinned localized clusters could also exist under metastable conditions, i.e., for  $Q_{\text{up}} < Q(\rho_{c2})$  and  $Q_{\text{tot}} < Q_{\text{max}}$ , if the speed inside the PLC state was finite and the outflow higher than  $Q_{\text{out}} \approx 1,800$  vehicles per kilometer and lane. For the outflow  $Q_{\text{out}}^{\text{PLC}}(V)$  of a PLC state as a function of the average vehicle speed  $V$ , we conjecture that  $\lim_{V \rightarrow 0} Q_{\text{out}}^{\text{PLC}}(V) = Q_{\text{out}}$  and  $\lim_{V \rightarrow V_0} Q_{\text{out}}^{\text{PLC}}(V) = Q_{\text{max}}$  (cf. Figure 17). An analogous speed-dependence of the outflow  $Q_{\text{out}}$  would apply to other congested traffic states if the vehicles would still move forward with a finite average speed  $V$  inside the congested area.

extended congested traffic will always produce oscillating rather than homogeneous congested flow. *Stop-and-go waves* (SGW) are a special case of OCT. They should primarily occur for not-too-large average densities  $\rho_{\text{cong}}$  in the (partially) congested area—i.e., small enough bottleneck strengths  $\Delta Q$ , because they require free traffic flow between the congested areas.

## 5. Measurement of Congested Traffic States

### 5.1. Choice of the Measurement Site

In the following, we will study data from the German autobahn A5 close to Frankfurt/Main, because this measurement site includes several large road sections without on- and off-ramps, and regularly suffers from severe traffic congestion. Because this measurement site has been the basis of most of Kerner's empirical investigations, it also allows us to gain an independent judgement of the traffic dynamics on this freeway, and to assess the validity of Kerner's far-reaching conclusions. Such an independent study is needed, because Kerner has called all traffic models containing a fundamental diagram into question. We will show that this has been done partly for the wrong reasons (see §7.1).

Some of the issues that Kerner has not addressed in a satisfactory way are the following: What are the implications of freeway properties such as

1. significant gradients along the freeway,
2. the high truck fraction, which varies strongly,
3. the frequent occurrence of accidents or breakdowns of cars,
4. the operation of an “intelligent speed control system” (which adaptively limits the speeds in cases of an expected traffic breakdown to 120 km/h or 100 km/h and in cases of an actual breakdown gradually to 80 km/h or even 60 km/h), and
5. the existence of very long ramps (as they provide an additional lane)?

We will see that neglecting such aspects leads to special rather than general conclusions, and sometimes even to wrong ones:

1. For example, we will find a new mechanism generating stop-and-go waves at gradient-related bottlenecks (see §6.4).
2. Another mechanism producing stop-and-go waves is based on the boomerang effect, which is probably caused by overtaking maneuvers of trucks (see Figures 2, 20b, and 21a).
3. Breakdowns of cars and accidents often cause complex congestion patterns at unexpected locations, including homogeneous congested traffic (see §§6.2 and 6.3).
4. The intelligent speed control system is probably responsible for the homogeneous-in-speed type

of “synchronized” congested flow (Kerner 1998b); see §7.2.

5. Long off-ramps allow one to leave the freeway when congested traffic becomes visible downstream. This adaptive response of drivers to downstream congestion is probably related to the “pinch effect” (Kerner 1998a), see §6.7. Long on-ramps may cause downstream flows that are higher than the outflow  $Q_{\text{out}}$  from congested traffic, because a certain proportion  $\alpha(t)$  of the on-ramp flow may be able to enter the freeway downstream of the congestion front that is caused by the remaining proportion  $1 - \alpha(t)$  of the on-ramp flow.<sup>7</sup>

Moreover, the following questions must be addressed: What implications does it have if empirical data are compared with results of macroscopic simulation data without considering the measurement (data aggregation) method? Would empirical findings such as the scattering of synchronized flows be naturally reproduced by a macroscopic traffic flow model if multiclass and multilane effects are taken into account? We will come back to this in §7.1.

## 5.2. Description of the Measurement Site and Traffic Data

In the following investigation, we will focus on velocity data aggregated over one-minute intervals for a freeway section frequently suffering from serious congestion. Measurements were available for all freeway lanes and most ramps. However, apart from the investigation of exceptional situations such as accidents, we have arithmetically averaged the speeds over all freeway lanes (which is particularly justified for synchronized flows or other congested traffic states). A sketch of the measurement site is presented in Figure 6. It shows an approximately 30-kilometer-long freeway section, which for the most part has three lanes into each direction, three intersections with other freeways, and one junction. At the intersections, there are additional merging and diverging lanes, some of which are more than one kilometer long. Between the intersections “Frankfurt North-West” and “Bad Homburg,” but also between the latter section and the junction “Friedberg,” there are two approximately 10-kilometer-long three-lane freeway sections without disturbances by on- or off-ramps. However, the freeway crosses a valley between kilometers 478 and 480 with gradients of about 2%–3%, which causes an additional flow-conserving bottleneck in the northern direction.

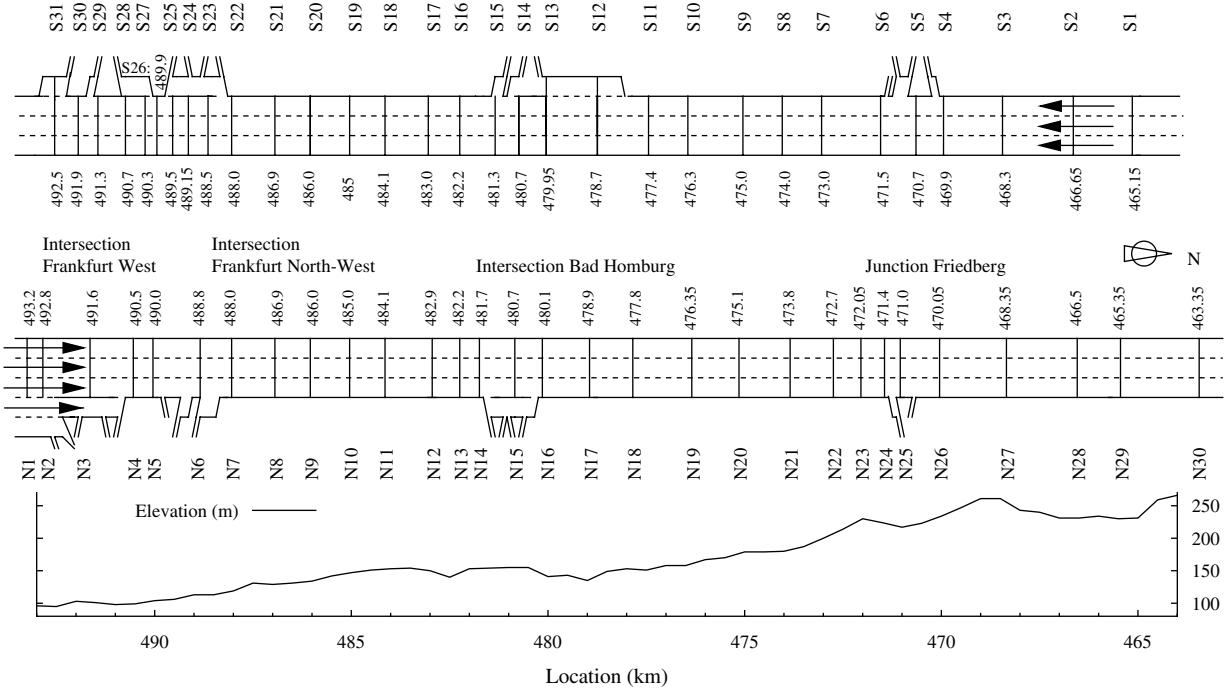
<sup>7</sup> The description of a scenario with such a particular on-ramp design is as follows: If  $\alpha(t)Q_{\text{rmp}}(t)$  with  $0 \leq \alpha(t) \leq 1$  is the fraction of the ramp flow  $Q_{\text{rmp}}(t)$  entering the freeway downstream of the congestion front, the effective bottleneck strength is  $\Delta Q(t) = [1 - \alpha(t)]Q_{\text{rmp}}(t)/n$ , and the flow downstream of the bottleneck is  $Q_{\text{down}}(t) = Q_{\text{out}} + \alpha(t)Q_{\text{rmp}}(t)/n$ .

Moreover, there is a relatively steep hill (“Köpperner Berg”) between kilometers 471 and 472.5, with gradients up to 5%, but its effect on traffic flow in the southern direction is small compared to the slightly upstream located Junction Friedberg (although the average vehicle speed at Köpperner Berg is often as small as 60 km/h, possibly also due to the operation of the intelligent speed control). The main bottlenecks of the studied freeway section in the southern direction are the Junction Friedberg at kilometer 471 and the Intersection Bad Homburg at kilometer 481 (see top of Figure 6), while the main bottlenecks in the northern direction are the Intersection Frankfurt North-West at kilometer 489 and the bottleneck at kilometer 478 due to the uphill gradient (see middle of Figure 6).

We have investigated the available data for both driving directions of the above freeway between 04/01/2001 and 09/30/2001 (165 days), which allows us to identify the typical features of traffic flow and to make statistical analyses. On these days (distinguishing both directions), we identified more than 240 breakdowns of traffic flow at several bottlenecks, mostly during the morning rush hour between about 7 A.M. and 9 A.M. and during the afternoon rush hour between about 3 P.M. and 7 P.M. (on Fridays, between about 1 P.M. and 7 P.M.). Breakdowns were also induced by holiday traffic or accidents. The latter normally occurred outside of the typical bottleneck areas and were protocolled by the road authorities. During the investigated time period, about 500 accidents or breakdowns of vehicles have occurred. However, most of them had only a minor impact on the traffic flow.<sup>8</sup> Finally, note that the average truck fraction during rush hours was between 10% and 15% (with a standard deviation of about 5%), and higher during other time periods (apart from Sundays, which are truck-free until night). The truck fraction may be relevant because trucks are allowed to overtake other vehicles, and their speeds are generally limited to 80 km/h, while speeds of cars are not.

Our further study is based on aggregate double-loop detector data containing, among other information, the arithmetically averaged vehicle velocities and traffic flows at 30 cross-sections (in direction North) or 31 cross-sections (in direction South), see Figure 6. We have preprocessed our one-minute data by an error correction procedure which is followed by a smoothing routine. Except for the few times in a month when none of the detectors recorded any data for a certain time period due to maintenance work or a breakdown of the computer facilities, the data showed very few errors of two types:

<sup>8</sup> In many cases, the cars involved can be moved to the emergency lane and do not block any of the freeway lanes.



**Figure 6 Top:** Sketch of the Investigated Freeway Section, Showing the Two Directions of the German Three-Lane Freeway A5 Near Frankfurt/Main  
**Bottom:** Height Profile Along the Freeway

*Notes.* Each measurement cross-section of the freeway is marked by a vertical line. It is named with the initial of the driving direction (“N” for north or “S” for south), followed by a number increasing in driving direction. The geographical position of each detector is given in kilometers according to the official notation of the road authorities.

- Sometimes the data of a cross-section consisted only of error bits for a single minute or two subsequent minutes. In this case (amounting to approximately 1% of the data), we applied a linear interpolation in time. If error bits for time intervals longer than two minutes were found, all values were set to zero. After visualization of the data, these events were clearly visible, because errors of this type occurred simultaneously for all detectors of a direction.

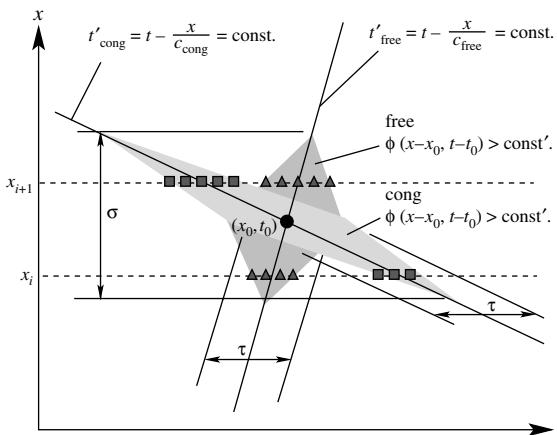
- In other cases, some detector (measuring a single lane at a specific cross-section) failed, which sometimes lasted for a few weeks. When traffic around this cross-section was congested, we averaged the data of the other lanes. Because of similar (synchronized) velocities in neighboring lanes during congestion (Kerner and Rehborn 1996b), this procedure leads to consistent results for the velocity. If there was free traffic, we interpolated the data of the preceding cross-section and the following one for the same lane.

### 5.3. Adaptive Smoothing Method

In our empirical evaluations, we used an adaptive smoothing algorithm (Treiber and Helbing 2002) as an alternative to the use of cumulative plots (Newell 1982; Cassidy and Windover 1995; Coifman 2002;

Muñoz and Daganzo 2002b; Bertini et al. 2004).<sup>9</sup> This algorithm gives a good spatiotemporal impression of traffic patterns based on the data of a few measurement cross-sections. The adaptive smoothing method uses an exponential filter  $\phi(x, t)$  that smoothes over an average time window  $\tau$  and an average spatial interval  $\sigma$ . In this study, we have always chosen  $\tau = 1.2$  min and  $\sigma = 0.6$  km. The particular feature of this method is the smoothing into the respective propagation direction of perturbations (i.e., along the “characteristic lines”). In free traffic, perturbations are assumed to propagate forward (downstream) with a speed of approximately  $c_{\text{free}} = 80$  kilometers per hour, while in congested traffic, the perturbations travel upstream with a speed of about  $c_{\text{cong}} = c_0 = -16$  kilometers per hour (cf. Figure 7). These values have been calibrated in a way that minimizes discontinuities in the three-dimensional representation of the average velocity along the freeway in the course of time and agree well with other observations (Kerner and Rehborn 1996a). The adaptive smoothing method switches automatically between the free and the congested regime based on a certain criterion, so that there is no subjective element in this method of data preprocessing; for details, see Treiber and Helbing

<sup>9</sup> For an investigation of traffic data from the same freeway stretch, but with the method of cumulative plots, see Bertini et al. (2004).



**Figure 7 Illustration of the Effects of Two Linear Homogeneous Filters with the Kernels  $\phi^{\text{free}}(x, t)$  and  $\phi^{\text{cong}}(x, t)$ , Respectively, (After Treiber and Helbing 2002)**

*Notes.* The shaded areas denote the regions considered in the calculation of a data point at  $(x, t)$ . Triangles denote the mainly contributing input data sampled in free traffic, squares the ones sampled in congested traffic.

(2002). The main properties and advantages of this method are illustrated by Figure 8.

We have applied this method to the representation of different quantities such as the average velocity  $V(x, t)$ , its inverse  $1/V(x, t)$ , the traffic flow  $Q(x, t)$ , or the vehicle density determined by  $\rho(x, t) = Q(x, t)/V(x, t)$ . It turned out that the most intuitive picture of the traffic situation was obtained by showing the average velocity over space and time, but with high values on the bottom and small values on the top. Therefore, this data representation is based on a quantity that is easy and reliable to measure (in contrast to the vehicle density). In the respective plots of the inverted average speed over location and time, congestion corresponds to “hills,” similar to a typical three-dimensional visualization of the density. In other words, the higher the value on the vertical axis of the graph, the higher is the resistance to the driver (the smaller is the velocity).

Furthermore, the method smooths out statistical fluctuations of the measurements (which are due to the fact that one-minute data average over a small number of vehicles only). In this way, the main systematic features of traffic patterns become more easily visible. Note that if the assumptions of the LWR model were true (according to which the flow and average speed would be a direct function of the density), perturbations in plots obtained with the adaptive smoothing method should not grow along homogeneous freeway sections.

## 6. Empirical Features of Congested Traffic States

In the following section, we will discuss different kinds of spatiotemporal congestion patterns that have

been observed on the investigated section of the German freeway A5.

### 6.1. Pinned Localized Cluster (PLC)

Let us start with a discussion of the particular kind of spatiotemporal congestion pattern illustrated in Figure 9. This pattern normally appears during rush hours at bottlenecks of the freeway such as on-ramps or gradients and does not move up- or downstream. Its location is fixed to the location of the bottleneck and its spatial extension is limited.

Pinned localized clusters (PLC) are characterized by a localized reduction of velocity and a higher density.<sup>10</sup> They are mostly triggered by a downstream travelling perturbation (see Figure 9a) or they are caused by an upstream travelling one that stops at the location of the bottleneck (see Figure 9b). Upstream and downstream of the PLC we have free traffic flow. When the traffic volume becomes too high, PLCs start to expand in space, which gives rise to other spatially extended congestion patterns (see Figure 20a, Figure 21b, and §6.8).

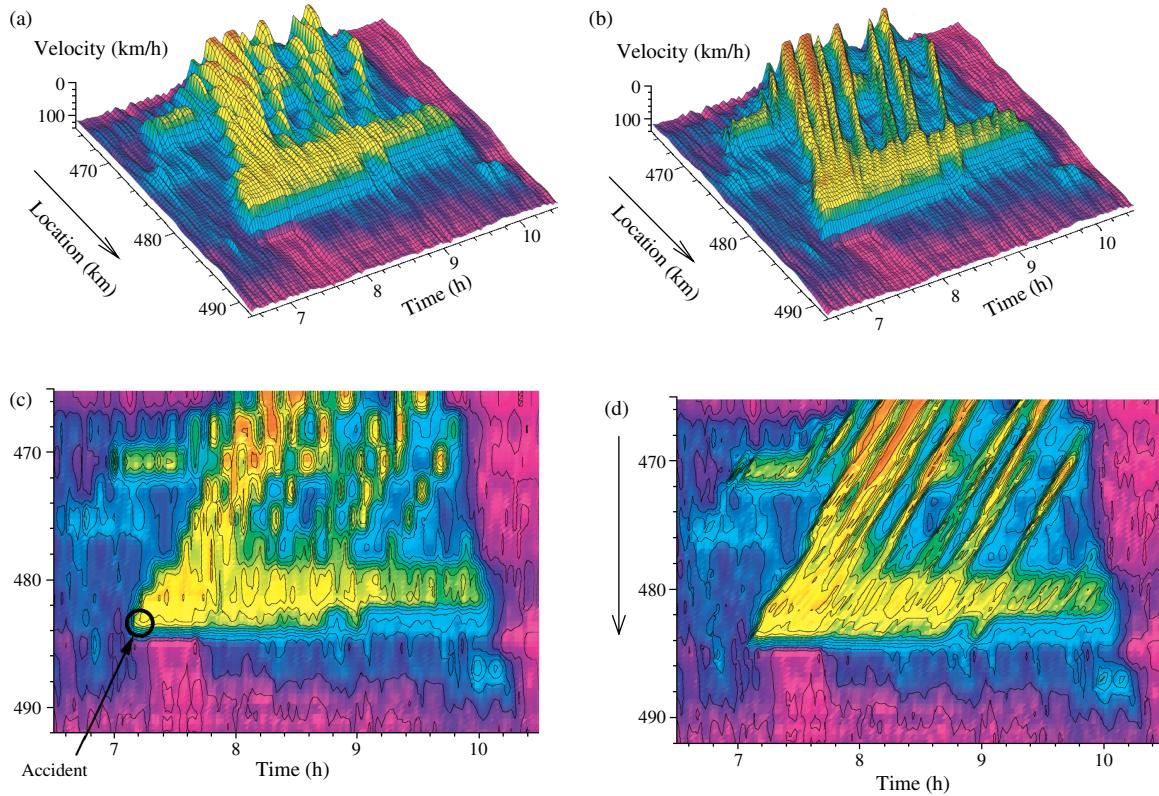
### 6.2. Homogeneous Congested Traffic (HCT)

One kind of spatially extended congested traffic is homogeneous congested traffic (HCT). We observed such patterns only after serious accidents or during holiday traffic. In Figure 10a, for example, we have diagnosed a complete closing of all three lanes by data analysis, which has been confirmed by the responsible road authorities. After a first accident occurred at 13:50 at kilometer 477.08, 16 other cars were immediately involved in six subsequent accidents. A quarter of an hour later, two more accidents happened at about the same position. The congestion pattern in Figure 10b was also caused by an accident. It occurred at 19:15 at kilometer 478.736.

In homogeneous congested traffic, the (smoothed) velocity is very low and more or less constant (i.e. homogeneous) over a longer section of the freeway. The downstream front—i.e., the small road section on which the vehicles accelerate again—is located slightly downstream of the upstream end of a serious bottleneck. The upstream jam front moves upstream, which gives rise to a spatially extended congestion pattern that grows in time.

Downstream of homogeneous congested traffic, one usually finds free traffic. Once the bottleneck is removed (e.g., the accident and the lanes blocked by it are cleared), the downstream front of congested traffic moves upstream with a speed  $C$ , which normally

<sup>10</sup> This speed reduction is probably a result of interaction-based obstruction effects, e.g., due to weaving flows—i.e., frequent lane changes close to freeway ramps.



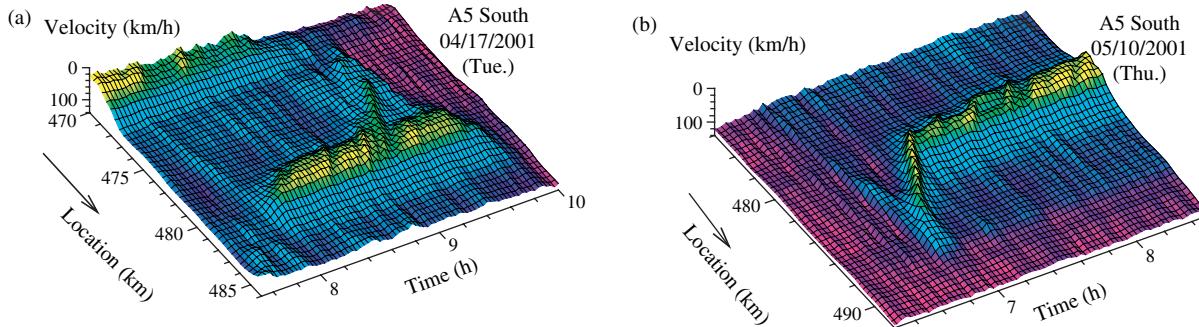
**Figure 8 Three-Dimensional Representations of the Average Speed (Top) and Velocity Contour Plots (Bottom) of Stop-and-Go Waves Observed on Thursday, 04/05/2001, in Direction South**

*Notes.* For a discussion of the traffic pattern, see §6.9. Left: Isotropic smoothing parallel to the  $x$ - and  $t$ -axes, yielding discontinuous patterns. Right: Same data using the adaptive smoothing method, where travelling waves become clearly visible. (For better illustration, only every second detector has been used here, resulting in an effective mean distance between neighboring detectors of about 2 km, while the actual mean distance is approximately 1 km, cf. Figure 6.) (b) Gives the best overview of the details of the velocity profile. The propagation speeds of the 11 main stop-waves are (from left to right, determined as described in the caption of Figure 14):  $-17.2 \text{ km/h}$ ,  $-16.3 \text{ km/h}$ ,  $-16.9 \text{ km/h}$ ,  $-15.4 \text{ km/h}$ ,  $-16.6 \text{ km/h}$ ,  $-16.77 \text{ km/h}$ ,  $-17.7 \text{ km/h}$ ,  $-17.4 \text{ km/h}$ ,  $-17.7 \text{ km/h}$ ,  $-15.7 \text{ km/h}$ , and  $-17.1 \text{ km/h}$ . Note that, while travelling through Junction Friedberg at km 471, three pairs of waves merge to form three single waves afterwards.

agrees with  $c_{\text{cong}} = c_0 \approx -16 \text{ km/h}$ . The spatial extension of homogeneous congested traffic shrinks as soon as  $|c| > |C|$  (see Figure 10b). The time of disappearance can, in principle, be forecasted when the clearing time and the time-dependent upstream flow  $Q_{\text{up}}(t)$  are known.

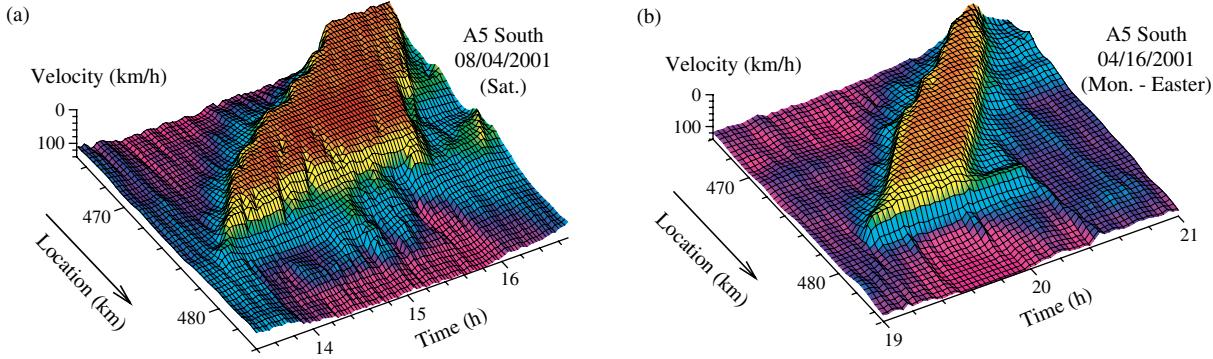
### 6.3. Oscillating Congested Traffic (OCT)

Oscillating congested traffic is another kind of extended congestion pattern. It has similar features to homogeneous congested traffic regarding its development, growth, and dissolution mechanism. However, the congested area shows more or less regular oscillations.



**Figure 9 Representative Examples of Pinned Localized Clusters at Intersection Bad Homburg**

*Notes.* For a detailed discussion of the congestion pattern displayed in (a), see §6.8. The pinned localized cluster shown in (b) is triggered by a boomerang effect (see §2.5). The PLC state starts at 7:03 A.M., when a large perturbation travelling with the speed  $-18.26 \text{ km/h}$  reaches the on-ramp at kilometer 481.3 (cross-section S15).



**Figure 10 Two Representative Examples of Homogeneous Congested Traffic Caused by Accidents**

**Notes.** In (a), the freeway is effectively blocked because the outflow reaching the intersection at cross-section S14 is only 90 vehicles per hour and lane. The initial mean speed of the upstream shockwave is  $C_1 = -(7.87 \pm 0.27)$  km/h, until it reaches the second on-ramp of Junction Friedberg at kilometer 471.5. Further upstream, after 2 P.M., it is  $C_2 = -(4.15 \pm 0.11)$  km/h. The speed of the downstream congestion front is  $c_1 = -(24.18 \pm 2.27)$  km/h on the section between kilometer 476.5 and 471.5; further upstream it is  $c_2 = -(12.42 \pm 0.24)$  km/h. In (b), a pinned localized cluster with smaller density persists for some time after the accident site is cleared, possibly because of continuing efforts in the emergency lane. The speed of the upstream shockwave between kilometer 478.7 and km 473.0 is  $C_1 = -10.5 \pm 0.4$  km/h; later on, from kilometer 470.69 to 465.15 it is  $C_2 = -(8.20 \pm 0.24)$  km/h. The speed of the downstream shock front between kilometer 477.4 and 473.0 amounts to  $c_1 = -(13.01 \pm 0.35)$  km/h. Later on, from kilometer 470.69 to 465.15 it is  $c_2 = -(16.31 \pm 1.09)$  km/h. While the speeds of the upstream shock fronts are expected to vary according to Equation (6), the change in the speed of the downstream shock front is unexpected. We suggest that this may be due to a higher maximum density  $\rho_{\text{jam}}$  or time gap  $T$  along the gradient of the freeway.

lations of speed with a frequency and amplitude that stay roughly constant over a certain period of time. The oscillations propagate upstream with a velocity  $c \approx c_{\text{cong}} = -16$  km/h, cf. Figure 11.

Oscillating congested traffic is often surrounded by free traffic flow and may be triggered by a perturbation, but it can also arise when the traffic volume exceeds a certain value. There are two different dissolution mechanisms:

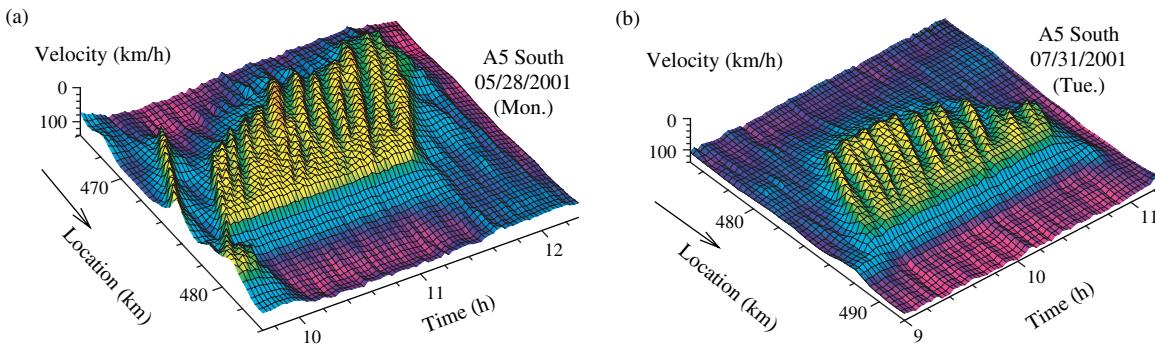
- If a temporary bottleneck is lifted (e.g., an accident site is cleared), the freeway section can again cope with the overall traffic volume and the bottleneck becomes inactivated. In this case, the downstream congestion front starts to move upstream with a speed  $c \approx c_0 = c_{\text{cong}}$ , until the upstream front is

reached and the congestion is thereby dissolved (see Figure 11a).

- If the time-dependent upstream flow  $Q_{\text{up}}$  becomes smaller than  $Q_{\text{cong}}$ , the vehicle queue shrinks and the upstream congestion front moves downstream. If  $\Delta Q$  stays long enough above zero, the upstream front will finally reach the bottleneck, terminating the congestion.

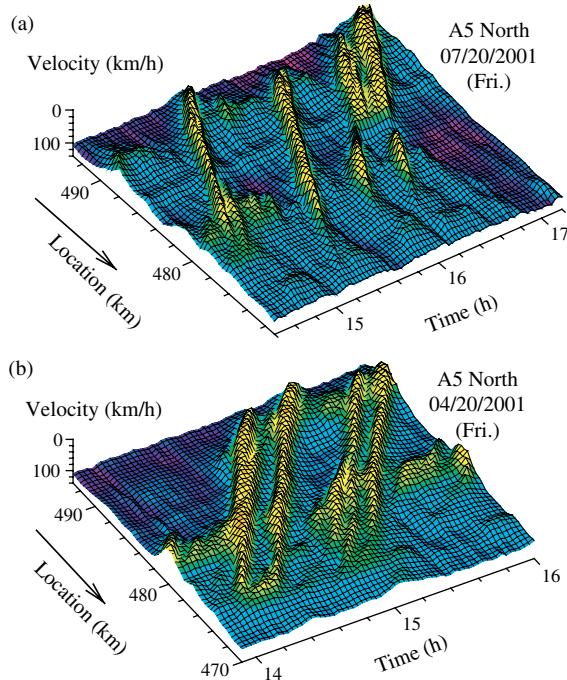
#### 6.4. Stop-and-Go Waves (SGW)

Another form of congested traffic, which has spatially extended and localized features at the same time, are stop-and-go waves. They are related to oscillating congested traffic, but have a large characteristic wavelength, while there is no typical wavelength. Stop-and-go waves (SGW) consist of a sequence of traffic



**Figure 11 Representative Examples of Oscillating Congested Traffic**

**Notes.** (a) This congestion pattern is caused by an accident at kilometer 478.325 at 9:50 A.M.. The mean propagation velocity of the 12 stop waves is  $\bar{c} = 15.7 \pm 1.2$  km/h, and the standard errors in the determination of the velocities of the individual stop waves vary between 0.1 and 1.2 km/h. (b) This oscillating congested traffic is the result of an obstruction on the fast lane between kilometers 486.0 and 486.9. The mean propagation velocity of the 11 large stop waves amounts to  $\bar{c} = 17.6 \pm 1.4$  km/h.



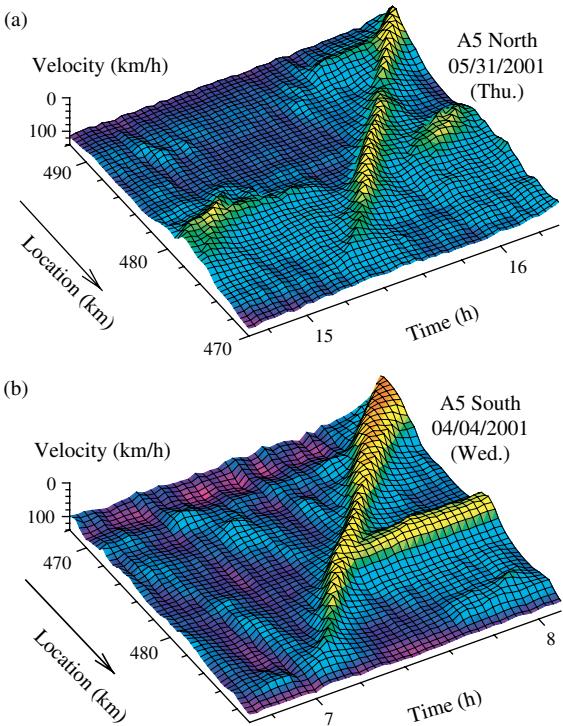
**Figure 12 Representative Examples of Stop-and-Go Waves with Propagation Speeds of (a)  $16.6 \text{ km/h} \pm 1.1 \text{ km/h}$  and (b)  $16.5 \text{ km/h} \pm 1.0 \text{ km/h}$**

*Notes.* See Figure 14 for the measurement method. In (a), stop-and-go waves appear between 14:30 and 16:10 at the uphill gradient, generating a flow-conserving bottleneck between kilometer 475 and kilometer 477 (see Figure 6). The interruption of the waves at 16:20 is caused by an accident at kilometer 482.8. (b) The stop-and-go waves are created at the uphill gradient “Köpperner Berg” between kilometer 474 and kilometer 472 from 14:00 to 15:10. In both cases, the stop-and-go waves disappear at Intersection Frankfurt West due to several off-ramps.

jams with free traffic in between. The traffic jams are localized (i.e., spatially confined in space and time), and they propagate upstream from the generating bottleneck with a velocity of  $c_{\text{cong}} \approx -16 \text{ km/h}$ . The spatial and temporal distance between two successive traffic jams vary significantly. Stop-and-go waves are sometimes triggered by small perturbations in the traffic flow or may originate from an area of pinned localized clusters. (see Figures 12, 20, 21, and §6.7). The propagation speed of stop-and-go waves does not change when they travel through pinned localized clusters (see Figures 20b, 21a) or spatially extended congested traffic (Kerner 2000a, b, 2002b; Helbing et al. 2003). Their propagation typically ends in free traffic, mostly when passing off-ramps.

### 6.5. Moving Localized Cluster (MLC)

In the case of a single moving traffic jam (instead of a sequence of them), we talk about a moving localized cluster (MLC). As for a PLC the width of an MLC state is limited, but MLC is propagating with the speed  $c_{\text{cong}} \approx -16 \text{ km/h}$  rather than staying at a particular place.



**Figure 13 Representative Examples of Moving Localized Clusters**

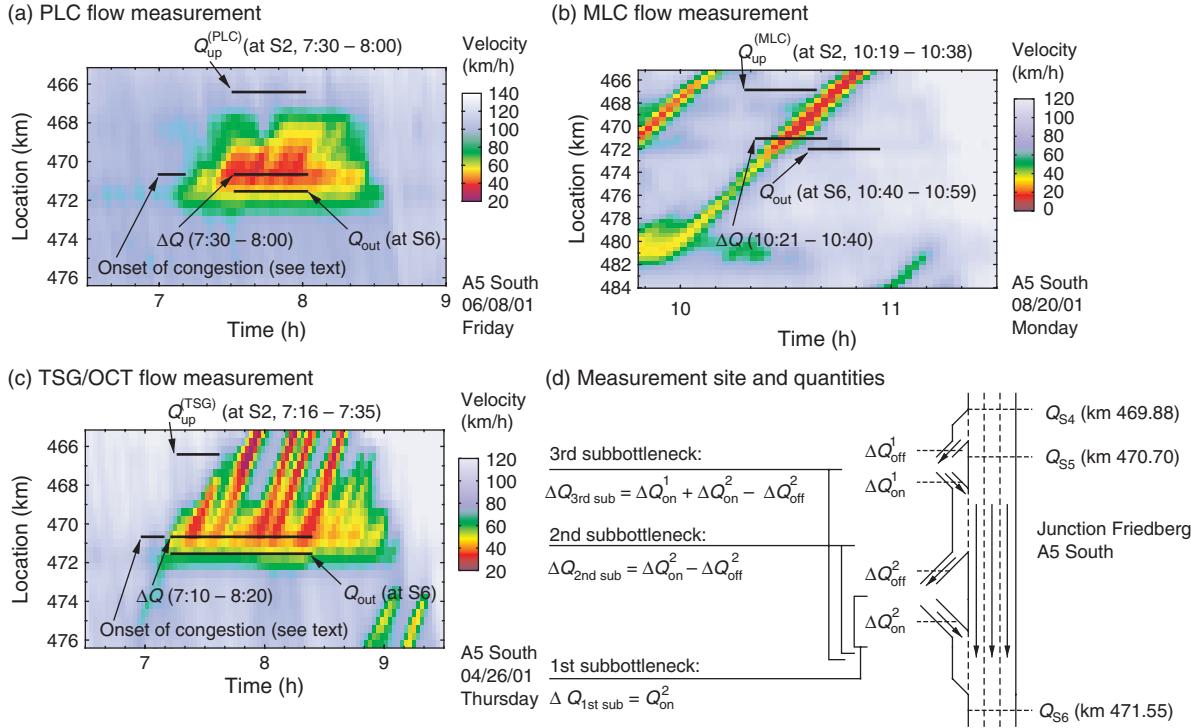
*Notes.* In (a), the MLC state emerges at the uphill gradient Köpperner Berg (kilometer 474 to kilometer 472, cf. Figure 6). The velocity of the stop wave is  $c = (-18.0 \pm 0.2) \text{ km/h}$ , which has been determined as described in the caption of Figure 2. (b) At 7:10 A.M., the moving localized cluster is triggered at kilometer 488 by a “boomerang effect” (see §2.5). The triggering is due to a peak in the flow at kilometer 486.9 (cross-section S21) between 7:01 A.M. and 7:06 A.M. with a value of 2,178 veh/h/lane, which is probably caused by a platoon of cars behind some overtaking trucks. Note that when passing the Intersection Bad Homburg at kilometer 481, the MLC state meets a small, downstream moving perturbation that initiates a PLC state at the on-ramp.

An MLC state usually originates at a bottleneck from a (large) perturbation of traffic flow. One can distinguish two different cases:

- The perturbation is large enough to travel upstream from the very beginning (see Figure 13a).
- The perturbation approaches from upstream and appears in the form of a high-flow time interval that initiates a growing, upstream traveling moving localized cluster when arriving at a bottleneck, see Figure 13a at kilometer 473 and Figure 13b at kilometer 488. Normally, this high-flow time interval is very distinct (Figure 13b) and a boomerang effect (see §2.5) can easily be observed. If the high flow at the bottleneck persists for a longer time period, it is more likely to observe a stop-and-go wave pattern or multiple MLC states rather than a single MLC; see Figure 12.

### 6.6. Empirical Phase Diagram

In contrast to previous evaluations (Schönhof and Helbing 2004; Helbing et al. 2003), the following results were obtained with a semiautomated measurement method, which is described in Figure 14. The



**Figure 14 The Bottleneck Strength  $\Delta Q$  and the Upstream Flow  $Q_{up}$  on the Freeway as the Quantities Required for Generating the Empirical Phase Diagram Were Measured by Averaging Flow Values Over a Certain Time Window, Depending on the Type of Congestion**

**Notes.** As a trapeziform weight function reduces the sensitivity to the beginning and end of the time window, we have weighted values inside the time window with 1, the boundary values with 2/3, and the closest values outside of the time window with 1/3. (a) Pinned localized clusters at Junction Friedberg, see (d), were often triggered by peaks in the flow, but very soon after the traffic breakdown a stationary state occurred, where congestion was spread over the whole junction and the cross-section S4 at km 470 was situated inside the congested area. In this stationary state, the actual bottleneck strength was assumed to be the net ramp inflow  $\Delta Q = \max(\Delta Q_{on}^1 - \Delta Q_{off}^1, 0) + \max(\Delta Q_{on}^2 - \Delta Q_{off}^2, 0)$ . The cross-sections directly upstream of the bottleneck (S4, S3) suffered from the congestion and displayed larger fluctuations. Therefore, the upstream flow was measured  $\approx 3$  km further upstream at cross-section S2 (kilometer 466.65) for the same time window as the bottleneck strength  $\Delta Q$ , exploiting that the travel time between S2 and the junction is negligible and the number of vehicles on this freeway section is conserved. The time window was chosen as the interval during which the congestion (involving detector S4) was stationary. It was also the basis for the measurement of the outflow  $Q_{out}$  at cross-section S6. (b) The inflow  $Q_{up}$  to a moving localized cluster was averaged over a time interval of 20-minutes, ending at the time when the velocity at cross-section S2 dropped below 90 km/h. The bottleneck strength was determined with the above formula for  $\Delta Q$ , averaging over a 20-minute window centered at the time, when the moving localized cluster entered the junction, i.e., when the velocity at cross-section S6 exceeded 60 km/h again. The outflow was measured at cross-section S6 during the 20 minutes after the jam passed the junction, i.e., the velocity at S4 exceeded 60 km/h again. (c) For stop-and-go waves and oscillating congested traffic,  $Q_{up}$  was determined as for moving localized clusters, while  $\Delta Q$  and  $Q_{out}$  were obtained as for pinned localized clusters. For the averaging time window we chose the period during which the congestion pattern had its characteristic form, i.e., the moments of high initial flows triggering the congestion patterns (see short thick line, “onset of congestion”) did not enter into our measurements. This method differs from our previous measurement method (Schönhof and Helbing 2004; Helbing et al. 2003), in which we focussed on the initial flows. Finally, outflows  $Q_{out}^{hindered}$  of flow-conserving bottlenecks at accident sites were measured at the next downstream cross-section during the obstruction period. The reference outflow  $Q_{out}$ , measured at this place immediately after the velocity increased again (after the accident was cleared), was used to estimate the bottleneck strength via  $\Delta Q \approx Q_{out} - Q_{out}^{hindered}$ .

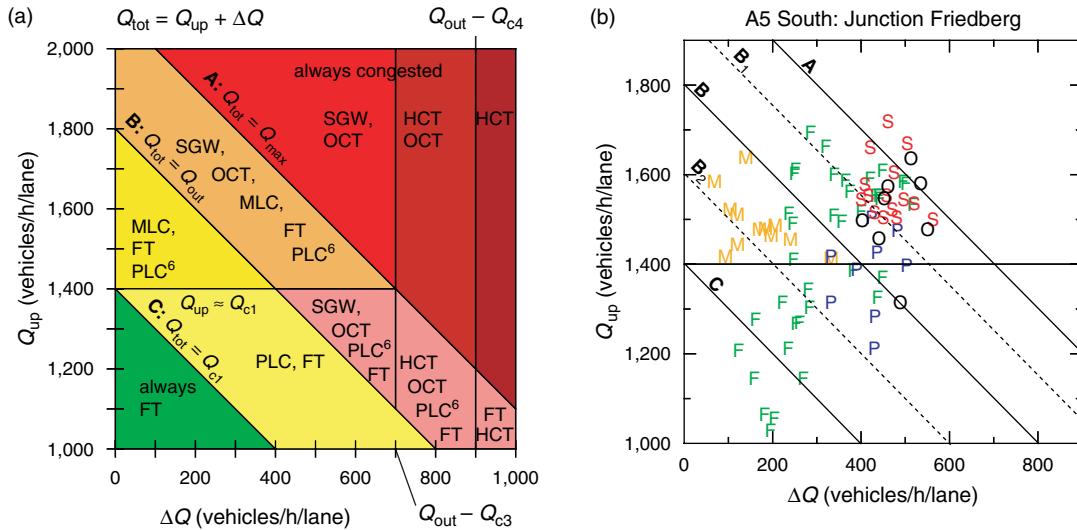
empirical phase diagram for the congestion patterns at Junction Friedberg that are depicted in Figure 15, and the observed flows for the (other) congestion patterns presented in the different figures throughout this paper, are surprisingly well compatible with the theoretical one (see Figure 4):

- For values of the overall flow  $Q_{tot} = Q_{up} + Q_{rmp}/(nL)$  below 1,400 vehicles per hour and lane, traffic is always free, while above 2,100 vehicles per hour and lane, it is always congested.
- MLC occur for upstream flows above 1,400 vehicles per hour and lane and for overall flows  $Q_{tot}$  below 1,800 vehicles per hour and lane.

- For  $1,800 \text{ veh/h/lane} \leq Q_{tot} \leq 2,100 \text{ veh/h/lane}$  and  $Q_{up} \geq 1,400 \text{ veh/h/lane}$ , we find a coexistence of FT, OCT, and SGWs.

- HCT is only found after serious accidents with lane closures, with a bottleneck strength  $\Delta Q \geq 700$  vehicles per hour and lane, cf. Figures 8, 10, and 21b. Homogeneous congested traffic has not been observed upstream of Junction Friedberg (i.e. its most upstream located off-ramp).

Although altogether there is a good agreement between the theoretical and empirical traffic states, and the empirical phase diagram is qualitatively compatible with the theoretical phase diagram, we



**Figure 15 Comparison of the Theoretical and Empirical Phase Diagram**

*Notes.* (a) The superposition of the theoretical phase diagrams of the nonlocal GKT model for negligible and large perturbations (see Figure 4) illustrates the multistable nature of traffic flows (see §4.2). (For the possible existence of PLC states up to  $Q_{\text{tot}} = Q_{\max}$  see Footnote 6.) (b) Phase points of empirically observed traffic states at the Junction Friedberg as a function of the effective ramp flow  $\Delta Q$  per freeway lane and the upstream freeway flow  $Q_{\text{up}}$  ( $F$  = free traffic,  $M$  = moving localized cluster,  $S$  = stop-and-go waves,  $O$  = oscillating congested traffic).

The flows were measured as described in Figure 14, and free traffic was evaluated over a 20-minute interval before the flow broke down, starting 40 minutes before the velocity dropped below 90 km/h at cross-section S5. Solid lines are guides to the eyes to support a comparison with the theoretical phase diagram displayed in (a). The dashed lines  $B_1$  and  $B_2$  reflect the empirical variation of the value of the outflow  $Q_{\text{out}}$  from congested traffic, see Figure 17.

want to mention one inconsistency: There are a few PLC states with overall flow  $Q_{\text{tot}}$  above 1,800 veh/h/lane. This can have several reasons:

1. The PLC state may be incorrectly classified: If  $Q_{\text{tot}} = Q_{\text{up}} + \Delta Q \approx Q_{\text{out}}$ , the difference between the upstream flow  $Q_{\text{up}}$  and the flow  $Q_{\text{cong}} = Q_{\text{out}} - \Delta Q$  in the congested area is small (of the order of 100 vehicles per hour and lane). As a consequence, the upstream congestion front moves very slowly (see Equation (6) and the discussion of Figure 9a in §6.8). Therefore, within the time period of the rush hour, the congestion front may not reach the next upstream detector, which is about one kilometer away. In conclusion, the traffic states with a total flow greater than 1,800 veh/h/lane that were classified as PLC states are potentially extended congested traffic states for which the upstream front did not reach the next detector before the end of the rush hour.

2. The outflow  $Q_{\text{out}}$  from congested traffic varies considerably for all congested traffic states (see Figure 17a), which implies that line B representing the condition  $Q_{\text{tot}} = Q_{\text{out}}$  in the phase diagram must instead be replaced by an area delimited by a lower line  $B_1$  and an upper line  $B_2$ . In fact, scaling the flows underlying the phase diagram by the empirically measured outflow  $Q_{\text{out}}$  gives a very clear picture: The nonextended traffic states are then scattered around the line  $Q_{\text{up}}/Q_{\text{out}} + \Delta Q/Q_{\text{out}} = 1$ , while the extended traffic states are above this line (see Figure 16), as expected.

The clearer separation of different congested traffic states after the scaling with the outflows  $Q_{\text{out}}$  indicates the importance of understanding the varying outflows  $Q_{\text{out}}$ . In fact, the relationship

$$Q_{\text{out}} = Q_{\text{cong}}(V_{\text{cong}}) + \Delta Q \quad (23)$$

suggests that  $Q_{\text{out}}$  could be specified as a function of the average speed  $V_{\text{cong}}$  in the congested area. In other words, the capacity drop would be lower the higher the vehicle speed in the congested area is, which sounds plausible.

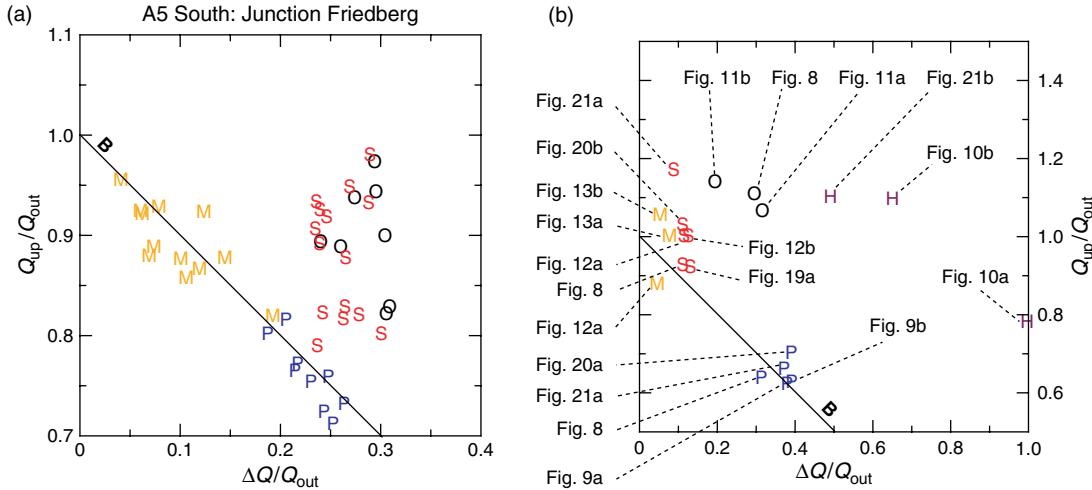
Figure 17 supports this interpretation: First, the empirical relationship  $Q_{\text{cong}}(V_{\text{cong}})$  is quite compatible with the theoretical relationship

$$Q_{\text{cong}}(V_{\text{cong}}) = \frac{\rho_{\max}|c|}{1 + |c|/V_{\text{cong}}} \quad (24)$$

that would follow from the “jam line”

$$Q_{\text{cong}}(\rho) = \frac{1}{T} \left( 1 - \frac{\rho}{\rho_{\max}} \right) \quad (25)$$

and  $V_{\text{cong}}(\rho) = Q_{\text{cong}}(\rho)/\rho$ . Here,  $\rho_{\max}$  again denotes the maximum vehicle density per lane,  $|c| = 16 \text{ km/h}$  the propagation speed of a downstream congestion front, and  $T = 1/\rho_{\max}/|c|$  the safe time gap. Second,  $Q_{\text{out}}$  clearly grows with the average speed  $V_{\text{cong}}$  in the congested area (see Figure 17a). Third, Figure 17b suggests that the flow  $Q_{\text{cong}} = Q_{\text{out}} - \Delta Q$ , and therefore



**Figure 16 Empirical Phase Diagrams, Where the Flows Have Been Scaled by the Respective Outflows  $Q_{\text{out}}$  (See Figure 14 for the Measurement Method)**

*Notes.* Illustration (a) shows the congested traffic states observed at Junction Friedberg in direction South, while (b) illustrates the flows corresponding to the traffic states displayed in the figures throughout this paper ( $M$  = moving localized cluster,  $S$  = stop-and-go waves,  $O$  = oscillating congested traffic,  $P$  = pinned localized clusters). It can be clearly seen that the nonextended traffic states are scattered around the line  $Q_{\text{up}}/Q_{\text{out}} + \Delta Q/Q_{\text{out}} = 1$ , as expected, whereas the extended traffic states are above this line. Moreover, pinned localized clusters, moving localized clusters, stop-and-go waves/oscillating congested traffic, and homogeneous congested traffic are well separated from each other.

$Q_{\text{out}}$  itself, tends to increase with the sum of the off-ramp flows  $\Delta Q_{\text{off}} > 0$  along Junction Friedberg and with the on-ramp flow  $\Delta Q_{\text{on}}^1$  of the first on-ramp. In accordance with observations, this surprising result suggests that it is *not* correct to treat all the freeway lanes in the same way.

Rather, the left lane is used by vehicles passing the junction, while the right lane is mainly used by vehicles entering and leaving along Junction Friedberg, and by trucks. The middle lane is characterized by mixed use. Considering this, it becomes understandable that the off-ramp flows can increase the outflow  $Q_{\text{out}}$  as they leave gaps between the vehicles remaining on the right lane, which facilitate entering the right and middle lanes of the freeway. The on-ramp flow  $\Delta Q_{\text{on}}^1$  of the first on-ramp seems to increase the outflow further downstream, as it enhances the flow separation between the right lane and the other two lanes. This particularly underscores that Junction Friedberg cannot be treated like an isolated bottleneck, in contrast to Kerner's view (2002b). Rather, it is understandable only by a detailed analysis of the interactions between several on- and off-ramps. During congested time periods, Junction Friedberg has to cope with an overall on-ramp flow of as much as 1,700 veh/h (which requires the capacity of at least one full lane), while the overall off-ramp flow is only 375 veh/h.

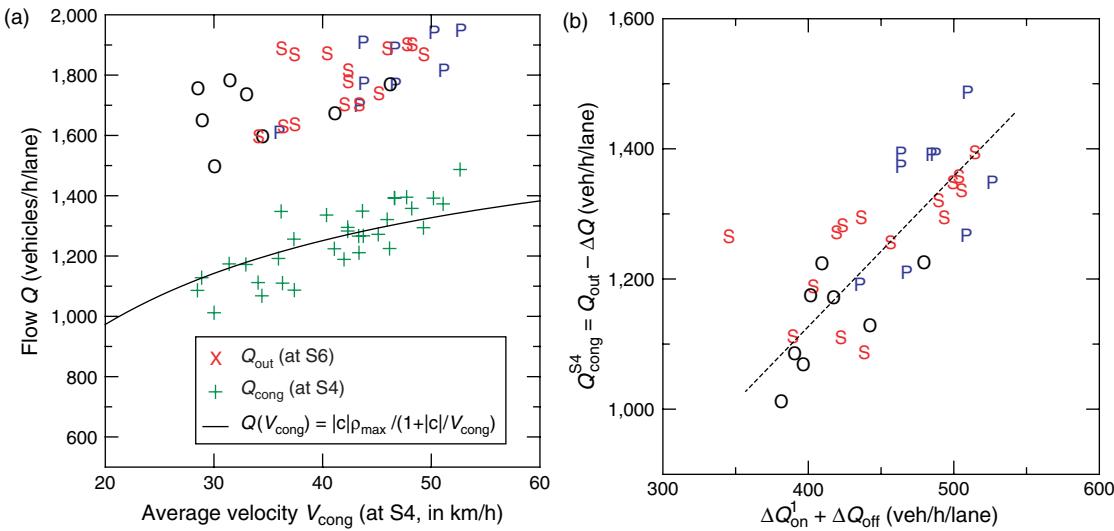
## 6.7. Spatial Transitions Between Different Congestion States and “Pinch Effect”

Freeways are not at all spatially homogeneous. Due to the existence of on- and off-ramps, gradients, curves,

etc., a freeway is composed of road sections of different capacities, even if the number of lanes is constant. If the capacity is reduced from one road section to the next one in a downstream direction, we speak of a bottleneck. Upstream of bottlenecks, there is a danger of queue formation or congestion if the traffic volume becomes too high, while downstream of the upstream end of a bottleneck, one mostly observes free traffic if there is not another activated bottleneck further downstream.

The phase diagrams depicted in Figure 4 assume only a single bottleneck. In principle, however, phase diagrams could also be generated for more complex freeway geometries with several bottlenecks. These would be multidimensional (with one additional dimension per bottleneck). For this reason, there are many more possibilities for classifying different congested traffic states for freeways with several bottlenecks. In the simplest case, particularly for largely separated bottlenecks, these states would just correspond to the possible combinations of the six traffic states occurring at a single bottleneck, namely FT, PLC, MLC, SGW, OCT, and HCT. It could, however, happen that spatially extended traffic states influence the bottlenecks and traffic states upstream due to spillover effects.

For illustration, let us focus here on the combination of an on-ramp with an off-ramp further upstream (see Figure 18 for a sketch). This freeway design is often chosen to reduce the magnitude of traffic breakdowns, because it is favorable when vehicles leave the freeway before new ones enter. What would a bottleneck analysis analogous to the one in §4.2 predict for



**Figure 17** Traffic Flows at Junction Friedberg During Congestion

*Notes.* (a) The outflows  $Q_{\text{out}}$  and the upstream congested flow  $Q_{\text{cong}}$  for the PLC, SGW, and OCT states (represented by  $\times = \text{P}$ ,  $\text{S}$ , and  $\text{O}$ , respectively) at Junction Friedberg depend on the average velocity  $V_{\text{cong}}$  in the congested area. With a maximum density of  $\rho_{\max} = 109 \text{ veh/km/lane}$ , the average dependence  $Q_{\text{cong}}(V_{\text{cong}})$  is well consistent with theoretically predicted formula (24). (b) The flow  $Q_{\text{cong}}$  at Junction Friedberg tends to grow with the sum of the off-ramp flows  $\Delta Q_{\text{off}} > 0$  and the on-ramp flow of the first on-ramp denoted by  $\Delta Q_{\text{on}}^1$ . The dashed line serves as guide for the eye and has the slope +2. The subsequent on- and off-ramp flows obviously facilitate a higher average velocity of the vehicles and a higher traffic efficiency within the congestion area (see main text).

this setup? In order to discuss this, let us again denote the outflow capacity downstream of the on-ramp by  $Q_{\text{out}}$ , its bottleneck strength equivalent to the on-ramp flow  $Q_{\text{rmp}}$  by  $\Delta Q_{\text{on}} = Q_{\text{rmp}}/n > 0$ , the upstream flow by  $Q'_{\text{up}}$ , and the average congested flow resulting immediately upstream of the on-ramp by  $Q_{\text{cong}}$ . In contrast, we will denote the same quantities relating to the area of the off-ramp by primes ('), but we will introduce the abbreviation  $-\Delta Q_{\text{off}} = -Q'_{\text{rmp}}/n' < 0$  for the effect of the off-ramp flow  $Q'_{\text{rmp}} > 0$ .

We expect the following dynamics: Traffic will break down at the on-ramp bottleneck first. If  $Q'_{\text{up}} > Q_{\text{cong}} = Q_{\text{out}} - \Delta Q_{\text{on}}$ , we will have a growing area of extended congested traffic, and it will eventually reach the location of the off-ramp (cf. Figure 18a). What will happen there? From that time on, when the off-ramp is completely congested (cf. Figure 18b), the capacity downstream of the off-ramp will suddenly go down from  $Q'_{\text{out}} = Q_{\text{out}}$  to the congested flow

$$Q'_{\text{out}} = Q_{\text{cong}} = Q_{\text{out}} - \Delta Q_{\text{on}} \quad (26)$$

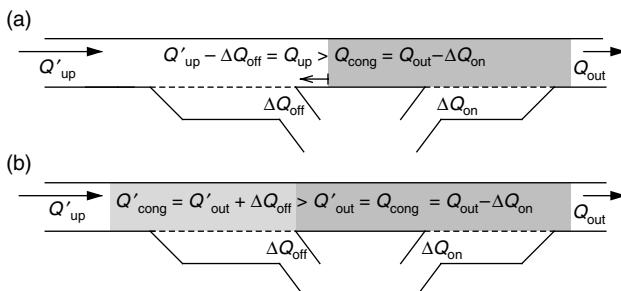
due to a spillover effect. This abrupt change in the bottleneck capacity restricts the capacity *upstream* of the off-ramp to  $Q_{\text{cong}} + \Delta Q_{\text{off}}$ . For  $Q_{\text{cong}} + \Delta Q_{\text{off}} < Q_{\text{out}}$ , a bottleneck at this position exists, and its effective strength  $\Delta Q$  is given by the difference

$$\Delta Q = Q_{\text{out}} - (Q_{\text{cong}} + \Delta Q_{\text{off}}) = \Delta Q_{\text{on}} - \Delta Q_{\text{off}} \quad (27)$$

of these values, because this is the amount of outflow from congested traffic that cannot be served by the off-ramp and the downstream freeway flow. For  $Q_{\text{cong}} + \Delta Q_{\text{off}} > Q_{\text{out}}$ , no bottleneck occurs, which corresponds to a bottleneck strength  $\Delta Q = 0$ . This finally results in the expression

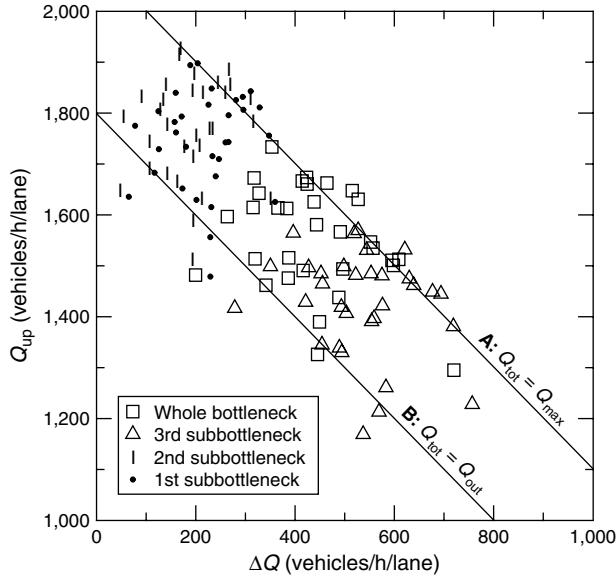
$$\begin{aligned} \Delta Q &= \max(\Delta Q_{\text{on}} - \Delta Q_{\text{off}}, 0) \\ &= \max\left(\frac{Q_{\text{rmp}}}{n} - \frac{Q'_{\text{rmp}}}{n'}, 0\right) \leq \Delta Q_{\text{on}} \end{aligned} \quad (28)$$

(Helbing 2003). Whenever  $\Delta Q_{\text{on}} - \Delta Q_{\text{off}} < 0$ , the off-ramp bottleneck is deactivated. For  $\Delta Q > 0$ , the



**Figure 18** Illustration of Our Notations for a Combination of an On-Ramp Bottleneck with an Upstream Off-Ramp

*Notes.* Traffic flow is from left to right. (a) When the upstream flow  $Q'_{\text{up}}$  upstream of the off-ramp minus the off-ramp flow  $\Delta Q_{\text{off}}$  exceeds the outflow  $Q_{\text{out}}$  from congested traffic minus the on-ramp flow  $\Delta Q_{\text{on}}$ , congested traffic upstream of the on-ramp (grey area) is expected to grow. (b) As soon as the congested area extends up to the location of the off-ramp, the off-ramp bottleneck is activated. Its outflow  $Q'_{\text{out}}$  is given by the congested flow  $Q_{\text{cong}}$  upstream of the on-ramp, while the congested flow  $Q'_{\text{cong}}$  upstream of the off-ramp is higher by the amount  $\Delta Q_{\text{off}}$  corresponding to the off-ramp flow. The flow upstream of the congestion and the off-ramp is denoted by  $Q'_{\text{up}}$ .



**Figure 19 Flow Data for the Different Subbottlenecks of Junction Friedberg in Direction South (See Figure 14d)**

*Notes.* According to our theory, congestion patterns upstream of the off-ramp at kilometer 470 are expected when all subbottlenecks of the whole effective bottleneck between cross-sections S4 and S6 suffer from extended congested traffic, specifically the on-ramp upstream of cross-section S6 (subbottleneck 1), the combination of this on-ramp with the upstream off-ramp (subbottleneck 2), and their combination with the next on-ramp upstream (subbottleneck 3). For this, the corresponding values of  $Q_{up} + \Delta Q$  should all be located above the diagonal corresponding to  $Q_{up} + \Delta Q = Q_{out}$  (solid line). This is actually the case for the majority of data points, and the remaining points are still within the area of possible outflows from extended congestion states ( $Q_{tot} > 1,600$  veh/h/lane, cf. Figure 17a). In fact, after the first signs of a traffic breakdown (namely, reduced speeds at kilometer 471.55, i.e., cross-section S6), congestion at Junction Friedberg spreads very rapidly and typically extends over the whole junction up to cross-section S4 at kilometer 469.9 within a few minutes. The flow measurements at the subbottlenecks were made as follows: The inflow  $Q_{up}$  and the bottleneck strength  $\Delta Q$  were determined over a time window of 10 minutes directly before the velocity dropped below 90 km/h at cross-section S5. The upstream flow  $Q_{up}^{3rd\ sub}$  of the 3rd subbottleneck was obtained by the flow measured at cross-section S5. Because there are no detectors to measure the upstream flows  $Q_{up}$  of the 1st and 2nd subbottlenecks directly, we applied the formulas  $Q_{up}^{2nd\ sub} = Q_{S5} + \Delta Q_{on}^1$  and  $Q_{up}^{1st\ sub} = Q_{S5} + \Delta Q_{on}^1 - \Delta Q_{off}^2$  (see Figure 14d, where the bottleneck strength  $\Delta Q$  for the subbottlenecks is also defined).

resulting congested flow upstream of the off-ramp is expected to be

$$\begin{aligned} Q'_{cong} &= Q'_{out} + \Delta Q_{off} = Q_{out} - \Delta Q_{on} + \Delta Q_{off} \\ &= Q_{out} - \Delta Q = Q_{cong} + \Delta Q_{off}. \end{aligned} \quad (29)$$

Moreover, we find

$$Q'_{up} - Q_{up} = \Delta Q_{off} = Q'_{cong} - Q_{cong}. \quad (30)$$

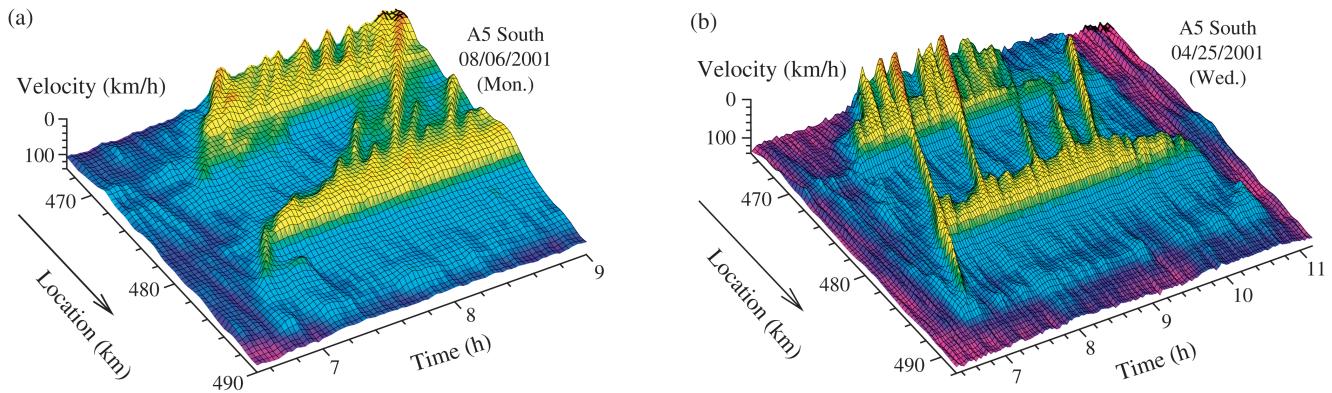
In conclusion, if congested traffic upstream of an on-ramp reaches an upstream off-ramp, the off-ramp becomes a bottleneck of strength  $\Delta Q$ , which is given by the difference between the on-ramp and the off-ramp flows (or zero, if this difference would be negative). Figure 19 illustrates the validity of this picture.<sup>11</sup>

Because  $\Delta Q \leq \Delta Q_{on}$  according to Equation (28), and because of  $Q'_{cong} \geq Q_{cong}$  according to Equation (29), the congestion upstream of the off-ramp is expected

to be milder than the congestion upstream of the on-ramp. Note, however, that in reality the situation becomes more complicated due to the following effects: On the one hand, spatially extended congestion upstream of the on-ramp, depending on the off-ramp and upstream flows, may pass the location of the off-ramp. This will obstruct the flow of departing vehicles, i.e., reduce the off-ramp flow  $\Delta Q_{off}$  compared to the situation with free flow at the off-ramp. On the other hand, congested traffic upstream of the off-ramp will motivate a larger fraction of drivers to leave the freeway once they reach the off-ramp. This may adjust the bottleneck strength of the off-ramp bottleneck in such a way that the upstream congestion front fluctuates around the upstream end of the off-ramp (cf. the discussion of Figure 20b in §6.9). The fluctuation and the resulting flips of  $\Delta Q$  can be a source of oscillatory dynamics upstream of the off-ramp, which is often observed (see Figures 8, 20, and 21a).

The above considerations suggest a new interpretation of the “pinch effect” or the “general pattern,” respectively, the “general pattern” (Kerner 1998a; 2002a, b): After 7:00 A.M., Figure 20a shows a congestion pattern at the on-ramp of Intersection Bad Homburg at kilometer 481.3, which until 7:40 A.M. looks

<sup>11</sup> At Junction Friedberg and Intersection Bad Homburg, the situation is actually even more complicated, because these are made up of a series of several on- and off-ramps; see Figures 6 and 14d. However, the flow balance is again expected to determine the effective bottleneck strength  $\Delta Q$  as long as the resulting values are nonnegative (otherwise, the off-ramp bottleneck is deactivated).



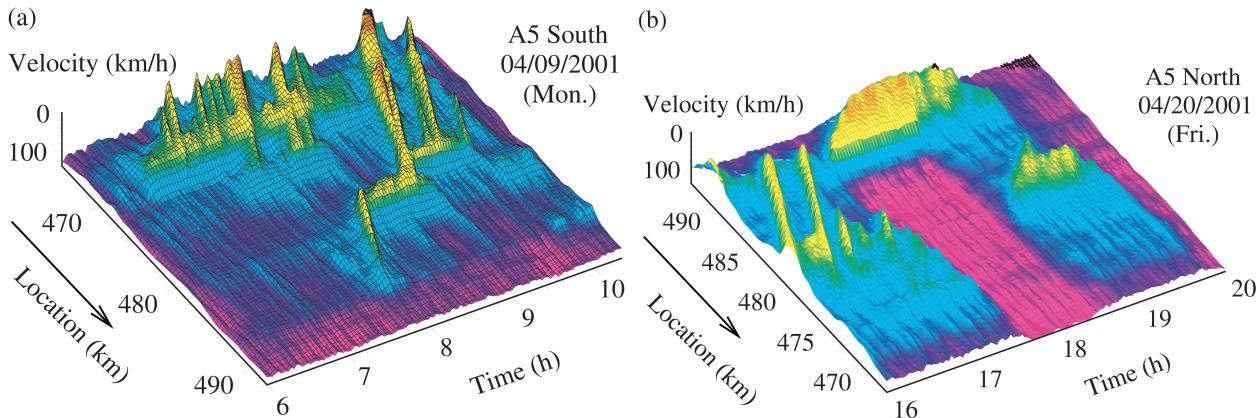
**Figure 20 Empirical Example of the “Pinch Effect”**

*Notes.* (a) Representative example of the “pinch effect” upstream of Intersection Bad Homburg at kilometer 481. One can clearly see the typical spatial sequence of homogeneous congested, oscillating congested, and stop-and-go traffic in upstream direction. (b) In this similarly looking congestion pattern, two stop-and-go waves pass the next bottleneck Junction Friedberg at kilometer 471, grow wider over there, and continue with a slightly changed velocity ( $15.6 \pm 0.3$  km/h compared to  $18.0 \pm 0.5$  km/h for the first stop wave and  $15.2 \pm 0.5$  km/h compared to  $16.2 \pm 0.4$  km/h for the second).

like a PLC. This classification fits with the overall bottleneck strength  $\Delta Q = 606$  veh/h/lane for the two on-ramps and the inflow of  $Q_{\text{up}} = 1,100$  veh/h/lane (values averaged over the time interval 7:10 A.M. to 7:40 A.M.). However, for this time interval, the outflow at cross-section S16 downstream of the Intersection Bad Homburg is only  $Q_{\text{out}} = 1,568$  veh/h/lane. Thus, the total inflow of  $Q_{\text{tot}} = Q_{\text{up}} + \Delta Q = 1,706$  veh/h/lane actually leads to a slowly growing congestion pattern with the appearance of oscillating or even homogeneous congested traffic (the velocities within the congested area fluctuate between 10 km/h and 30 km/h). At 7:53 A.M., the velocity drops at cross-section S11, indicating that the whole intersection is congested and that the off-ramp bottleneck is activated. For this new situation, the inflow (measured three kilometer upstream at cross-section S8 as described in Figure 14c) is given by  $Q_{\text{up}} = 1,487$  veh/h/lane, and the bottleneck strength by  $\Delta Q = 206$  veh/h/lane (averaged over the next hour,

see Figure 14c). The point in the phase diagram is located inside the MLC region, but relative to the low outflow of  $Q_{\text{out}} = 1,616$  veh/h/lane (see Figure 16), the point instead corresponds to stop-and-go waves, as observed. While travelling upstream, some of these waves vanish due to the small upstream flow of 1,487 veh/h/lane, whereas one of them persists. This surviving SGW enters another area of congested traffic upstream of Junction Friedberg at kilometer 471.5. Note that around kilometer 478, we do not really see a merging of small oscillations in favor of a few remaining moving traffic jams, in contrast to the suggestion by Kerner (1998a). Rather, the oscillations seem to disappear, i.e., they seem to be dissolved in free traffic.

Although Kerner views the pinch effect as a feature of a general pattern of congestion (Kerner 2002b), we did not frequently observe it. The dissolution or merging of narrow jams does not occur very often. This may be due to the applied smoothing procedure. Structures on a smaller scale than  $\sigma = 0.6$  km



**Figure 21 Combinations of Different Congested Traffic States**

*Notes.* See main text for details.

may actually merge, but from our point of view, it is hard to tell them apart from fluctuations. Therefore, we instead propose to interpret the pinch effect as result of the particular dynamics generated at a bottleneck composed of subsequent off- and on-ramps, as explained above.

### 6.8. Transitions Between Congested Traffic States in Time

Due to the metastability of traffic flows, transitions between different congested traffic states in time, but at the same bottleneck, do not occur frequently. Nevertheless, the phase diagram predicts the possibility of such transitions. Normally, in the absence of large perturbations (see Figure 4a) we expect a direct transition from free flow to spatially extended congested traffic when the overall traffic flow  $Q_{\text{tot}}$  increases. This mostly implies a direct transition from free to OCT or SGWs, which may sometimes be hard to distinguish empirically. Homogeneous congested traffic is only expected when a bottleneck becomes extremely large, as for accidents with lane closures or major holiday traffic. If an accident site is lifted during the rush hour, an HCT state may become an OCT or SGW state (see Figure 8). In cases where the total flow  $Q_{\text{tot}}$  exceeds a certain value, a PLC should be able to grow and to generate an extended form of congestion (see Figure 21a at km 471.5). The opposite should also be possible if the flow goes down (see Figure 8).

As an example for more complex transitions in time, let us discuss Figure 9a in detail: At 7:56 A.M., the traffic breaks down at the on-ramp of Junction Bad Homburg. Due to a queueing process, the congested area extends, and at 8:15 A.M. reaches cross-section S13 at kilometer 480. During this time interval, the total flow should exceed the outflow, i.e.,  $Q_{\text{tot}} > Q_{\text{out}}$ . The upstream flow is given by the inflow measured at cross-section S11 minus the off-ramp flow at km 480.8. It amounts to  $Q_{\text{up}} = 1,750 \text{ veh/h/lane} - 484 \text{ veh/h/lane} = 1,266 \text{ veh/h/lane}$ . The bottleneck strength  $\Delta_Q = 693/\text{h/lane}$  is determined by dividing the on-ramp flow by the number of freeway lanes. For these flow values, the phase diagram predicts oscillating or even homogeneous congested traffic. The difference between the total flow  $Q_{\text{tot}} = Q_{\text{up}} + \Delta_Q = 1,959 \text{ veh/h/lane}$  and the measured outflow at cross-section S16 of 1,766 veh/h/lane causes the upstream congestion front to expand by a distance of 1,300m in 20 minutes. Between 8:15 A.M. and 8:55 A.M., it travels from cross-section S13 to S12. During this time period, there is a stationary phase from 8:30 A.M. to 8:50 A.M., where the total flow  $Q_{\text{tot}} = Q_{\text{up}} + \Delta_Q = 1,893 \text{ veh/h/lane}$  matches the outflow  $Q_{\text{out}} = 1,899 \text{ veh/h/lane}$ . The corresponding upstream flow and bottleneck strength are  $Q_{\text{up}} = 1,180 \text{ veh/h/lane}$  and  $\Delta_Q = 713 \text{ veh/h/lane}$ , respectively. In Figure 16b, this state has been classified

as PLC, although it may also be interpreted as a homogeneous congested traffic state that is stationary, i.e., not extending any further. The reason for the upstream jam front to cross-section S12 and to finally arrive at cross-section S11 at 9:05 A.M. is a fluctuation of the inflow ( $Q_{\text{S11}}^{8:55-9:05 \text{ A.M.}} = 1,901 \text{ veh/h/lane} > Q_{\text{S11}}^{8:30-8:50 \text{ A.M.}} = 1,765 \text{ veh/h/lane}$ ). After the off-ramp bottleneck is activated around 9:00 A.M., its upstream flow  $Q_{\text{up}} = 1,867 \text{ veh/h/lane}$  and the resulting overall bottleneck strength  $\Delta_Q = 55 \text{ veh/h/lane}$  allow a light SGW state during the next 10 minutes. However, after the first SGW has passed cross-section S11, the flow values of  $Q_{\text{up}}$  and  $\Delta_Q$  have already decreased and lead to a slow dissolution of congestion rather than persistent SGW.

### 6.9. Combinations of Different Congestion States Along the Freeway

Congestion normally starts to form upstream of the freeway section for which the difference between the arriving traffic volume and the section capacity (bottleneck capacity) is highest. With increasing traffic volume, more and more bottlenecks are activated. Therefore, we normally have a sequence of different congested traffic patterns along the freeway, and the spatiotemporal patterns are given by interactions between different bottlenecks. The bottlenecks influence each other in a downstream direction by determining the upstream flow of the next bottleneck, and in an upstream direction by the propagation of congested traffic patterns (spillback effects). In the following, we will discuss several examples.

Let us start with the traffic pattern shown in Figure 20b. There, a boomerang effect (see §2.5) triggers an MLC at about 6:55 A.M. Around 7:10 A.M., this MLC state arrives at kilometer 481.3 (cross-section S15 of Intersection Bad Homburg), where it triggers a PLC at the location of the on-ramp, which eventually extends over the whole intersection ( $\Delta_Q = 698 \text{ veh/h/lane}$ ). Within the next hours, the upstream front of the localized cluster fluctuates between cross-section S12 (which is always congested after 7:50 A.M.) and cross-section S11 upstream of the long off-ramp, where the velocity stays, overall, above 80 km/h, apart from the times when an MLC passes. It seems that by fluctuations in the upstream flow, the off-ramp bottleneck is alternately activated and deactivated (see §6.7). The bottleneck strength between 7:20 A.M. and 9:40 A.M. is given by  $\Delta_Q = \Delta Q_{\text{on}} - \Delta Q_{\text{off}} \approx 187 \text{ veh/h/lane}$  and the upstream flow (measured before the first SGW at cross-section S8 from 7:19 A.M. to 7:38 A.M.) amounts to  $Q_{\text{up}} = 1,791 \text{ veh/h/lane}$ , which falls into the area of SGW/OCT states (see Figure 16b). Around 9:00 A.M., when the morning rush

hour calms down, some of the SGW and the congestion at kilometer 471 (Junction Friedberg) start to disappear, whereas the congestion at kilometer 480 (at Intersection Bad Homburg) persists even longer.

We will now continue with Figure 8. According to police reports, at 7:20 A.M. an accident occurred at kilometer 483.1. This accident is clearly visible in the data, because the downstream front of the jam is located between the cross-sections S17 (km 483.0) and S18 (km 484.1), where the freeway usually has no bottleneck. In our data, the flow at cross-section S17, along with the velocity, dropped at 7:08 A.M., whereas the flow drop at cross-section S18 occurred one minute later. This is a strong hint that the time of the accident was not reported exactly. During the first half hour after the accident (7:10 A.M. to 7:40 A.M.), the outflow was 1,231 veh/h/lane. This very light traffic resulted in high velocities downstream of the accident location, which can be seen in the plots. Between 7:40 A.M. and 10:00 A.M., the outflow reached normal values again (1,781 veh/h/lane). Between 7:10 A.M. and 7:40 A.M., the bottleneck strength was estimated as  $\Delta Q = Q_{\text{out}}^{\text{normal}} - Q_{\text{out}}^{\text{hindered}} = (1,781 - 1,231)$  veh/h/lane = 550 veh/h/lane. At 9:00 A.M., the bottleneck forming the downstream congestion front moved from the accident site at kilometer 483.1 to the on-ramp of the Intersection Bad Homburg further upstream (cross-section S15 at kilometer 481.3). The inflow measured between 6:48 A.M. and 7:07 A.M. at cross-section S16 (kilometer 482.2), before congestion reached this point, was  $Q_{\text{up}} = 1,971$  veh/h/lane. According to the phase diagram, these flows are related to oscillating congested traffic in the neighborhood of homogeneous congested traffic. Upstream of the accident site at cross-section S16, the average velocities of the three lanes fluctuate between 12 km/h and 35 km/h; upstream of the Intersection Bad Homburg the fluctuations become larger, resulting in clear oscillating congested traffic. The flow conditions of the subsequent SGWs and PLC states are displayed in Figure 16b.

Figure 21a shows the breakdown of free traffic flow at kilometer 471.5 (Junction Friedberg). Congestion starts at the second on-ramp at 6:23 A.M. with a velocity drop at cross-section S6. Five minutes later, the velocity upstream of the first on-ramp at cross-section S4 breaks down as well, and the junction is completely congested. After 10 more minutes, during which the bottleneck strength increases, SGW are finally formed (see §6.7). At about 7:50 A.M., we observe the formation of a moving localized cluster around kilometer 486 at the off-ramp of Intersection Frankfurt North-West, which is triggered by the boomerang effect (see §2.5). The MLC turns into a PLC when it reaches cross-section S14 (Intersection Bad Homburg) at 8:10 A.M. because of the small

inflow: From 8:00 A.M. to 8:10 A.M., the upstream flow  $Q_{\text{up}}$  measured at cross-section S13 is just 1,079 veh/h/lane, as many vehicles are leaving over the off-ramp located there (see Figure 6). However, the following freeway flow (8:10 A.M. to 8:30 A.M.) at this cross-section of 1,266 veh/h/lane and the effective bottleneck strength of  $\Delta Q = 704$  veh/h/lane caused by the ramps still allow the PLC to exist, considering the large perturbation by the MLC. According to the phase diagram of Figure 16b, the LC should even be slowly growing. In fact, the the velocity drops at cross-section S13 at 8:23 A.M. and at cross-section S12 at 8:48 A.M. At 8:53 A.M., when the congestion reaches cross-section S11 and extends over the whole junction, we expect a different traffic state (see §6.7). In fact, moving localized traffic becomes visible, which triggers an accident at 8:50 A.M. between cross-sections S12 and S13, namely, at kilometer 479.2. This results in a bottleneck with an outflow of  $Q_{\text{out}}^{\text{hindered}} = 1,431$  veh/h/lane, so that the downstream congestion at Intersection Bad Homburg disappears. The bottleneck strength of the accident location is  $\Delta Q = Q_{\text{out}} - Q_{\text{out}}^{\text{hindered}} = (1,551 - 1,431)$  veh/h/lane  $\approx 120$  veh/h/lane, where we have estimated the outflow  $Q_{\text{out}} \approx 1,551$  veh/h/lane by the outflow at cross-section S11 between 9:47 A.M. and 9:56 A.M., i.e., after the accident was cleared. The upstream flow measured at cross-section S8 was  $Q_{\text{up}} = 1,816$  veh/h/lane, which is consistent with the observed stop-and-go waves (see Figure 16b). These cross Junction Friedberg at kilometer 471.5, because the flow conditions over there (with  $Q_{\text{up}} = 1,638$  veh/h/lane and  $\Delta Q = 136$  veh/h/lane) allow the survival of MLCs (see Figure 14). Note that the last MLC dissolved before it reaches Junction Friedberg, as its outflow  $Q_{\text{out}} \approx 1,860$  veh/h/lane is greater than its inflow  $Q_{\text{up}} \approx 1,660$  veh/h/lane, which is the outflow of the upstream MLC.

Figure 21b displays SGWs upstream of Intersection Bad Homburg; this congestion pattern is generated at the uphill gradient shortly downstream of this intersection. Around 17:20, homogeneous congested traffic forms at the Intersection Frankfurt North-West as a result of an accident at kilometer 487.5 at 17:13. The downstream front of this HCT pattern is pinned between cross-sections N7 and N8, and the outflow of the accident site is just  $Q_{\text{out}}^{\text{hindered}} = 873$  veh/h/lane, so that we find free traffic flow downstream. If we compare this value with the normal outflow of  $Q_{\text{out}} \approx 1700$  veh/h/lane in the first minutes after the accident has been cleared, the estimated bottleneck strength is  $\Delta Q \approx 827$  veh/h/lane, which is consistent with the phase diagram (see Figure 15a). After the accident has been cleared, the downstream congestion front moves upstream until it reaches the next bottlenecks at the Intersections Frankfurt North-West and

Frankfurt West. These are less severe due to existing off-ramps, and cause OCT. At the same time, the increased outflow causes a congestion pattern for 45 minutes at the uphill gradient after Intersection Bad Homburg, where traffic was already congested before the accident.

## 7. Summary, Discussion, and Conclusions

In this contribution, we critically discussed the arguments in favor of and against first- and second-order traffic models. In particular, we pointed out that it is possible to formulate a macroscopic traffic model with a dynamical velocity equation that is theoretically consistent—i.e., overcomes the criticisms raised in Daganzo's Requiem (Daganzo 1995b) and other publications. Such a model is the nonlocal GKT model (Treiber, Hennecke, and Helbing 1999), which implies the existence of several different kinds of congested traffic states. The conditions for the possible occurrence of these states have been illustrated by a phase diagram as a function of the bottleneck strength and the upstream traffic flow. One important point of this phase diagram is the multistable nature of traffic flows. That is, several different traffic states can exist under similar conditions. Their occurrence depends on the initial and boundary conditions, particularly on whether there occurs a large perturbation or not (compare Figures 4a and 4b). Similar predictions have been made by microscopic traffic models with stable, metastable, and linearly unstable traffic regimes (Treiber, Hennecke, and Helbing 2000; Helbing et al. 2002). However, a traffic model without linearly unstable flow regimes, but stable and metastable density regimes only, would imply the same congestion patterns.<sup>12</sup> Furthermore, the validity of the phase diagram is not restricted to a particular traffic model. The phase diagram approach can also be used without linking it to *any* model.

In order to test the predictions of the LWR model and the phase diagram, and in order to assess Kerner's findings, we have carried out an extensive study of empirical traffic states observed at a 30-kilometer-long section of the German freeway A5 near Frankfurt. Despite two freeway intersections, there were approximately 10-kilometer-long freeway sections without any on- or off-ramps. To obtain a more detailed picture of the spatiotemporal dynamics of traffic flows, we have used an adaptive smoothing algorithm that interpolates and smooths traffic data from successive cross-sections of the freeway, taking

into account the propagation speeds of perturbations in free and congested traffic. The method is based on measurements of average vehicle speeds by double-loop detectors, which are quite reliable. Our inverted speed representation is reminiscent of density plots, which are easy to interpret.

Along the freeway section we studied, we have found a rich variety of congested traffic states, but the great majority of them could, in fact, be interpreted as a spatial combination of, altogether, six different, elementary traffic states: free traffic (FT), pinned localized clusters (PLC), moving localized clusters (MLC), stop-and-go waves (SGW), oscillating congested traffic (OCT), and homogeneous congested traffic (HCT); see §§6.1 to 6.5. The same states have been found at other freeways that we studied (Treiber, Hennecke, and Helbing 2000; Helbing et al. 2002). A detailed understanding of complex traffic patterns consisting of several elementary congestion states is made possible by considering combinations of several bottlenecks, such as on- and off-ramps (see §§6.7 to 6.9).

### 7.1. Facing Kerner's Criticisms

Because Kerner has generally criticized all models with a fundamental diagram (i.e., most traffic models ever developed in the history of traffic science), and because he has specifically questioned some features of the GKT model, let us address these points in the following:

1. According to Kerner (1998b), models containing a fundamental diagram could not explain the wide scattering of flow-density data observed for synchronized congested traffic flow.
2. The phase diagram approach would predict the transitions free traffic → moving or pinned localized cluster → stop-and-go traffic or oscillating congested traffic → homogeneous congested traffic as the ramp flow or the overall traffic volume  $Q_{\text{tot}}$  increases. However, homogeneous congested traffic would not exist (Kerner 2002a, b), and the actual transition to moving jams would always require an advance transition to synchronized flow.

Let us comment on these points one by one in the same order:

1. Kerner's criticism is based on a comparison of his empirical data (usually one-minute aggregates) with simulation results that do not simulate the measurement (aggregation) process and do not contain effects of different vehicle classes, different driver behaviors, and lane changes. However, it has been shown that all three aspects produce a scattering of flow-density data in the congested regime (Treiber and Helbing 1999; Nishinari, Treiber, and Helbing 2003). Quite realistic fluctuations of "synchronized" congested flow can already be obtained by simulating a mixture of two vehicle classes, each obeying

<sup>12</sup> The main difference would be that OCT would not be expected for  $\Delta Q < Q_{\text{out}} - Q_{\text{c3}}$ . According to Figure 15b, this would be compatible with  $Q_{\text{out}} - Q_{\text{c3}} \approx 400 \text{ veh/h/lane}$ .

a macroscopic traffic model with a fundamental diagram (Treiber and Helbing 1999). Moreover, it has been found that the wide scattering of flow-density data in synchronized congestion states can be very well reproduced by the formula

$$Q(t) = \frac{1}{T(t)} \left( 1 - \frac{\rho(t)}{\rho_{\text{jam}}(t)} \right) \quad (31)$$

for the jam line (Kerner and Konhäuser 1994b; Kerner and Rehborn 1996b) if not only the vehicle density  $\rho(t)$  is varied in time, but also the time gap  $T(t)$  and the jam density  $\rho_{\text{max}}(t)$  (Nishinari, Treiber, and Helbing 2003), as demanded by Banks (1999). These vary naturally due to fluctuating truck fractions and driver behaviors (Treiber and Helbing 1999).

2. HCT clearly exists (see Figure 10), but it occurs very rarely and only for extremely large bottleneck strengths exceeding  $\Delta Q \approx 700$  vehicles/h/lane. Because freeways are dimensioned in a way to avoid bottlenecks of this size, HCT occurs primarily when freeway lanes are closed after a serious accident. In other words, when excluding cases of accidents from the data set, HCT states will normally not be found. Moreover, Kerner is wrong in claiming that our theoretical phase diagram would require MLCs to occur before the transition to SGWs or OCT. This misunderstanding might have occurred by ignoring the dependence of the traffic state predicted by the theoretical phase diagram on the perturbation size. Figure 4a clearly shows that a direct transition from FT flow to OCT is predicted in cases of small perturbations.

In summary, according to our analysis, Kerner has rejected other theories for wrong reasons, namely because alternative explanations of empirical facts (and the data themselves) were not checked carefully enough. Therefore, we currently do not see any reasons to replace well-justified macroscopic traffic models such as the nonlocal GKT model by more complicated models trying to explain the same facts with a higher number of parameters.

## 7.2. Criticism of the Three-Phase Theory and Its Extensions

**7.2.1. Three-Phase Theory and “Moving Synchronized Patterns.”** Kerner’s three phase-theory is a well-known attempt at a systematic classification of different traffic states (Kerner and Rehborn 1996b). This classification is phenomenological and distinguishes free traffic, synchronized flow, and wide moving jams (Kerner et al. 2004). In contrast to synchronized flow, wide moving jams are localized. They have a characteristic outflow and a constant downstream propagation speed (Kerner and Rehborn 1996a). Synchronized flow has further been subdivided into three different kinds (Kerner and

Rehborn 1996b): (i) *stationary and homogeneous states*, where both the average speed and the flow rate are stationary during a relatively long time interval, (ii) *“homogeneous-in-speed states,”* see also Kerner (1998b), and (iii) *nonstationary and nonhomogeneous states*, see also Kerner (1998b).

Comparing Kerner’s classification with ours, homogeneous congested traffic (HCT) may correspond to “synchronized” traffic flow of type (i), whereas oscillating congested traffic (OCT) seems to relate to synchronized flow of type (iii). Moreover, homogeneous-in-speed states resemble the free branch of the fundamental diagram, but with a reduced free velocity. Therefore, this synchronized flow of type (ii) bears features of both free and congested traffic. This may reflect the influence of the “intelligent speed control” operated at the studied freeway stretch, or it may point to recovering traffic downstream of bottlenecks (see Treiber and Helbing 1999, Figure 4; Tilch and Helbing 2000, Figure 3). Finally, Kerner’s wide moving jams seem to be comparable to moving localized clusters (MLC). A major technical difference between the classification of Kerner and ourselves seems to be that our classification is based on an analysis of the spatiotemporal traffic dynamics, while Kerner’s is based on the dynamics at single freeway cross-sections or comparisons of the dynamics at subsequent sections. Moreover, our phase diagram defines quantitative existence criteria, whereas the three-phase theory stays qualitative.

Although Kerner claims that the three-phase theory is the only theory that can describe the empirical phenomena accurately, he keeps reporting new spatiotemporal traffic patterns. The general pattern and the pinch effect will be critically discussed in the next section. Recently, Kerner has introduced the notion of a “moving synchronized pattern” (MSP) to resolve inconsistencies with his three-phase theory. The reason for this is Kerner’s microsimulation of localized moving jam structures that are stopping at a bottleneck, as we have discovered them in our empirical data (see Figures 9b, 20a, and 21a). These localized patterns can obviously not be called wide moving jams, because these should propagate through all traffic states.

MSPs travel with the same speed as wide moving jams, however, “...in contrast with a wide moving jam, if a MSP reaches a bottleneck, the MSP is caught at the bottleneck: The MSP that propagates upstream can exist only for a finite time” (Kerner 2005, p. 187). That is, the actual type of traffic congestion remains undefined until it reaches the next upstream bottleneck. It is certainly not scientific to distinguish two apparently identical states that cannot be told apart, at least if there is no upstream bottleneck at all. Even if there is a bottleneck allowing one to differentiate

between an MSP and a wide moving jam, the “catching” of the jam and thus the identity of the traffic phenomenon will depend on the bottleneck strength, which leaves the classification unsatisfactory. Thus, a clear identification of a wide moving jam in contrast to synchronized traffic is no longer possible, and the three-phase theory is severely questioned by this shortcoming.

**7.2.2. Criticism of the Pinch Effect and Generalized Pattern Generating Wide Moving Jams.** In his three-phase theory, Kerner also states that the occurrence of wide moving jams would necessarily require the previous transition from free to synchronized flow (Kerner 1998b; 2000a, b) and that the generation mechanism could be imagined as follows: In the “general pattern,” synchronized traffic upstream of a bottleneck would breed wide moving jams based on the pinch effect. That is, upstream of a section with synchronized congested traffic close to a bottleneck, a so-called *pinch region* should spontaneously give birth to narrow vehicle clusters. These perturbations should be growing while travelling further upstream. Eventually, wide moving jams should be formed by the merging or disappearance of narrow jams. Once formed, wide jams should suppress the occurrence of new narrow jams in between. Instead of forming wide jams, however, narrow jams would be able to coexist if their distance was larger than about 2.5 km (Kerner 1998a).

In our empirical analysis, we have found only a few examples of Kerner’s pinch effect, i.e., for the merging or dissolution of small narrow jams (see Figure 20). This may be a consequence of averaging out small perturbations by the adaptive smoothing method, but we think that such small perturbations are subject to interpretation and should not constitute the basis of a theory. We did, however, find congestion patterns resembling Kerner’s general pattern. However, we do not interpret them as congestion patterns that would *generally* occur at an isolated bottleneck far away from other bottlenecks or inhomogeneities, as Kerner (2002b) suggests. Kerner (2002b, p. 9) claims that “the effective bottleneck at D6 can indeed be considered as an isolated bottleneck at the on ramp.” We doubt this. According to our analysis, a general pattern instead occurs because of the complex interaction between different on- and off-ramps (and, additionally, with long ramp lanes). In particular, we interpret a general pattern as a combination of an extended form of congested traffic between an off-ramp and a subsequent on-ramp, with SGWs upstream of the off-ramp, which must be treated as a second, weaker bottleneck (see §6.7). In our data, the pinch effect and general patterns did not clearly occur at flow-conserving bottlenecks such as gradients or locations

of accidents, although some simulation models indicate the possibility of this (Helbing et al. 2002).

Moreover, we point out that the SGW and wide moving jams (MLC states) depicted in Figures 12 and 13 are clearly not emanating from synchronized flows or general patterns. We have instead identified two other mechanisms that generate moving localized clusters:

1. SGWs like the ones depicted in Figure 12 and single MLCs, as shown in Figure 13a seem to appear spontaneously at a gradient due to unstable traffic flows.

2. The MLC patterns in Figures 9b, 13b, 20b, and 21a have been initiated by the boomerang effect, which is based on a significant downstream travelling perturbation. The boomerang effect was the origin of at least 18 out of 245 cases of traffic breakdowns.

### 7.3. Conclusions

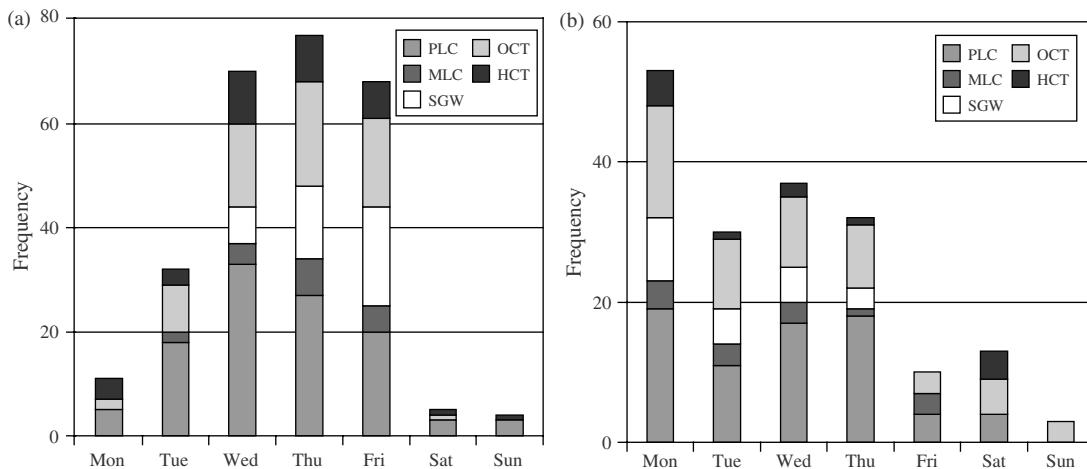
Applying the adaptive smoothing method, which can visualize small perturbations and distinguish them from fluctuations by avoiding that continuously moving patterns are split up into artificial, discontinuously looking structures (see Figure 8), we obtained a more detailed picture of the spatiotemporal dynamics of traffic flows. We identified growing perturbations on freeway stretches without ramps or changes in the number of lanes. This questions the validity of the LWR model and supports traffic models with an unstable traffic regime. The LWR model also does not provide a good explanation for the different observed congestion states.

We have shown that the typical traffic pattern depends on the upstream freeway flow  $Q_{up}$  and the bottleneck strength  $\Delta Q$ .<sup>13</sup> Therefore, the frequency of different traffic patterns also depends on the weekday (see Figure 22). However, there is usually no unique relationship between the flow values ( $\Delta Q, Q_{up}$ ) and the type of congestion (see Figure 15b). Due to multistability, initial and boundary conditions are also relevant (see Figure 4). Therefore, many traffic breakdowns are triggered by perturbations in the traffic flow, e.g., overtaking maneuvers of trucks (see Figure 2b).

In our study, less than 5% of the congested traffic states remained unexplained, considering the bottlenecks along the freeway known to us.<sup>14</sup> However, most of them could be related to accidents based

<sup>13</sup> The bottlenecks that we identified along the freeway were related to freeway on- or off-ramps, gradients, accidents, or reductions in the number of lanes. However, they could have other origins as well, like speed limits or road work, accidents in the opposite lanes (“rubberneck effect”), bad road conditions (possibly due to rain, fog, or ice), bad visibility (e.g., because of blinding sun or tunnel entrances), distracting views, or curves.

<sup>14</sup> Note that we did not have a protocol of road works performed.



**Figure 22** Absolute Frequencies of Congested Traffic States on the German Freeway A5 Close to Frankfurt (a) in Direction North and (b) in Direction South

**Notes.** One can see a typical dependence of the absolute frequency of congestion and the relative frequency of different kinds of congestion on weekdays. On Saturdays and Sundays, congestion occurred less frequently compared to work days, as expected. Moreover, in direction North, congestion patterns appeared rarely on Mondays compared to Tuesdays, Wednesdays, or Thursdays, while in direction South, the frequency of traffic breakdowns was surprisingly low on Fridays. This could be explained by commuters spending their weekends in the North and living in Frankfurt during the week.

on police reports, or they were cases of forwardly moving “phantom” bottlenecks (Gazis and Herman 1992; Muñoz and Daganzo 2002b), among them two congestion patterns moving forward at speeds of 1.6 km/h and 2.7 km/h (not shown). The most frequent states at the investigated freeway section were PLC and OCT states (see Figure 22), while large HCT states occurred mainly after serious accidents with lane closures.<sup>15</sup> The downstream fronts of these congestion patterns were located at bottlenecks, while for MLC and SGW states, they mostly propagated upstream at a constant speed  $c_0 \approx -16 \pm 2$  km/h. However, we also found changes in the propagation speed that seem to depend on the gradient (see Figure 20b).

The explanation of complex congestion patterns requires us to consider the combination of several bottlenecks. For example, on-ramp bottlenecks are frequently combined with a closeby upstream off-ramp. In such cases, the upstream front of extended forms of congested traffic (such as HCT or OCT) is often fixed at the location of the off-ramp (Treiber, Hennecke, and Helbing 2000; Lee, Lee, and Kim 2000). Moreover, the congestion type typically changes to a milder form upstream, for example, SGWs. Then, the overall appearance resembles of the pinch effect or general pattern (Kerner 1998a, 2002b). However, SGWs may also be generated by the boomerang effect or appear spontaneously at a bottleneck (see Figures 9b, 12, 13, 20b, and 21a.) in contrast to Kerner’s three phase theory (Kerner et al. 1998a).

In summary, we have found that the congestion patterns observed along a section of the freeway A5

close to Frankfurt/Main, Germany, were very consistent with the predictions of the phase diagram of traffic states. This phase diagram was motivated by properties of the nonlocal GKT model, which we have shown to avoid the inconsistencies of second-order traffic models (see §4), and which qualitatively agrees with phase diagrams obtained for a wide class of car-following models. Moreover, we did not find Kerner’s criticism of models with a fundamental diagram and the GKT model convincing. Therefore, it is not necessary to discard all previous knowledge accumulated in traffic modeling as suggested by Kerner. The phase diagram of traffic states instead offers a simple, straightforward, and consistent explanation of a wide range of empirical observations in the tradition of classical traffic flow modeling.

### Acknowledgments

The authors would like to thank the Volkswagen AG within the BMBF research initiative INVENT for financial support, the *Hessisches Landesamt für Straßen und Verkehrswesen* for providing the freeway data, and Martin Treiber for valuable discussions.

### References

- Banks, J. H. 1991a. Two-capacity phenomenon at freeway bottlenecks: A basis for ramp meeting? *Transportation Res. Record* **1320** 83–90.
- Banks, J. H. 1991b. The two-capacity phenomenon: Some theoretical issues. *Transportation Res. Record* **1320** 234–241.
- Banks, J. H. 1999. An investigation of some characteristics of congested flow. *Transportation Res. Record* **1678** 128–134.
- Bertini, R. L., R. Lindgren, D. Helbing, M. Schönhof. 2004. Empirical analysis of flow features on a German autobahn. *Proc. 83rd Annual Meeting Transportation Res. Board*. Washington, D.C.

<sup>15</sup>Otherwise, the freeway would not be appropriately dimensioned.

- Cassidy, M. J., R. L. Bertini. 1999. Some traffic features at freeway bottlenecks. *Transportation Res. Part B* **33** 25–42.
- Cassidy, M. J., M. Mauch. 2001. An observed traffic pattern in long freeway queues. *Transportation Res. Part A* **35** 143–156.
- Cassidy, M. J., J. R. Windover. 1995. Methodology for assessing dynamics of freeway traffic flow. *Transportation Res. Record* **1484** 73–79.
- Coifman, B. 2002. Estimating travel times and vehicle trajectories in freeways using dual loop detectors. *Transportation Res. Part A* **36** 351–364.
- Cremer, M. 1979. *Der Verkehrsfluss auf Schnellstraßen. Modelle, Überwachung, Regelung.* Springer, Berlin, Germany.
- Daganzo, C. F. 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Res. Part B* **28** 269–287.
- Daganzo, C. F. 1995a. A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Res. Part B* **29** 261–276.
- Daganzo, C. F. 1995b. Requiem for second-order fluid approximations of traffic flow. *Transportation Res. Part B* **29** 277–286.
- Daganzo, C. F. 2002a. A behavioral theory of multi-lane traffic flow. I: Long homogeneous freeway sections. *Transportation Res. Part B* **36** 131–158.
- Daganzo, C. F. 2002b. A behavioral theory of multi-lane traffic flow. II: Merges and the onset of congestion. *Transportation Res. Part B* **36** 159–169.
- Daganzo, C. F., M. J. Cassidy, R. L. Bertini. 1999. Possible explanations of phase transitions in highway traffic. *Transportation Res. Part A* **33** 365–379.
- Edie, L. C., E. Baverez. 1965. Generation and propagation of stop-start traffic waves. L. Edie, R. Herman, R. Rothery, eds. *Vehicular Traffic Science, Proc. 3rd Internat. Sympos. Theory Traffic Flow*, New York, 26–37.
- Edie, L. C., R. S. Foote. 1960. Effect of shock waves on tunnel traffic flow. *Proc. Highway Res. Board*, Vol. 37. Transportation Research Board of the National Academies, 492–505.
- Gazis, P. C., R. Herman. 1992. The moving and “phantom” bottlenecks. *Transportation Sci.* **26** 223–229.
- Helbing, D. 1996. Gas-kinetic derivation of navier-stokes-like traffic equations. *Phys. Rev. E* **53** 2366–2381.
- Helbing, D. 1997a. Fundamentals of traffic flow. *Phys. Rev. E* **55** 3735–3738.
- Helbing, D. 1997b. Modelling multi-lane traffic flow with queuing effects. *Physica A* **242** 175–194.
- Helbing, D. 1997c. Traffic data and their implications for consistent traffic flow modeling. M. Papageorgiou, A. Pouliozos, eds. *Transportation Systems*. International Federation of Automatic Control, Chania, Greece, 809–814.
- Helbing, D. 1997d. *Verkehrsdynamik*. Springer, Berlin, Germany.
- Helbing, D. 1998. Structure and instability of high-density equations for traffic flow. *Phys. Rev. E* **57** 6176–6179.
- Helbing, D. 2001a. Die wundervolle Welt aktiver Vielteilchensysteme. *Physikalische Blätter* **57** 27–33.
- Helbing, D. 2001b. Traffic and related self-driven many-particle systems. *Rev. Modern Phys.* **73** 1067–1141.
- Helbing, D. 2003. A section-based queueing-theoretical traffic model for congestion and travel time analysis in networks. *J. Phys. A: Math. General* **36** L593–L598.
- Helbing, D., A. Greiner. 1997. Modelling and simulation of multi-lane traffic flow. *Phys. Rev. E* **55** 5498–5507.
- Helbing, D., M. Treiber. 1999. Numerical simulation of macroscopic traffic equations. *Comput. Sci. Engrg. (CiSE)* **5** 89–99.
- Helbing, D., A. Hennecke, M. Treiber. 1999. Phase diagram of traffic states in the presence of inhomogeneities. *Phys. Rev. Lett.* **82** 4360–4363.
- Helbing, D., A. Hennecke, V. Shvetsov, M. Treiber. 2001. MASTER: Macroscopic traffic simulation based on a gas-kinetic, non-local traffic model. *Transportation Res. Part B* **35** 183–211.
- Helbing, D., A. Hennecke, V. Shvetsov, M. Treiber. 2002. Micro- and macrosimulation of freeway traffic. *Math. Comput. Model.* **35** 517–547.
- Helbing, D., I. J. Farkas, D. Fasold, M. Treiber, T. Vicsek. 2003. Critical discussion of “synchronized flow,” pedestrian evacuation, and optimal production. M. Fukui, Y. Sugiyama, M. Schreckenberg, D. E. Wolf, eds. *Proc. Traffic Granular Flow '01*. Springer, Berlin, Germany, 511–530.
- Helbing, D., A. Johansson, J. Mathiesen, M. H. Jensen, A. Hansen. 2006. Analytical approach to continuous and intermittent bottleneck flows. *Physical Rev. Lett.* **97** 1–4.
- Hoogendoorn, S., P. H. L. Bovy. 1998. *Continuum Modelling of Multi-Lane Heterogeneous Traffic Flow Operations*. TU Delft, Delft, The Netherlands.
- Kerner, B. S. 1998a. Experimental features of self-organization in traffic flow. *Phys. Rev. Lett.* **81** 3797–3800.
- Kerner, B. S. 1998b. A theory of congested traffic flow. R. Rysgaard, ed. *Proc. 3rd Internat. Sympos. Highway Capacity*, Vol. 2. Road Directorate, Copenhagen, Denmark, 621–642.
- Kerner, B. S. 1999. Theory of congested traffic flow: Self-organization without bottlenecks. A. Ceder, ed. *Proc. 14th Internat. Sympos. Transportation Traffic Theory*. Pergamon, New York, 147–171.
- Kerner, B. S. 2000a. Phase transitions in traffic flow. D. Helbing, H. J. Herrmann, M. Schreckenberg, D. E. Wolf, eds. *Proc. Traffic Granular Flow '99: Social, Traffic and Granular Dynamics*. Springer, Berlin, Germany, 253–284.
- Kerner, B. S. 2000b. Theory of breakdown phenomenon at highway bottlenecks. *Transportation Res. Record* **1710** 136–144.
- Kerner, B. S. 2001. Verfahren zur Verkehrszustandsüberwachung für ein Verkehrsnetz mit effektiven Engstellen. German Patent DE 199 44 075 C2.
- Kerner, B. S. 2002a. Empirical features of congested patterns at highway bottlenecks. *Transportation Res. Record* **1802** 145–154.
- Kerner, B. S. 2002b. Empirical macroscopic features of spatiotemporal traffic patterns at highway bottlenecks. *Phys. Rev. E* **65** 1–30.
- Kerner, B. S. 2004. *The Physics of Traffic*. Springer, Heidelberg, Germany.
- Kerner, B. S. 2005. Microscopic three-phase traffic theory and its applications for freeway traffic control. H. S. Mahmassani, ed. *Proc. 16th Internat. Sympos. Transportation Traffic Theory*, Elsevier, Amsterdam, The Netherlands, 181–203.
- Kerner, B. S., P. Konhäuser. 1993. Cluster effect in initially homogeneous traffic flow. *Phys. Rev. E* **48** R2335–R2338.
- Kerner, B. S., P. Konhäuser. 1994a. Structure and parameters of clusters in traffic flow. *Phys. Rev. E* **50** 54–83.
- Kerner, B. S., P. Konhäuser. 1994b. Structure and parameters of clusters in traffic flow. *Physical Rev. E* **50** 54–83.
- Kerner, B. S., H. Rehborn. 1996a. Experimental features and characteristics of traffic jams. *Phys. Rev. E* **53** R1297–R1300.
- Kerner, B. S., H. Rehborn. 1996b. Experimental properties of complexity in traffic flow. *Phys. Rev. E* **53** R4275–R4278.
- Kerner, B. S., H. Rehborn. 1997. Experimental properties of phase transitions in traffic flow. *Phys. Rev. Lett.* **79** 4030–4033.
- Kerner, B. S., H. Rehborn. 1998. Über zwei Arten von Phasenübergängen im Verkehrsfluss und den daraus resultierenden drei Arten von Kapazitäten im Verkehr. *Internationales Verkehrswesen* **5** 347–348.
- Kerner, B. S., H. Rehborn. 2000. Verfahren zur Verkehrszustandsüberwachung und Fahrzeugzuflusssteuerung in einem Straßenverkehrsnetz. German Patent DE 198 35 979 A1.

- Kerner, B. S., M. Aleksic, M. Denneler. 2001. Verfahren und Vorrichtung zur Verkehrszustandsüberwachung. German Patent DE 199 44 077 C2.
- Kerner, B. S., H. Kirschfink, H. Rehborn. 1998. Verfahren zur automatischen Verkehrsüberwachung mit Staudynamikanalyse. German Patent DE 196 47 127 C2.
- Kerner, B. S., S. L. Klenov, P. Konhäuser. 1997. Asymptotic theory of traffic jams. *Phys. Rev. E* **56** 4200–4216.
- Kerner, B. S., H. Rehborn, M. Aleksic, A. Haug. 2004. Recognition and tracking of spatial-temporal congested traffic patterns on freeways. *Transportation Res. Part C* **12** 369–400.
- Klar, A., R. Wegener. 1999a. A hierarchy of models for multilane traffic. I: Modeling. *SIAM J. Appl. Math.* **59** 983–1001.
- Klar, A., R. Wegener. 1999b. A hierarchy of models for multilane traffic. II: Numerical investigations. *SIAM J. Appl. Math.* **59** 1002–1011.
- Koshi, M., M. Iwasaki, I. Ohkura. 1983. Some findings and an overview on vehicular flow characteristics. V. F. Hurdle, E. Hauer, G. N. Stewart, eds. *Proc. 8th Internat. Sympos. Transportation Traffic Theory*, University of Toronto Press, Toronto, Ontario, 403–426.
- Kronjäger, W., P. Konhäuser. 1997. Applied traffic flow simulation. M. Papageorgiou, A. Pouliozos, eds. *Transportation Systems*. Vol. II. International Federation of Automatic Controlm Chania, Greece, 805–808.
- Kühne, R. D. 1984. Macroscopic freeway model for dense traffic—stop-start waves and incident detection. I. Volmuller, R. Hamerslag, eds. *Proc. 9th Internat. Sympos. Transportation Traffic Theory*. VNU Science Press, Utrecht, The Netherlands, 21–42.
- Kühne, R. D. 1987. Freeway speed distribution and acceleration noise—calculations from a stochastic continuum theory and comparison with measurements. N. H. Gartner, N. H. M. Wilson, eds. *Proc. 10th Internat. Sympos. Transportation Traffic Theory*. Elsevier, New York, 119–137.
- Laval, J. A., C. F. Daganzo. 2004. Multi-lane hybrid traffic flow model: Quantifying the impacts of lane-changing maneuvers on traffic flow. Working Paper UCB-ITS-WP-2004-1, Institute of Transportation Studies, Berkeley, CA.
- Laval, J. A., C. F. Daganzo. 2006. Lane-changing in traffic streams. *Transportation Res. Part B* **40** 251–264.
- Lebacque, J. P. 1996. The Godunov scheme and what it means for first order traffic flow models. J. B. Lesort, ed. *Transportation and Traffic Theory. Proc. 13th Internat. Sympos. Transportation Traffic Theory*, Pergamon, Oxford, 647–677.
- Lebacque, J.-P., M. M. Khoshyaran. 2005. First-order macroscopic traffic flow models: Intersection modeling, network modeling. H. S. Mahmassani, ed. *Transportation and Traffic Theory—Flow, Dynamics and Human Interaction*, Proc. 16th Internat. Sympos. Transportation Traffic Theory, Elsevier, Amsterdam, The Netherlands, 365–386.
- Lee, H., H. Lee, D. Kim. 1999. Dynamic states of a continuum traffic equation with on-ramp. *Phys. Rev. E* **59** 5101–5111.
- Lee, H., H. Lee, D. Kim. 2000. Phase diagram of congested traffic flow: An empirical study. *Phys. Rev. E* **62** 4737–4741.
- Lighthill, M. J., G. B. Whitham. 1955. On kinematic waves: II. A theory of traffic on long crowded roads. *Proc. Roy. Soc. London A* **229** 317–345.
- Lin, W.-H., H. K. Lo. 2003. A theoretical probe of a German experiment on stationary moving traffic jams. *Transportation Res. Part B* **37** 251–261.
- Mauch, M., M. J. Cassidy. 2002. Freeway traffic oscillations: Observations and predictions. M. A. P. Taylor, ed. *Proc. 15th Internat. Sympos. Traffic Transportation Theory*, Elsevier, Amsterdam, The Netherlands, 653–674.
- Mika, H. S., J. B. Kreer, L. S. Yuan. 1969. Dual mode behavior of freeway traffic. *Highway Res. Record* **279** 1–13.
- Muñoz, J. C., C. F. Daganzo. 1999. Experimental observations upstream of an active bottleneck caused by a freeway off-ramp. PATH working paper, California Partners for Advanced Transit and Highways, Richmond, CA.
- Muñoz, J. C., C. F. Daganzo. 2002a. The bottleneck mechanism of a freeway diverge. *Transportation Res. Part A* **36** 483–505.
- Muñoz, J. C., C. F. Daganzo. 2002b. Fingerprinting traffic from static freeway sensors. *Cooper@tive Tr@nsport@ion Dyn@mics* **1** 1.1–1.11.
- Muñoz, J. C., C. F. Daganzo. 2003. Structure of the transition zone behind freeway queues. *Transportation Sci.* **37** 312–329.
- Nagel, K., P. Nelson. 2005. A critical comparison of the kinematic-wave model with observational data. H. S. Mahmassani, ed. *Proc. 16th Internat. Sympos. Transportation Traffic Flow Theory*, Elsevier, Amsterdam, The Netherlands, 145–163.
- Newell, G. F. 1982. *Applications of Queueing Theory*. Chapman and Hall, Cambridge, UK.
- Nishinari, K., M. Treiber, D. Helbing. 2003. Interpreting the wide scattering of synchronized traffic data by time gap statistics. *Phys. Rev. E* **68** 1–4.
- Papageorgiou, M. 1983. *Applications of Automatic Control Concepts to Traffic Flow Modelling and Control*. Springer, Berlin, Germany.
- Paveri-Fontana, S. L. 1975. On Boltzmann-like treatments for traffic flow. A critical review of the basic model and an alternative proposal for dilute traffic analysis. *Transportation Res.* **9** 225–235.
- Payne, H. 1971. Models of freeway traffic and control. G. A. Bekey, ed. *Mathematical Models of Public Systems*, Vol. 1. Simulation Council, La Jolla, CA, 51–61.
- Payne, H. 1979. FREFLO: A macroscopic simulation model of freeway traffic. *Transportation Res. Record* **722** 68–77.
- Persaud, B. N., V. F. Hurdle. 1991. Freeway capacity: Definition and measurement issues. U. Brannolte, ed. *Proc. 3rd Internat. Sympos. Highway Capacity Level Service*. A. A. Balkema Press, Rotterdam, The Netherlands, 289–307.
- Persaud, B. N., S. Yagar, R. Brownlee. 1998. Exploration of the breakdown phenomenon in freeway traffic. *Transportation Res. Record* **1634** 64–69.
- Phillips, W. F. 1977. Kinetic model for traffic flow. Tech. Rep. DOT/RSPD/DPB/50-77/17, National Technical Information Service, Springfield, VA 22161.
- Phillips, W. F. 1979a. A kinetic model for traffic flow with continuum implications. *Transportation Planning Tech.* **5** 131–138.
- Phillips, W. F. 1979b. A new continuum traffic model obtained from kinetic theory. *Proc. 1978 IEEE Conf. Decision Control*. IEEE, New York, 1032–1036.
- Prigogine, I. 1961. A Boltzmann-like approach to the statistical theory of traffic flow. R. Herman, ed. *Sympos. Theory Traffic Flow 1959*, Warren, MI, Elsevier, New York, 148–164.
- Richards, P. I. 1956. Shock waves on the highway. *Oper. Res.* **4** 42–51.
- Schönhof, M., D. Helbing. 2004. Empirical features of congested traffic states and their implications for traffic modeling. *cond-mat/0408138*.
- Shvetsov, V., D. Helbing. 1999. Macroscopic dynamics of multi-lane traffic. *Phys. Rev. E* **59** 6328–6339.
- Sugiyama, Y., A. Nakayama, M. Fukui, K. Hasebe, M. Kikuchi, K. Nishinari, S. I. Tadaki, S. Yukawa. 2005. Observation, theory and experiment for freeway traffic as physics of many-body system. S. P. Hoogendoorn, S. Luding, P. H. L. Bovy, M. Schreckenberg, D. E. Wolf, eds. *Proc. Traffic Granular Flow '03*, Springer, Berlin, Germany, 45–58.
- Tilch, B., D. Helbing. 2000. Evaluation of single vehicle data in dependence of the vehicle-type, lane, and site. D. Helbing, H. J. Herrmann, M. Schreckenberg, D. E. Wolf, eds. *Proc. Traffic Granular Flow '99*, Springer, Berlin, Germany, 333–338.
- Treiber, M., D. Helbing. 1999. Macroscopic simulation of widely scattered synchronized traffic states. *J. Phys. A: Math. General* **32** L17–L32.

- Treiber, M., D. Helbing. 2002. Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooperative Transportation Dyn@mics* 1 3.1–3.24.
- Treiber, M., A. Hennecke, D. Helbing. 1999. Derivation, properties, and simulation of a gas-kinetic-based, non-local traffic model. *Phys. Rev. E* **59** 239–253.
- Treiber, M., A. Hennecke, D. Helbing. 2000. Congested traffic states in empirical observations and microscopic simulations. *Phys. Rev. E* **62** 1805–1824.
- Treiber, M., A. Kesting, D. Helbing. 2006. Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps. *Phys. Rev. E* **74**.
- Treiterer, J., J. A. Myers. 1974. The hysteresis phenomenon in traffic flow. D. Buckley, ed. *Proc. 6th Internat. Sympos. Transportation Traffic Theory*, Reed, London, UK, 13–38.
- Treiterer, J., J. I. Taylor. 1966. Traffic flow investigations by photogrammetric techniques. *Highway Res. Record* **142** 1–12.
- Wagner, C., C. Hoffmann, R. Sollacher, J. Wagenhuber, B. Schürmann. 1996. Second-order continuum traffic flow model. *Phys. Rev. E* **54** 5073–5085.
- Westland, D. 1998. Effect of highway geometry on freeway queuing at merge sections. R. Rysgaard, ed. *Proc. 3rd Internat. Sympos. Highway Capacity*, Vol. 2. Road Directorate, Denmark, 1095–1116.
- Whitham, G. 1974. *Linear and Nonlinear Waves*. Wiley, New York.
- Windover, J. R. 1998. Empirical studies of the dynamic features of freeway traffic. Ph.D. thesis, University of California, Berkeley, CA.
- Yi, J., H. Lin, L. Alvarez, R. Horowitz. 2003. Stability of macroscopic traffic flow modeling through wavefront expansion. *Transportation Res. Part B* **37** 661–679.