INVITED PAPER

STILL FLOWING: APPROACHES TO TRAFFIC FLOW AND TRAFFIC JAM MODELING

KAI NAGEL

Swiss Federal Institute of Technology (ETH), Department of Computer Science, ETH Zentrum, 8092 Zürich, Switzerland, nagel@inf.ethz.ch

PETER WAGNER and RICHARD WOESLER

German Aerospace Center (DLR), Institute of Transportation Research, Rutherfordstrasse 2, 12489 Berlin, Germany richard.woesler@dlr.de • peter.wagner@dlr.de

Certain aspects of traffic flow measurements imply the existence of a phase transition. Models known from chaos and fractals, such as nonlinear analysis of coupled differential equations, cellular automata, or coupled maps, can generate behavior which indeed resembles a phase transition in the flow behavior. Other measurements point out that the same behavior could be generated by geometrical constraints of the scenario. This paper looks at some of the empirical evidence, but mostly focuses on different modeling approaches. The theory of traffic jam dynamics is reviewed in some detail, starting from the well-established theory of kinematic waves and then veering into the area of phase transitions. One aspect of the theory of phase transitions is that, by changing one single parameter, a system can be moved from displaying a phase transition to not displaying a phase transition. This implies that models for traffic can be tuned so that they display a phase transition or not.

This paper focuses on microscopic modeling, i.e., coupled differential equations, cellular automata, and coupled maps. The phase transition behavior of these models, as far as it is known, is discussed. Similarly, fluid-dynamical models for the same questions are considered. A large portion of this paper is given to the discussion of extensions and open questions, which makes clear that the question of traffic jam dynamics is, albeit important, only a small part of an interesting and vibrant field. As our outlook shows, the whole field is moving away from a rather static view of traffic toward a dynamic view, which uses simulation as an important tool.

Received March 2001; revisions received July 2002, January 2003; accepted January 2003.
Subject classifications: Transportation, models, traffic. Simulation: traffic flow models.
Area of review: Invited Papers.

1. INTRODUCTION

Driving is simple, isn't it? Drivers have two goals: (i) to arrive quickly, (ii) without crashing into another car. Both goals seem to be met by drivers with sufficient satisfaction and an astonishingly low error rate: In medium and dense traffic any car interacts with the car in front every few seconds (car distances when following another car oscillate on this time scale). For example, in Germany cars are driving some 10^{11} km annually while generating about 10^6 crashes per year. Using further data (the average velocity), this results in an estimate of the error rate in car following of about 10^{-6} to 10^{-7} per interaction.

Therefore, it seems that modeling should be simple, too. Unfortunately, this is not the case, and it is not easy to uncover the reasons for this complexity in the modeling which has kept the traffic flow modelers busy for more than seven decades (e.g. Bachmann 1938, Reuschel 1950, Wehner 1939, Greenshields 1935). This paper will look at modeling from the perspective of traffic jams and traffic jam formation. Different perspectives can be found, e.g., in Chowdhury et al. (2000a), Helbing (2001), Kerner (1999b), and Lebacque and Lesort (1999).

Already in the 1950s, the theory of kinematic waves (Lighthill and Whitham 1955, Richards 1956) has provided fundamental insight. As will be seen later, the theory of kinematic waves predicts stable jams only under some specific assumptions. However, measurements often display data clouds which at first glance seem not compatible with the theory of kinematic waves; instead, they display two clearly separated regimes. This has triggered speculations about an abrupt transition in traffic flow (Treiterer and Myers 1974, Koshi et al. 1983, Acha-Daza and Hall 1993), which were recently backed up by models (Kühne 1984, Kerner and Konhäuser 1994, Bando et al. 1994). In addition, the modeling technique of cellular automata (CA), also already introduced in the 1950s (Gerlough 1956, Cremer and Ludwig 1986), was used increasingly in the last decade (Biham et al. 1992, Nagel and Schreckenberg 1992, Takayasu and Takayasu 1993). Although there is no fundamental reason behind this, CA modeling often included stochastic aspects while the other approaches often did not. It turned out to be surprisingly hard to clarify the question of a possible (phase) transition in stochastic CA models (Nagel 1994, Nagel and Paczuski 1995, Roters et al. 1999, Sasvari and Kertesz 1997, Krauß et al. 1997, Chowdhury et al. 2000b). We will get back to this point

0030-364X/03/5105-0681
1526-5463 electronic ISSN

Operations Research © 2003 INFORMS
Vol. 51, No. 5, September–October 2003, pp. 681–710

in §3.3. Other works (Hall et al. 1992, Cassidy 1998, Windover and Cassidy 2001, Munoz and Daganzo 2003) point out that many aspects of the regime separation in measurements could also be caused by geometric constraints (see §2), and the authors imply that the kinematic theory is complete enough to explain the most relevant aspects of traffic flow. The question of the possible existence of a phase transition, and how this is reflected (or not) in the models, will be an important one throughout this paper.

After giving a short overview of empirical facts (§2), this paper will look at intuitive and formal arguments regarding traffic jams, including their formation and their stability (§3). That section will contain the most important aspects of the kinematic theory. A relatively long section (§4) then will look at microscopic models for traffic, including coupled differential equation models, CA models, and coupled maps (models which are continuous in space but discrete in time). Section 5 gives more information about so-called second-order fluid-dynamical models for traffic. Section 6 shortly discusses many additional aspects, such as so-called synchronized traffic, kinetic theory, or multilane traffic. The paper is concluded by a summary (§7).

2. SOME EMPIRICAL FINDINGS

The typical measurement of traffic flow is the so-called fundamental diagram. It displays traffic flow q (e.g., in vehicles per hour; also called throughput or volume) as a function of density ρ (e.g., in vehicles per km). Reasonably, there is no flow when there is no car on the road, $\rho = 0$, and there is also no flow when there is a dense jam, $\rho = \rho_{\max}$. In between, the flow reaches a maximum value $q_{\max}$ at some density $\rho_{q\max}$. Often, two regimes can be identified, separated by a gap: a "free-flow" regime at low densities, and a "congested" regime at high densities (see Figure 1).

For the free-flow regime it is possible to assign to the data a function $Q(\rho)$, whose detailed form depends on the external conditions (e.g., Wu 2000).[1] Basically, it is linear for small densities, but becomes more complicated for higher densities, presumably because the interactions between cars become more important (Newell 1955).

In contrast, the interpretation of the congested regime is not unequivocal. Because the scatter of the data is very large, it seems questionable if one can assign a function at all. In terms of density, the two regimes can overlap: For some intermediate densities, traffic can be in the free-flow regime or in the congested regime (Edie 1961, Hall and Agyemang-Duah 1991, Kerner and Rehborn 1996a, p. R1300). It is useful to follow the dynamics of the measurements, that is, to follow the *sequence* of points in the fundamental diagram. In Figure 1, in the congested regime this is indicated by lines between the data points. Often it is better to plot density ρ, flow q, and velocity v as functions of time (Figure 2). These measurements show that the typical transition is from the free-flow regime to a regime where throughput is virtually undiminished but densities are much higher, meaning much lower velocities (Mika et al. 1969, Kerner and Rehborn 1996b, Kerner 1998, 1999b). In Figure 1, this corresponds to a transition from the "free" branch around $\rho = 25$ veh/km to values on the data cloud marked "congested" around $\rho = 40$ veh/km. This type of congested traffic is sometimes termed "synchronized" because it often goes along with synchronized speeds between lanes.

For even larger densities the system becomes jammed, with small throughputs, small speeds, and large densities. In the jammed regime, throughput and density are again strongly correlated (not shown). For a more detailed description of this see Knospe et al. (1999b, 2002). Jams cannot be directly observed in the fundamental diagram, but they can be seen in spatio-temporal measurements.

Figure 1. Empirical fundamental diagram, as recorded on the German freeway A43.

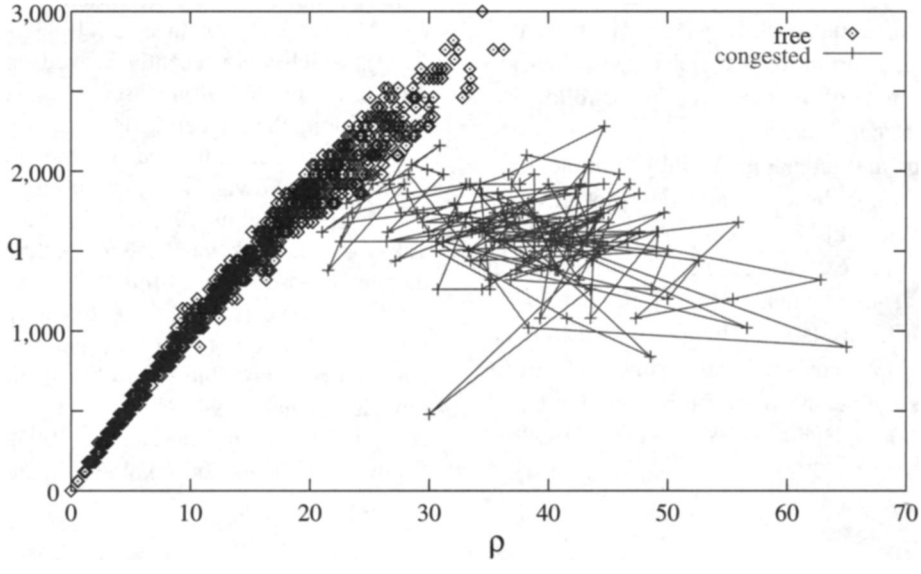
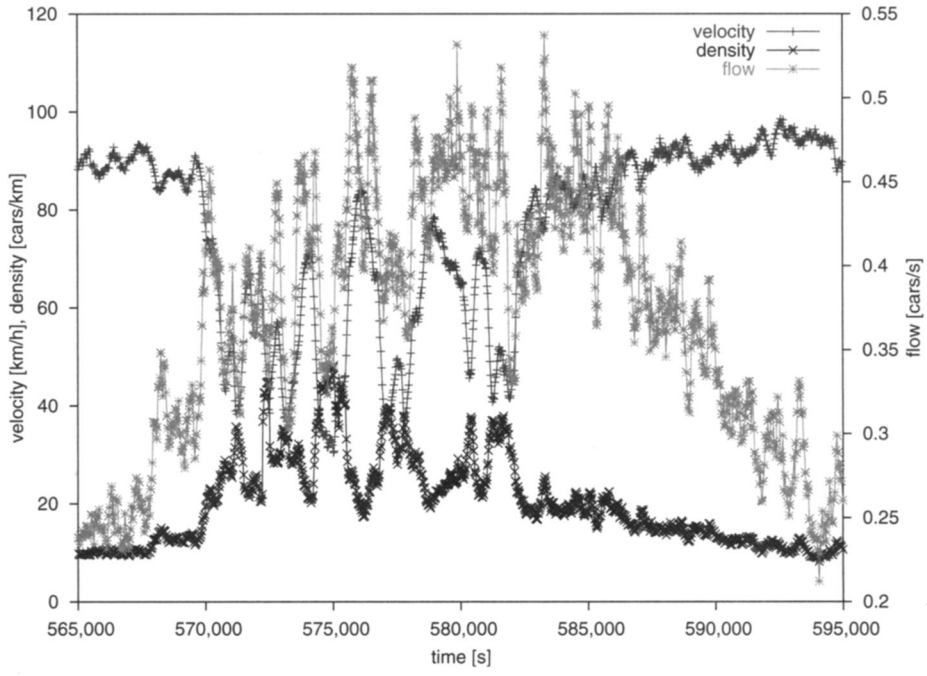


Figure 2. Time series of the three fundamental variables of traffic flow.



Note. The data was recorded on the German freeway A1 near an intersection with German freeway A59 in June 1996.

Jams typically move against the direction of traffic. If one assumes that cars move slowly or not at all inside jams, then this is easy to understand: Incoming cars join the jam at its upstream front, which moves this front in the upstream direction. On the downstream front, cars leave the jam, which also moves this front in the upstream direction. If such a jam moves across a fixed measurement location, then one will initially measure free traffic, then jammed traffic, then free traffic again. This leads, in the fundamental diagram, to some data points on the free-flow branch, then to some data points at high densities and low flows, and then again to some data points on the free-flow branch. The empirical findings about such wide jams can be summarized as follows (e.g., Newell 1962, Kerner and Rehborn 1997): (1) Traffic jams, once created, are fairly stable and can move without major changes in their form for several hours against the flow of traffic. (2) The flow *out* of a jam is a stable, reproducible quantity, and maximally possible flows can be up to 50% larger (see, e.g., Kerner and Rehborn 1996a, p. R1300, for empirical evidence).

The interpretation of these observations is less clear. One interpretation is that there are three phases: free flow, congested or synchronized, and jammed (Kerner and Rehborn 1996b, Kerner 1999b, Kerner and Klenov 2002, Kerner et al. 2002). An alternative explanation is that the effects are caused by geometrical constraints. For example, a bottleneck downstream of a measurement location can cause the following temporal sequence of measurements:

1. The system starts with low flow at low densities.
2. Both flow and density keep increasing, along the "free-flow" branch of the fundamental diagram.
3. This flow can be larger than what can flow through the bottleneck. Then, a queue starts forming upstream of the bottleneck, but that does not immediately influence the measurement.
4. Eventually, the queue will reach the measurement location. At that point in time, data points will move to a much higher density, while the flow value will be given by the bottleneck capacity.

That mechanism will generate plots similar to Figures 1 and 2, and one can argue that without these geometrical constraints the fundamental diagram would display a smooth function everywhere (Windover and Cassidy 2001, Daganzo et al. 1999, Muñoz and Daganzo in press, Cassidy 1998). However, although this second argument can explain occurrences of capacity drops in a fundamental diagram, it does not in our view explain the capacity drop from Kerner and Rehborn (1996a, p. R1300), which gives a more complete spatio-temporal picture, and where the flow into the jam is clearly larger than the flow out of the jam.

No matter what the origin of the different regimes, measurements indicate that all three possible transitions, that is, from free flow to synchronized, from synchronized to jammed, and from free flow to jammed, can happen (Kerner 1999a). A typical case of the transition from free flow to jammed is the one where a jam exists within a "background" of free flow (Kerner and Rehborn 1996a, p. R1300); similarly, a jam can exist in a background of synchronized flow (Figure 5 in Kerner 1999c).

Regime transitions which are visible in a fundamental diagram can be explained by the fact that a separating boundary between two regimes has moved across the sensor. But how do the different regimes come into being at

all? For example, how can jams emerge; that is, how can traffic which is in the free-flow regime transform to a mix of free-flow traffic and jams, called stop-and-go traffic?

This question is difficult to answer from existing data. Detailed analysis shows that there are often external reasons for breakdown (Daganzo et al. 1999); on the other hand, experiments with cars driving around in a circle show spontaneous breakdown (Kikuchi et al. 2001). However, even in this rather artificial case it took about 10 minutes until the breakdown happened. The drivers were instructed to drive as fast as possible. In some cases, it may be a matter of interpretation: Is the slow truck on the right lane a stochastic fluctuation or an external reason?

Regarding the "synchronized" regime, there is not even agreement on the question if the strong scatter of the data has a dynamical origin, e.g., being caused by an on-ramp, or a statistical origin, such as being caused by a mixture of cars and trucks which display different driving characteristics. In §6.1, we will discuss some of these hypotheses.

Complications come from the fact that most existing data are not well suited to decide these questions. Values for the fundamental diagram are averaged over time, i.e., $\langle q \rangle_T$ and $\langle \rho \rangle_T$, where T is in the range of several minutes. This is the first complication, because traffic regimes can change within minutes, thus resulting in data points which have contributions from more than one regime.

Next, traffic measurements are typically done as local (point) measurements, but density, as opposed to flow or velocity, is not a local quantity but needs space averaging. In practice, one often *computes* density from the two local quantities, i.e., either as $\rho = \langle q \rangle / \langle v \rangle$ or as $\rho = \langle q \rangle \langle 1/v \rangle$. The latter has theoretical advantages, but is difficult in practice, because most devices record $\langle v \rangle$ only, and also because small errors at low velocities lead to large errors of the density estimate (Gerlough and Huber 1975, Leutzbach 1988, Helbing 2001).

Further complications are due to the fact that traffic is strongly inhomogeneous, for example, consisting of cars and trucks, aggressive and relaxed drivers, or fast versus slow cars. This indicates that it would be better to collect single vehicle data instead of aggregated quantities, and leave the data aggregation to the analyst. A possible result of this is cumulative curves such as used in Daganzo et al. (1999).

Also, it is imperative that field measurements report sufficient additional information, such as the geometry of the road upstream and downstream, inflows or outflows upstream or downstream, additional measurements upstream and downstream, etc. We have already mentioned the fact that a bottleneck downstream will considerably influence the measurement. It would be best to have the complete spatio-temporal picture.

3. TRAFFIC JAMS

Given the fact that the interpretation of empirical findings is not clear, it is desirable to look at possible theories. This paper will do exactly that: look through models for traffic and see to what kind of predictions they lead. That is, it will concentrate on theoretical work (some analytical, mostly computational), and here on models which remain simple in the sense that complexity emerges from the interaction of simple ingredients. As will be seen, many of those models, from microscopic (car-following) to macroscopic (fluid-dynamical), can display a similar mechanism of instability. This instability triggers a transition from the free-flow regime to the regime that includes jams.

It is probable that the basic mechanism for this is determined by a few basic ingredients, such as excluded volume, inertia, and slow acceleration; the ultimate goal is to find them. Additional complexity will have to be added to make real-world forecasts. Nevertheless, the theoretical construct is useful. This is similar to the theory of phase transitions in physics, which is a useful theoretical construct because of its universality; yet, for real-world systems with gravity, inhomogeneities, condensation nuclei, etc., adjustments need to be made.

For most of the models, it is known that this mechanism can be switched on and off by changing the parameters of the models. One of the questions, therefore, is if traffic is better described with that mechanism switched on or switched off. As was pointed out in §2, in our view current field results are not sufficient to answer that question.

3.1. Intuitive Jam Stability

Assume an infinite system, with relatively low homogeneous density and homogeneous flow, except for a region of length ξ, where we have a jam (density close to $\rho_{\max}$). Under which circumstances will that jam grow? The intuitive answer is: when more vehicles are added at the upstream front than are released at the downstream front.

If the jam is caused by a geometric constraint, such as a red light, an accident, or a bottleneck, then this condition is easy to fulfill. If the bottleneck has service rate q_{bn} and the inflow is q_{in}, then if $q_{\text{in}} > q_{\text{bn}}$, the jam will grow. This is similar to standard queuing theory, except for the fact that the traffic jam has a spatial extension, i.e., it grows backward in space. The spatial growth can be described by the theory of kinematic waves (§3.2).

More unusual are jams which are not pinched to some geometric constraint, that is, they exist on a homogeneous stretch of road, embedded in free traffic upstream and downstream (e.g. Kerner and Rehborn 1996a, p. R1300). Let us, for the time being, skip the question of how this jam got started; assume, for example, that it was caused by a traffic accident which got cleared after a few minutes.

The intuition for such jams is the following. Consider five vehicles of velocity zero close to each other (=a small jam). In the first time step, only the first vehicle can move. In the second time step, the second vehicle can start, etc. However, in the meantime it can happen that other vehicles join the queue at the upstream end.

Given the right conditions (no less vehicles joining at the upstream end than leaving at the downstream end), this

results in a cluster of vehicles of velocity zero and that cluster will move against the traffic direction. The vehicle composition of this cluster is constantly changing—from the perspective of a driver, you join the jam from the end, the jam "moves through you", and then you can accelerate again. This is a standard wave phenomenon, and it is again described by the theory of kinematic waves (§3.2).

As pointed out in §2, the outflow from such a jam is a fixed quantity, which is approximately one vehicle every two seconds per lane. This is the average time it takes between two consecutive cars starting to move. Let us denote this fixed quantity by q^*. The jam will grow if $q_{\text{in}} > q^*$.

This implies that the formation of a growing jam in the outflow of another jam will be difficult. For example, if an accident gets cleared away, then under homogeneous conditions in the outflow from that accident significant jams should not occur. If in the outflow another jam was started, its inflow would be q^* (because it is the outflow from the first jam), and its outflow would also be q^*. As a result, the number of cars in that second jam would follow a random walk, with an absorbing boundary at zero.

Thus, the question arises whether traffic flow can reach sustained flows above q^* (call them supercritical flows), and if so, how. There is a long history of observations and discussions about "reverse lambda shape of the fundamental diagram" (Koshi et al. 1983), "hysteresis" (Treiterer and Myers 1974), "catastrophe theory" (Persaud and Hall 1989), "capacity drop" (Hall and Agyemang-Duah 1991) and the like, pointing to the possibility of supercritical flows. Other research (Hall et al. 1992, Windover and Cassidy 2001) maintains that such measurements are caused by geometric conditions, i.e., a large capacity road leading into lower capacity (see §2).

Let us entertain the idea that sustained supercritical flows were possible. Because, under homogeneous conditions, they cannot be caused as outflow from another jam, how could they come into existence? One possibility is by the merging of subcritical flows. For example, one could have an outflow from a jam, and then further vehicles merge via on-ramps into the flow. This means that outflow traffic would have space for additional vehicles to squeeze in. In other words, drivers at high speeds would accept smaller time headways than when accelerating out of a jam.

This is, however, still not the full answer. A second necessary ingredient, besides $q_{\text{in}} > q_{\text{out}}$, is that the density inside the jam should not decrease. How could the density inside the jam decrease? One way to imagine this is a situation in which as many vehicles are added upstream to the jam as are released downstream, but the two fronts are moving apart. In this situation, the same number of vehicles inside the jam is stretched over more and more space, until the density inside the jam is eventually the same as the density outside the jam, and the jam has disappeared.

3.2. Mathematical Jam Stability

To gain some formal intuition, let us look at the theory of kinematic waves (Lighthill and Whitham 1955, Haberman 1977, Haight 1963, Daganzo 1999). This theory starts with an equation of continuity,

$$\partial_t \int_{x_1}^{x_2} \rho(x,t)\,dx + q(x_2,t) - q(x_1,t) = 0, \tag{1}$$

where $q = \rho v$ is the current (also called flow or throughput or volume). For regular points x, where $\partial_t \rho$ and $\partial_x q$ are bounded, we get

$$\partial_t \rho + \partial_x q = 0. \tag{2}$$

This equation is the standard way to state mass conservation in fluid dynamics; it states that if, at a specific location, the flow coming in from the left is less than the flow going out to the right ($\partial_x q > 0$), then density has to decrease at that location ($\partial_t \rho < 0$).

The theory of kinematic waves now makes the assumption that q depends on ρ only, $q(x,t) = Q(\rho(x,t))$. With this assumption, $\partial_x q = (dQ/d\rho)\,\partial_x \rho$, and therefore

$$\partial_t \rho + Q'\,\partial_x \rho = 0, \tag{3}$$

with $Q' := dQ/d\rho$.

By assuming that ρ is a constant plus some disturbances, i.e., $\rho(x,t) = \bar{\rho} + \tilde{\rho}(x,t)$, the following equation is obtained:

$$\partial_t \tilde{\rho}(x,t) + \left(Q'(\bar{\rho}) + \tilde{\rho}(x,t)Q''(\bar{\rho})\right)\partial_x \tilde{\rho}(x,t) = 0, \tag{4}$$

and after retaining terms of at most linear order in $\tilde{\rho}$

$$\partial_t \tilde{\rho}(x,t) + Q'(\bar{\rho})\,\partial_x \tilde{\rho}(x,t) = 0. \tag{5}$$

A wave ansatz $\tilde{\rho}(x,t) = Ae^{i(kx-\omega t)}$ yields a wave solution of

$$\tilde{\rho}(x,t) = Ae^{ik(x-ct)} \tag{6}$$

with

$$c = Q'(\bar{\rho}). \tag{7}$$

That is, a disturbance on top of a homogeneous density of $\bar{\rho}$ travels with speed $c = Q'(\bar{\rho})$. This is the slope of a tangent applied to the diagram of $Q(\rho)$ at the position $\rho = \bar{\rho}$.

If there is a macroscopic shock, then linearization is no longer applicable. One can, however, solve directly for a shock that is moving but constant in its shape, by noting that, in the moving frame of the shock, the flow coming from the left has to be the same as the flow going to the right, because otherwise the shape of the shock could not remain constant. Technically one would solve Equation (1). Back in the fixed reference frame, the flow on the left is $\rho_1(v_1 - c)$, and the flow on the right is $\rho_2(v_2 - c)$, where c is the velocity of the shock. Setting this equal and solving for c yields

$$c = \frac{\rho_2 v_2 - \rho_1 v_1}{\rho_2 - \rho_1} = \frac{q_2 - q_1}{\rho_2 - \rho_1} = \frac{\Delta q}{\Delta \rho}. \tag{8}$$

That is, for a macroscopic shock, the velocity of the shock front is the slope of the secant connecting (ρ_1, q_1) to

(ρ_2, q_2) on the fundamental diagram. The limit of this for small disturbances yields the tangent which was the result of the linear analysis.

Equation (6) states that, under linearization, disturbances are marginally stable, that is, the amplitude of a disturbance neither grows nor shrinks. Let us make this point even clearer, by adding a diffusion term to the kinematic wave equation

$$\partial_t \rho + \partial_x Q(\rho) = D\,\partial_x^2 \rho, \tag{9}$$

where D is the diffusion constant and we have assumed that there is also a diffusive contribution to the flow, $q = Q(\rho) - D\,\partial_x \rho$. The linearized solution of Equation (9) is

$$\tilde{\rho}(x,t) = A e^{-Dk^2 t} e^{ik(x-ct)}. \tag{10}$$

This equation describes a wave, $e^{ik(x-ct)}$, multiplied by an exponentially decaying amplitude, $Ae^{-Dk^2 t}$. Disturbances with a short wave length (large k) will decay faster than disturbances with a long wave length. For $D \to 0$, this decay goes away and we recover the case from before. This is what is meant here by "marginally stable": Equation (6) is exactly on the margin between exponential growth and exponential decay.

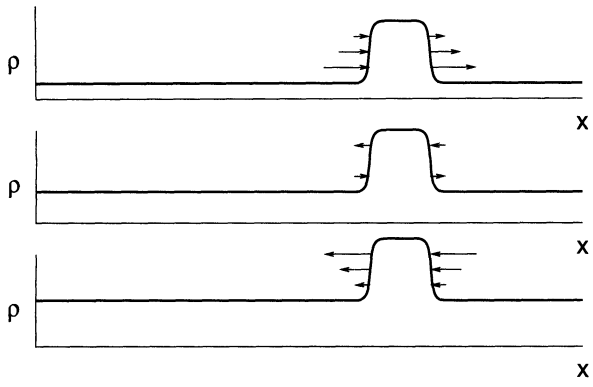
There is a version of Equation (9) which can be solved exactly. This equation is the so-called Burgers equation

$$\partial_t \rho + \partial_x \big(v_f \rho(1-\rho)\big) = D\,\partial_x^2 \rho, \tag{11}$$

which is Equation (9) with the specific choice of $Q(\rho) = v_f \rho(1-\rho)$. v_f is the free speed, as can be seen from the relation $v = q/\rho$ and taking the limit of $\rho \to 0$. We already know, from the above analysis, that small disturbances decay exponentially over time, but the Burgers equation can be solved analytically in its entirety, i.e., also for large disturbances (Burgers 1974).[2] The result for any localized initial disturbance is that it develops into a characteristic triangular structure, which grows in width as $\sim t^{1/2}$ and shrinks in height as $\sim t^{-1/2}$.

Without going through the analytical solution of the Burgers equation, the correct intuition is provided by Figure 3: Assume a "lump" with density ρ_{lump} and flow

Figure 3. Kinematic wave stability. The three graphs show different situations, with overall density increasing from top to bottom.



q_{lump} surrounded by traffic with a lower density ρ_0 and flow q_0. The velocity of both fronts is $c_0 = \Delta q/\Delta\rho < 0$. The fronts have different stability, though. For this, assume that the fronts have smoothed out a little bit, that is, they have finite slope. Remember that for finite slopes, the phase velocity is given by $c(\rho) = dQ/d\rho$, and let us assume strict monotonic decrease of $c(\rho)$ with increasing density (i.e., $dc/d\rho < 0$). Then:

- At the upstream front, c decreases from low to high density. This means that for the low-density parts of the front the phase velocity is larger than it is for the high-density parts of the front, meaning that the front makes itself steeper.
- The inverse is true for the downstream front: c increases from high to low density, which makes the low-density parts of the front faster than the high-density parts and thus smears out the front.

This means that, from the perspective of kinematic waves, under the assumed circumstances upstream fronts of traffic jams are stable while downstream fronts smear out; and if one thinks about a jam caused by a traffic light which just turned green, this appears plausible. This result is more general than just for the Burgers equation case, because only strict convexity of the $Q(\rho)$ curve is needed (Lax 1973).

If one gives up strictness, the result is changed. Assume a piecewise linear $Q(\rho)$ relation with a triangular shape, as given in Figure 13. This is now qualitatively different from the Burgers equation assumption. There are two cases:

- Both the density inside the jam and the surrounding density are on the same linear piece of the $Q(\rho)$ relation. Then, all parts of the front have the same phase velocity. This is true both for the inflow and the outflow front. That is, by making the outflow front more stable, the inflow front has become less stable. Without diffusion, i.e., when $D = 0$, both fronts are marginally stable, now in the sense that any disturbance of their shape will neither heal out nor amplify. When diffusion is added, i.e., when $D > 0$, then both fronts will slowly (diffusively) smear out.
- The density in the jam is on a different linear piece of the $Q(\rho)$ relation than the surrounding traffic. In this case, one again finds that the inflow front is stable while the outflow front is unstable. More precisely, this is true for those parts of the outflow front where the density is lower than the density at maximum flow, $\rho_{q\max}$. That is, the outflow front will smear out until outflow has reached the density where the linear piece of the $Q(\rho)$ relation starts.

Within this theory, true stability of the downstream front can only be achieved by a concave part of the fundamental diagram $Q(\rho)$. However, this would entail an unstable upstream front.

Given this, how does it come that there are observations of stable traffic jams surviving virtually unchanged in their profile for up to several hours (Kerner and Rehborn 1996a, §2)? There are several possible explanations:

- The theory of kinematic waves, with $D = 0$ and a piecewise linear fundamental diagram as explained above,

predicts that both fronts are marginally stable, i.e., that they will retain their shapes in the absence of external disturbances and in the absence of noise. Some researchers believe that this is sufficient as a full explanation. In such a system, the jam would not "pull together" if spread out by some external influence, but it would also not spread out by itself. By "pulling together" we mean that (1) steep fronts are stable both upstream and downstream, and (2) the jam has a preferred density which it attempts to restore if disturbed. Note, however, that we need $D = 0$ *and* a piecewise linear $Q(\rho)$ relation for this. Such a theory would not explain a capacity drop in the outflow of a jam. Recall that some researchers find such a capacity drop and some do not.

- The assumption for kinematic waves was that q is a function of ρ only. This also means that velocity v is a function of density ρ only. For example, it means that there should be no velocity adaptation delay effects when driving from light into dense traffic, or vice versa. Relaxing this assumption means the introduction of *inertia*, i.e., the effect that vehicles, similar to standard physics particles, have a tendency to retain their velocity even under the influence of adverse forces. It turns out that these new equations can possess a density range in which both jam fronts are stable. In other words, there is some kind of surface tension, which pulls *both* jam fronts toward the interior of the jam (see §3.3).
- A possibility, which has not yet been explored in detail, is that both fronts could possibly become truly stable simultaneously when the diffusion constant becomes negative for certain densities. This does not seem very intuitive but can happen in mathematical derivations (Nelson 2000). It is a recent development, and it is not clear in how far the "physics" behind the models is truly different from introducing inertia as outlined above.

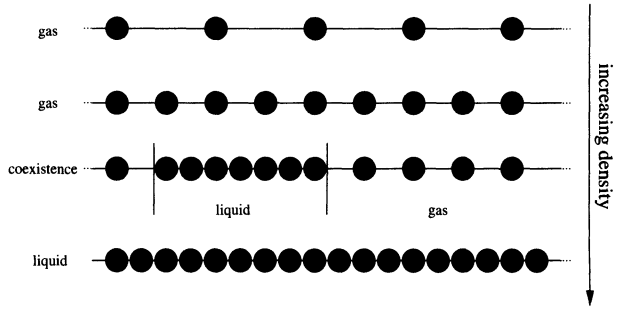
3.3. Stability, Breakdown, and Phase Transitions

In the last section, we have collected a number of arguments about traffic jam behavior. It would be nice to organize these arguments into a larger, consistent framework. An analogy with a gas-liquid phase transition may be able to provide that framework. Although that analogy has been discussed for many years (Prigogine 1961, Prigogine and Herman 1971, Reiss et al. 1986), no clear one-to-one correspondence has been established. The following paragraphs attempt to establish the analogy, based on our own recent research, but the results must be regarded as preliminary.

In the gas-liquid analogy, laminar traffic would correspond to the gas phase, jammed traffic would correspond to the liquid phase, and stop-and-go waves would correspond to the coexistence of liquid and gas. Based on that analogy, the expectation for increasing densities would be (see Figure 4):

1. At low densities, say below ρ_1, there is only the gas phase, corresponding to laminar traffic.

Figure 4. Schematic view of gas-liquid transition.



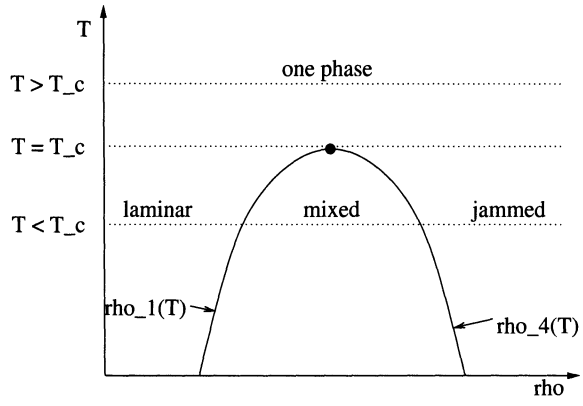
2. At intermediate densities, say between ρ_1 and ρ_2, droplets can form, corresponding to jams. Droplets are, however, only stable if they are above a certain size. In a deterministic system, a large enough external disturbance would be necessary to bring a droplet beyond that critical size. In a stochastic system, the same effect would eventually be achieved by a large enough fluctuation. As long as this has not happened, the system is *super-critical*. Once a droplet has formed, it will grow, and thus swallow additional molecules, until the density outside the droplet is back to ρ_1. The density inside the droplet will be at ρ_4, which is defined below.
3. At even larger densities, say between ρ_2 and ρ_3, droplets will form immediately. As in 2. above, they will grow until the density outside the droplet(s) is reduced to ρ_1. Droplets will coagulate until there is only a single droplet left. This is sometimes called the spinodal decomposition regime.
4. One would expect another regime, between ρ_3 and ρ_4, which behaves as the mirror image of the regime between ρ_1 and ρ_2: Sub-critical homogeneous liquid is possible until either a stochastic fluctuation or a large enough external disturbance starts the formation of one or more areas of gas. After that, the gas phase expands until the liquid phase is moved to a density of ρ_4. The gas phase will be of density ρ_1.
5. Above ρ_4, only the liquid phase exists.

For the gas-liquid transition, the densities ρ_1 to ρ_4 can be varied by changing the temperature. Increasing the temperature will move them closer together, until at a certain temperature T_c they eventually all meet in a so-called critical point (Figure 5). Above T_c, the differences between the phases have vanished; only one phase is left, independent of the density.

It now turns out that there are deterministic traffic models which behave as follows (Kerner and Konhäuser 1994, Bando et al. 1995):

- For a density range $0 \leqslant \rho \leqslant \rho_1$ and another density range $\rho_4 \leqslant \rho \leqslant 1$, homogeneous traffic is stable. This means that any disturbance will "heal out", and traffic will return to the homogeneous state.
- For a density range $\rho_2 \leqslant \rho \leqslant \rho_3$, homogeneous traffic is (linearly) unstable. This means that *any* disturbance will initially grow in size, and it will not go away once it is there (Figure 6).

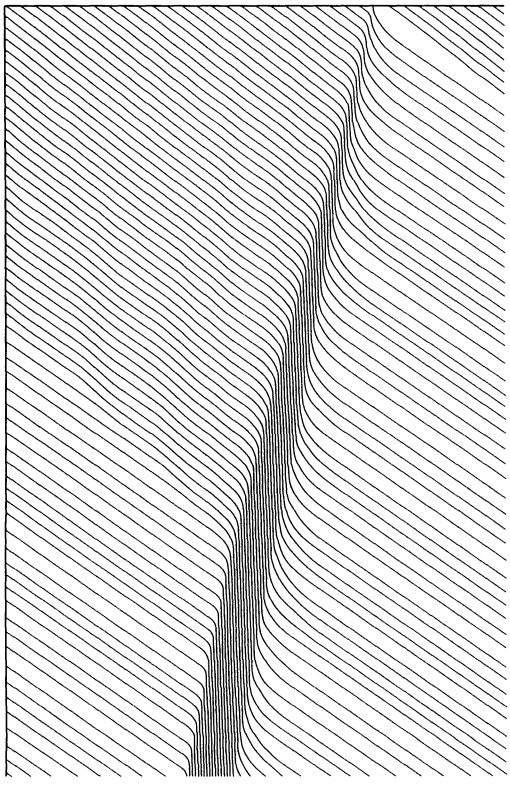
Figure 5. Phase diagram of gas-liquid transition.



• For density ranges $\rho_1 \leqslant \rho \leqslant \rho_2$ and $\rho_3 \leqslant \rho \leqslant \rho_4$, homogeneous traffic is unstable under large amplitude disturbances. This means that small disturbances heal out, but large enough amplitudes cause the system to move into an inhomogeneous state. Once in the inhomogeneous state, it will remain there except if pushed back into the homogeneous state by another large amplitude disturbance.

With respect to jam stability, both jam fronts are stable for all densities between ρ_1 and ρ_4.

Figure 6. Space-time plot of a spontaneous traffic jam.



Note. Lines show consecutive time steps of configurations on a simulated one-lane road; each pixel in a line is a different car; only every 10th car is plotted. The reduced outflow from the jam could be seen clearly. Because the simulation was run on a loop, this reduced outflow finally reaches the upstream front of the jam, creating a precarious equilibrium between the outflow and inflow of the jam.

The values for ρ_2 and ρ_3 are given by equations such as (Kerner and Konhäuser 1994)

$$\rho_i \frac{dV}{d\rho}(\rho_i) = \text{const} \tag{12}$$

or (Bando et al. 1994)

$$\frac{dV}{d\rho}(\rho_i) = \text{const}. \tag{13}$$

Here, $dV/d\rho$ is the derivative of the desired velocity with respect to the density.

In both cases, $|dV/d\rho|$ can be so small everywhere or const so large that the equation has no solution. This implies that, similar to the gas-liquid transition, one can change the parameters of the models such that the possibility of different phases eventually goes away.

As pointed out above, field results are inconclusive with regard to the question if this is a valid description, and if so, if real traffic is in the 1-phase or in the 2-phase regime. Kerner and Rehborn (1996a, p. R1300) state that "the flux out of the jam, which . . .equals the threshold for traffic jam existence, was considerably lower than the maximal flux in laminar traffic flow". They also find hysteresis loops, i.e., the transition from free to congested flow does not occur at the same density as the transition back. In contrast, Cassidy (1998) finds a uni-valued flow curve for all values of the density, which excludes hysteresis.

Deterministic models can generate different phases and different regimes, but they cannot say anything about spontaneous transitions between the regimes. More precisely, between ρ_1 and ρ_2 spontaneous transitions from the laminar to the coexistence state should be possible in a stochastic system. Differently said, between ρ_2 and ρ_3 the laminar state would immediately break down, while for $\rho_1 < \rho < \rho_2$ the system would be *metastable*, staying in the laminar state for some time until it eventually breaks down. Because the stochastic fluctuations need to conspire to produce an amplitude large enough to overcome the barrier, the waiting time until this happens increases the more one goes away from ρ_2 towards lower densities.

Crystallization cores, such as seeds of a different material, can reduce that waiting time. In traffic, there is a multitude of possible crystallization cores, such as slow trucks, bottlenecks, grade changes, changes of speed limit, etc. This means that the fact that traffic jams typically start at such crystallization cores by itself is not an argument against a 2-phase theory.

There is no systematic work to extend the above deterministic models into the stochastic regime. Work with stochastic approaches has instead concentrated on so-called CA models; some of these models will be presented later. A problem with stochastic models with respect to phase transitions is that it is impossible to readily identify the different phases. The most important reason for this is that gas-liquid phases are characterized by a density difference, but density is not a particle property, but a field property

(i.e., averaged over many particles). Also, the analogy cannot carry through completely because the traffic system is asymmetric, with drivers receiving more stimulus from ahead than from behind. This leads to the effect that one front of a jam can be stable while the other one is unstable; in a gas-liquid transition, either the whole interface is stable, or the whole interface is unstable.

Work by Jost (2002) and Jost and Nagel (2003) implies that phases can be unequivocally established as long as one is far enough away from T_c, but it becomes difficult once one approaches it. The consequence of this is that, *once one has identified* the traffic model parameter that corresponds to temperature, then some good analogy between the gas-liquid transition and laminar-jammed transition seems to be possible. However, without identification of that parameter, characterization of models is difficult.

This difficulty is related to the fact that close to the critical point, properties of the system look critical up to some length scale, no matter if one is on the 1-phase or on the 2-phase side of the critical point, i.e., above or below T_c in Figure 5. More precisely, from the theory of critical phase transitions, one would expect the following (e.g., Stauffer 1985):

- Exactly at the critical point, the system would show certain properties such as fractal behavior and critical slowing down. Critical slowing down means that one has to wait long times until systems are in equilibrium. Both aspects make computational work difficult.
- Away from the critical point, these properties remain in existence up to length and time scales (l^*, t^*) which depend on the distance $|T - T_c|$ from the critical point. The further one is away from the critical point, the shorter are these length and time scales. This means that one can be above the critical point ($T > T_c$), but the system may still display structure formation, i.e., jams embedded in laminar traffic. Only when averaging over length scales larger than l^*, one will no longer measure jams and laminar traffic, but a homogeneous system corresponding to one phase only.

This second effect has confounded research in the area for a long time (Roters et al. 1999, Chowdhury et al. 2000b). Current evidence (Jost 2002, Jost and Nagel 2003) implies that one of the often-used cellular automata models, the Stochastic Traffic Cellular Automata (STCA) model (§4.3.3), by pure chance, is close to but not exactly at a critical point.

Another complication arises with respect to breakdown behavior. According to the theory of phase transitions, one would expect that in a 2-phase model, laminar traffic between ρ_1 and ρ_2 would eventually break down (Figure 7, left). We found, however, that for one stochastic 2-phase model, supercritical laminar traffic is abnormally stable (Nagel et al. 2003) (Figure 7, right). It was impossible to establish via computational experimentation if this is a deviation from the theory or just a very large pre-factor in front of an otherwise consistent mathematical expression. These results refer to the Krauß model (§4.4) with parameters $(a, b, \epsilon) = (0.2, 0.6, 1)$.

Finally, for the STCA model, there is a limiting case (the so-called cruise control limit) where breakdown is artificially suppressed. In that situation, computational experiments have established fractal or near-fractal structures (Nagel and Paczuski 1995); similar results are found in related work (Gray and Griffeath 2001). It is unclear if these fractal structures are a coincidence or if there is a systematic reason behind them.

The consequence of all this is the following. Theory allows for models which, by varying their parameters, can go from 1-phase models to 2-phase models. When these models are close to that transition, they can be technically 1-phase models but display aspects of 2-phase models. Models based on kinematic theory alone (§3.2) are typically 1-phase models (i.e., all jams smear out), but specific versions of the kinematic theory, with a piecewise linear fundamental diagram and without diffusion, predict jams which do not smear out by themselves. This means that from a theoretical perspective several types of behavior are possible for traffic, and diligent experimentation, maybe combined with additional theoretical insight, will have to resolve this issue. Resolution of this issue seems important with respect to real-world traffic management. Considerable amounts of money are invested in systems which are supposed to prevent breakdown—a measure that only seems to make sense if breakdown is possible, which assumes 2-phase traffic.

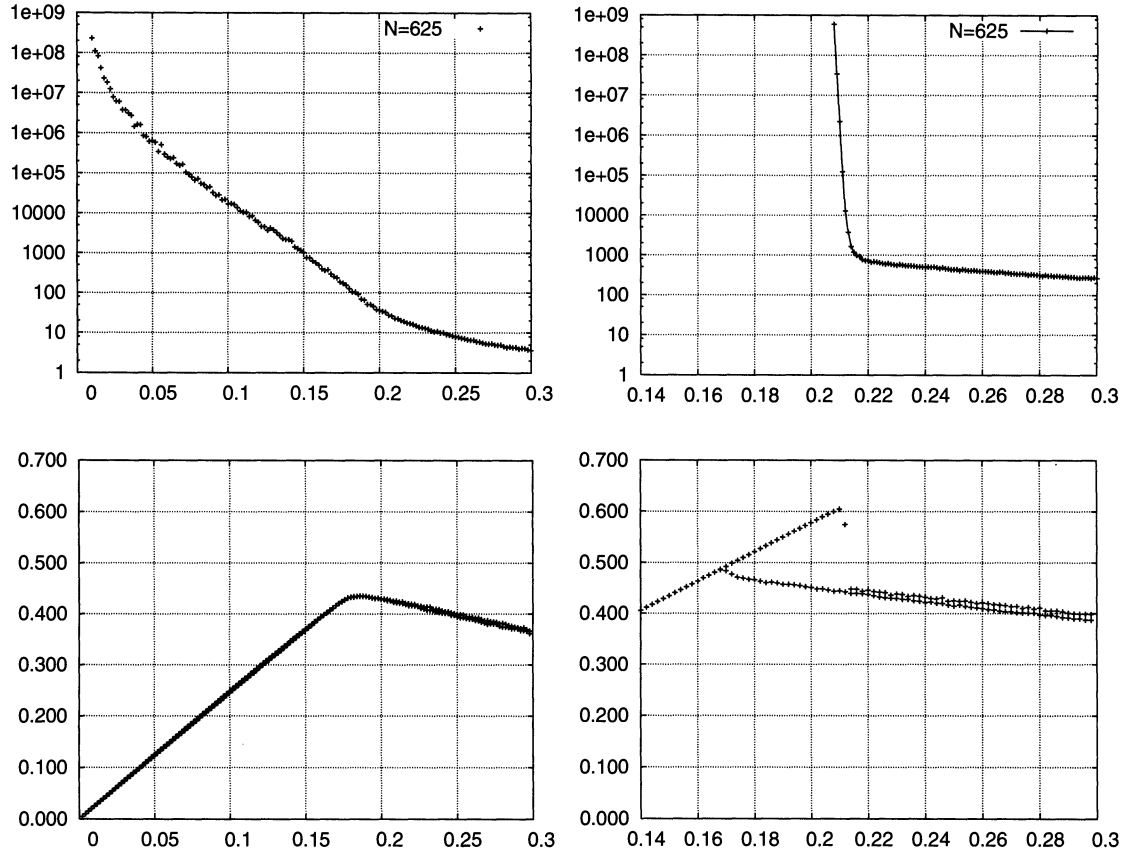
Most of the descriptions in this paper technically refer to "closed systems," i.e., idealized "traffic in a closed loop" where vehicles neither enter nor leave the system. This is done for the purpose of an easy exposition; all arguments can be extended to open systems. Nevertheless, open systems can show dynamics which go beyond the dynamics of closed ones. This can be seen for the CA models of §4.3, where analytical results for open systems exist (e.g., Rajewsky and Schreckenberg 1997). This will be the topic of §6.2.

4. MODELING TRAFFIC FLOW MICROSCOPICALLY

4.1. Preliminaries

For *dynamical* traffic modeling, in which we are interested here, there are two principal methods: microscopic and fluid dynamical. Microscopic means that every car is resolved individually as a particle. For the fluid-dynamical approach, some aggregation takes place, leading to aggregated quantities such as flow or density, and partial differential equations connecting these quantities. Fluid-dynamical approaches can be seen as a subclass of macroscopic approaches; an example of a macroscopic model which is not fluid dynamical is the volume-based link cost function which is used in static assignment (e.g., Sheffi 1985).

Figure 7. Top row: Times to breakdown. Bottom row: Fundamental diagrams after 5,000 time steps. Left column: Normal stochastic model. Right column: Stable stochastic model. From (Nagel et al. 2003).



Besides the reviews mentioned in §1, there are a number of Internet sites (SMARTEST, DLR/VF 2001) that give an overview of traffic models. Despite large differences in the approach and scope, most current models can be classified in the following way: they have an underlying function $V_{\text{des}}(\ldots)$ (the "driving relation") which defines a desired velocity as a function of distance g to the car ahead (microscopic) or of the car density ρ (macroscopic). This desired speed function serves as a fixed point for the car-following dynamics. An interesting topic is the stability of this fixed point: the models where this fixed point, i.e., the corresponding homogeneous state, becomes unstable at some density ρ_c are models where a phase transition occurs.

There are also models that display more complex behavior, e.g., Wiedemann (1974), Knospe et al. (2000), Tomer et al. (2000), Treiber et al. (1999a), and Kerner and Osipov (2002). We will comment on these approaches in §6.5. Besides being simplistic in their car-following dynamics, the models of this section are also all single lane. As mentioned before, the topic of multilane traffic is treated briefly in §6.6.

4.1.1. Reaction Time Argument. The basic argument for car-following theory is that vehicles follow each other with constant time headway, which means that space headway is proportional to speed. Many of us learn this at driving school as the "two-second rule" or "distance in meters equals half the speed in km/h." This may be against intuition, because braking distance is proportional to the *square* of the speed, but it can be justified as follows (cf., Figure 8):

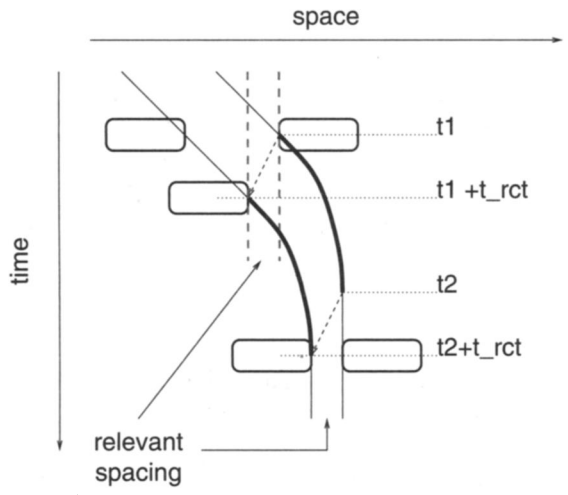
1. Assume two cars following each other with a constant distance and constant speed v.

2. Assume that the leading car suddenly brakes at time t_1. The back bumper of this car at this time is at position $\tilde{x}_1$. Assume further that the following car has a reaction delay of t_{rct} and in consequence starts braking at time $t_1 + t_{\text{rct}}$ and at position x_2. Reaction delay means an effective reaction time, which includes, for example, both anticipation and mechanical delay.

3. If both vehicles have exactly the same deceleration profile $b(v)$, then it means that they have exactly the same braking distance d. In consequence, the rear bumper of the leading car will come to a stop at $\tilde{x}_1 + d$, and the front bumper of the following car at $x_2 + d$. That is, if the front bumper of the following car was beyond $\tilde{x}_1$ (i.e., $x_2 > \tilde{x}_1$) when it started breaking, then it will crash into the leader ($x_2 + d > \tilde{x}_1 + d$).

The consequence of this argument is that one's space headway g needs to be larger than the distance vt_{rct} that one covers during one's reaction delay. This is the same as the requirement that one's time headway needs to be larger than one's reaction time.

Figure 8. Reaction time argument.



Note. The figure shows the trajectories of the rear bumper of the leader and the front bumper of the follower plus the position of the entire vehicle as some specific points in time. At t_1, the leader starts braking; at t_2, she has come to a standstill. The follower starts braking at $t_1 + t_{rct}$; and because his braking follows exactly the same characteristics, he comes to a standstill at $t_2 + t_{rct}$. The position of the follower's front bumper at $t_1 + t_{rct}$ with respect to the leader's rear bumper at t_1 is the spacing that will remain when the two cars are stopped. When it is negative, an accident cannot be avoided.

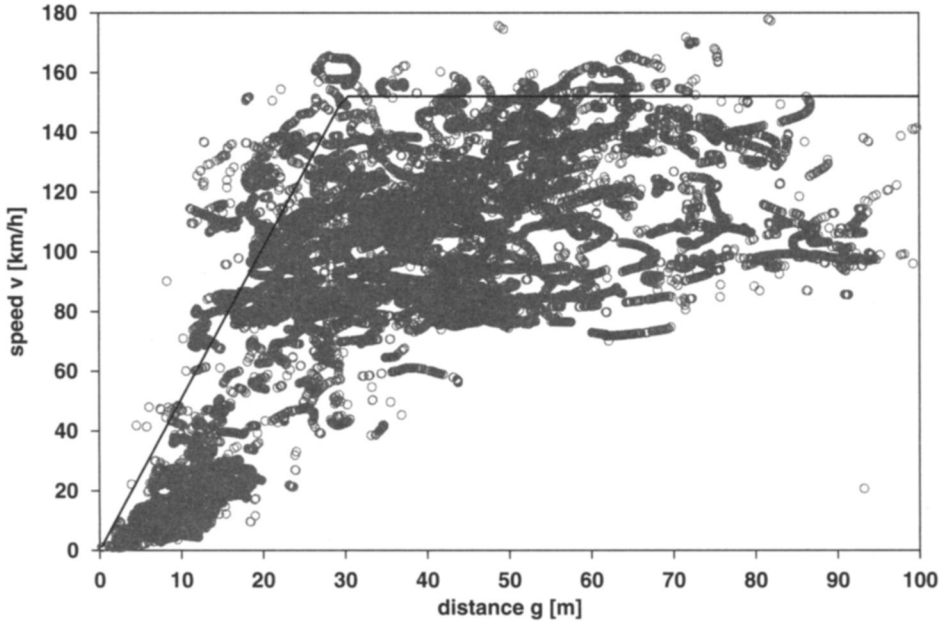
For jam stability, this means that one condition for stable jams is that the average time headway when escaping from a jam needs to be more than the average time headway during normal driving (Newell 1962, Krauß 1998). It is plausible that this is the case in reality.

4.1.2. Driving Relation and Homogeneous Solution. The last section implied that, under homogeneous conditions, time headways would be constant and therefore space headways proportional to speed, i.e., $g \propto v$. It is clear that this can only be an approximation; for example, in very dilute traffic one will have g very large while v does not grow beyond the maximum speed. As an extension of $g \propto v$, let us define the *driving relation*. By this we mean the desired distance as a function of speed under homogeneous conditions. In some average sense such a function exists even under stochastic conditions: At a given speed, people will follow other cars with a certain average distance. One could measure this, and average it over all drivers and all times (Figure 9). It is plausible to assume that this desired distance grows with speed; as a result, it is invertible and we obtain desired speed as a function of the distance to the car ahead.

Next, we note that the distance to the car ahead is related to the density. If we define Δx as the front-bumper-to-front-bumper distance and l as the space a car occupies in a jam (car length plus some additional space), then the gap $g := \Delta x - l$ is a reasonable measure of the "distance" that goes into the driving relation. In consequence, we have $\rho = 1/\Delta x = 1/(g + l)$. With this, we can define the driving relation as a function of the density, i.e., $V(\rho)$. Most car-following models and most fluid-dynamical theories contain such a relation $V(\rho)$ more or less explicitly. A noteworthy exception is the traditional car-following theory (§4.2.1), which allows any distance for any velocity.

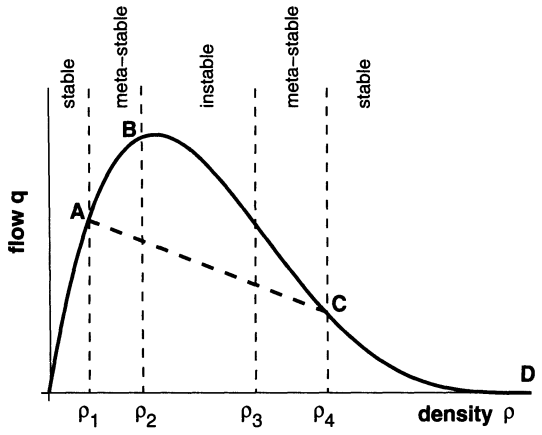
The driving relation yields a homogeneous solution, which for many car-following models is a fixed point of

Figure 9. Data from a car-following experiment.



Note. Shown are pairs of velocity and gap taken from a dataset consisting of the speed, gap, speed of the leading car, and acceleration sampled for three hours. Plotted are only those data where the speed difference between the cars was smaller than 5 km/h, and the acceleration was smaller than 0.5 m/s^2, respectively. Plotted is a fitted piecewise linear function with slope $1/\tau$, where $\tau = 1.4$ s. Note the big scatter in the data, despite the fact that only data from a near following situation was chosen.

Figure 10. Stylized fundamental diagram for many traffic flow models where the homogeneous solution has an unstable range, as described in the text.



Note. The plot shows the various points of instability, ρ_1, ρ_2, ρ_3, and ρ_4, respectively. Note that for homogeneous models, all ρ_i collapse into one, meaning that these models are stable for the entire density range.

the dynamics: If we have the density given, we line up all cars with equal distance and with the speed $v = V(\rho)$. Because all cars will drive with the same speed, they will have again equal distance after any amount of time has passed. From the driving relation, $V(\rho)$, follows a line in the flow-density diagram, $Q(\rho) = \rho V(\rho)$. This is the solid line in Figure 10. Some models are not stable everywhere on this line. As explained in §3.3, the generic way in which the instability exists in models is that there are densities ρ_1, ρ_2, ρ_3, and ρ_4 (Figure 10), such that the homogeneous solution is stable below ρ_1 (the laminar phase) and above ρ_4 (the jammed phase), it is meta-stable for $\rho_1 < \rho < \rho_2$ and for $\rho_3 < \rho < \rho_4$, and it is instable for $\rho_2 < \rho < \rho_3$.

Once the instability is triggered and traffic breaks down, the model traffic reorganizes into a coexistence of laminar and jammed regions. The densities of these laminar and jammed regions are at ρ_1 and ρ_4, respectively. It was already said before that not all researchers agree on the existence of this instability in reality.

4.2. Continuous Space and Time

In microscopic modeling, each vehicle is individually resolved. Vehicles are described by a vector of state variables, typically (x_n, v_n, a_n), meaning the spatial location, the speed, and the acceleration (all three along the one-dimensional road) of the nth vehicle. A model then consists of a set of rules or equations to update these quantities over time, depending on the states of other vehicles around.

Let us assume that vehicles move to the right, that is, towards increasing x, and that they are numbered from right to left, which corresponds to the temporal sequence in which they pass a standing observer. Other useful quantities, which will be used in the following, are front-bumper-to-front-bumper distance $\Delta x_n := x_{n-1} - x_n$, speed difference $\Delta v_n := v_{n-1} - v_n$, and the gap $g_n := \Delta x_n - l$ as above. Because $g = 0$ in a jam, $1/l$ is the jam density. Here we assume that l is the same for all vehicles. Of course, more realistic microscopic simulation tools do not have this restriction. Inhomogeneous fleets are discussed in §6.6.

A useful starting point results from the observation that time headways between vehicles are relatively constant (Richards 1956), resulting in $g \propto v$. Note that this follows also from the reaction time argument of §3.1. A simple update rule is (Newell 2002)

$$v(t) \propto g(t-\tau), \tag{14}$$

where $g(t-\tau)$ refers to the gap at time $t-\tau$. More generally, car-following models have the form

$$v_n(t) = V(v_n(t-\tau_v), \Delta x_n(t-\tau_v), \Delta v_n(t-\tau_v)) \tag{15}$$

or

$$\dot{v}_n(t) = A(v_n(t-\tau_a), \Delta x_n(t-\tau_a), \Delta v_n(t-\tau_a)), \tag{16}$$

where $\dot{v} = dv/dt \equiv a$ is the acceleration of the vehicle. There are sometimes mathematical and sometimes heuristic arguments which connect these two approaches.

4.2.1. "Classic" Car-Following Models. The "classic" car-following model family (Gerlough and Huber 1975, Gazis et al. 1961) is

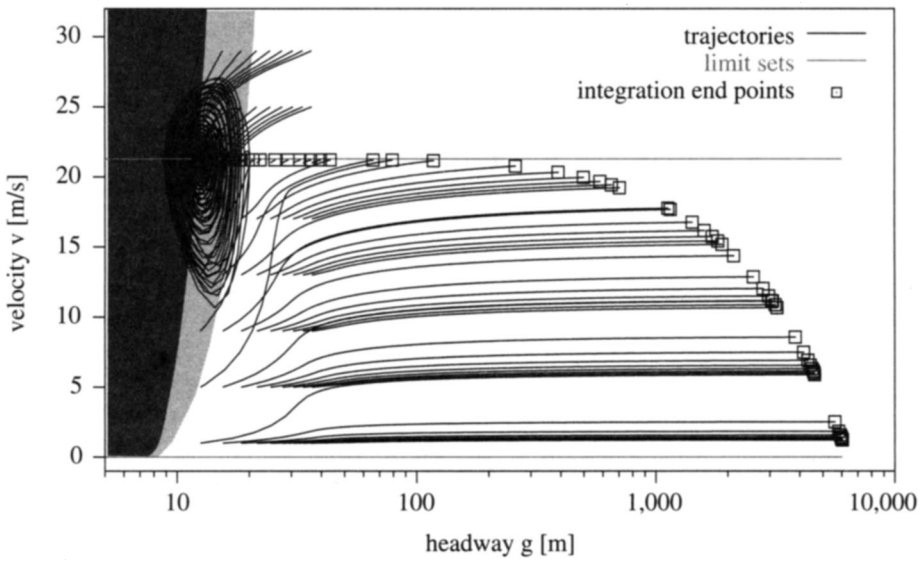
$$\dot{v}(t) = \alpha \frac{[v(t)]^l}{[\Delta x(t-\tau)]^m} \Delta v(t-\tau) \tag{17}$$

(dropping the car index to simplify notation), where m and l are free, usually fractional, exponents. Note that $\dot{v}(t) \propto \Delta v(t-\tau)$ can be obtained by taking the time derivative of Equation (14); the effect of this is, as will be seen, that the dynamics do not longer react to Δx, but only to changes of it.

This model was possibly the first to recognize the importance of instability of traffic flow and to capture this notion into a mathematical model. Let us analyze the situation of one car following a lead car that drives with constant speed $v_0 > 0$ (Herman et al. 1959). In this case, the system has three different limit sets, i.e., attraction states for Equation (17), which are obtained by setting the time derivative, $\dot{v}(t)$, to zero: (i) $\Delta v = 0$, (ii) $v = 0$, and (iii) $\Delta x \to \infty$ (see Figure 11).

(i) means that *any* space headway is a fixed-point solution ($\dot{v} = 0$) as long as the follower has the same speed as the leader. (ii) means that the second vehicle having no speed at all is a fixed point of Equation (17). The limit set (iii) with $\Delta x \to \infty$ may look surprising at first, but has a simple explanation: if the initial distance to the car ahead is too large, the following car is not able to follow the car ahead because the acceleration is not high enough; eventually, the follower will be left so far behind that, because of the large Δx, its speed will not change any more.

Figure 11. Phase-space portrait for Equation (17), with $\alpha = \tau = l = 1$ and $m = 1.7$.



Note. The velocity of the leading car is 22 m/s. Plotted are g and v for different initial conditions g_0, v_0. In addition to the trajectories with their end points, the plot shows the limit sets $v = 0$ and $\Delta v = 0$ (at $v = 22$ m/s). The initial conditions for any starting point in the (g, v)-plane is a function in the time interval $[-\tau, 0]$, where $(g(t) = g_0, v(t) = v_0)\ \forall t \in [-\tau, 0]$ has been chosen. Simulations starting in the dark region to the left are unstable, leading to growing oscillations and eventually crashes; no trajectories are shown here. Simulations starting in the lightly shaded region have oscillatory damped solutions, as can be seen by the circular trajectories. Simulations starting in the white region have overdamped solutions, which means that they approach $\Delta v = 0$ slowly or not at all. The trajectories have been integrated by using an adaptive step-size control algorithm (Owren and Zennaro 1992).

Certain combinations of parameters and initial conditions lead to a crash between the two cars. This is true if $m < 1$ and $v(t = 0)$ is large.

The stability analysis performed in Gazis et al. (1961) states that for $\alpha\tau v_0^l / g_0^m > \pi/2$, the solution $\Delta v = 0$ becomes unstable. When looking at more than two cars ("asymptotic stability", or better platoon stability), the model becomes unstable earlier, that is, for a smaller value of $\alpha\tau v_0^l / g_0^m > 1/2$ (Gazis et al. 1961, Zhang and Jarrett 1997).

If one checks explicitly for car crashes ($\Delta x < 0$), then attempts to numerically integrate the system with more than two vehicles are usually not successful. Only for a limited range of parameters and initial conditions (small sensitivity α, small "reaction" time τ, small velocity v_0 (for a given spacing), and large spacing g_0 (for a given speed)), do such simulations succeed. The model behavior can be summarized as follows (e.g., Helbing and Tilch 1998, Ceder 1977):

- There is no acceleration without a lead vehicle. Therefore, an additional acceleration term like $\dot{v} \propto v_{\max} - v$, is needed for realistic applications.
- The model is structurally unstable. For example, adding that acceleration term leads to a different model; in this case, the limit set $\Delta v = 0$ vanishes. Models of this type are unlikely to describe reality, because small changes in the equations can lead to large changes in the dynamical behavior.
- There is no preferred distance; any state where Δv vanishes is a stationary state.
- It is not capable of describing traffic *beyond* the onset of instabilities because it lacks a mechanism that limits oscillations to realistic values—i.e., to values limited by the acceleration and braking capabilities of vehicles.

An alternative to Equation (17) is (Newell 1961)

$$v(t) = G(\Delta x(t - \tau)). \tag{18}$$

It turns out that there are many relations between Equations (17) and (18) (Newell 1962), including the considerations regarding stability. This becomes clear when one takes the first derivative of Equation (18) with respect to time, yielding

$$a(t) = G'(\Delta x(t - \tau))\Delta v(t - \tau), \tag{19}$$

with $G'(X) = dG/dX$. This is indeed similar to Equation (17). One important difference is that the solution of Equation (17) which admits arbitrary distance as long as $\Delta v = 0$ is no longer present in Equation (18). Technically, this is due to an integration constant that does not depend on t; it indicates that Equation (19) will yield the same solution as Equation (18) when started with valid initial conditions and when no integration errors occur. The latter cannot be achieved by numerical integration, meaning that from a computational point of view, Equation (18) is much more useful than Equation (17) or (19).

4.2.2. The Optimal Velocity Model (OVM). A model that is in the tradition of Newell (1963) is the so-called optimal velocity model (OVM) (Bando et al. 1995). Its version of the acceleration equation (16) reads

$$\dot{v} = \alpha \cdot (V(\Delta x) - v), \tag{20}$$

where the index n is again left out for convenience, $V(\Delta x)$ is a sigmoid function that has the features $V(\Delta x) \to v_{\max}$ if $\Delta x \to \infty$, and $V(\Delta x) = 0$ at a certain distance Δx. A version of Equation (20) can be found by Taylor-expanding $v(t+\tau) \propto g$ (Equation (14)) in τ and dropping terms of second or higher order in τ.

This model has been analyzed in Bando et al. (1995). It is an example for a model that displays the deterministic breakdown picture described in §3.3 (see Figure 10). For small densities, only a homogeneous solution exists, where all cars drive with the same velocity $v = V(1/\rho)$. Above a certain density ρ_1, this homogeneous solution becomes unstable against large disturbances, which can kick it into a state where the system has one or more jams separated by regions of free flow. At a certain higher density ρ_2, the homogeneous solution becomes *linearly* unstable, which means that the tiniest disturbance will now move the system away from the homogeneous state. As said before, the densities ρ_2 and ρ_3 for linear instability are given by

$$\frac{dV}{d\rho}(\rho_i) = \text{const.} \tag{21}$$

As also said before, this equation has no solution when const becomes large or when $|dV/d\rho|$ is small everywhere.

In general, the OVM has the disadvantage that it is not completely crash free, and it displays unrealistically large accelerations.

4.2.3. Combined Models and Discussion. It seems that a better attempt could be made to make the development of car-following theory more systematic. This could be along the following lines. Assume that drivers have a *goal*, for example, to drive with some gap $G(\cdot)$, depending on the surroundings. Obvious values from the surroundings are the velocity of oneself, v, and the velocity of the car ahead, $\tilde{v}$.

One would assume that any such theory could be linearized at the operating point, leading to a desired gap of

$$G(\cdot) = C_0 + C_1 v - C_2 \tilde{v}, \tag{22}$$

where the signs are set such that one would expect the constants to be positive. A more complicated theory could have nonlinear influences and cross terms. Such models combine classic car-following and optimal velocity models.

Next, one needs an equation that models the approach towards the goal; this could be

$$a(t) = \frac{1}{\tau}(g(t) - G(\cdot)). \tag{23}$$

Using the linearized $G(\cdot)$ of Equation (22) yields

$$a(t) = \frac{1}{\tau}\big(g(t) - C_0 - C_1 v(t) + C_2 \tilde{v}(t)\big). \tag{24}$$

Most current models are formulated in terms of a desired velocity instead of in terms of a desired gap. One can indeed make a similar argument when starting from desired velocity. This yields, for example,

$$V(\cdot) = D_0 + D_1 g + D_2 \tilde{v}, \tag{25}$$
$$a(t) = \frac{1}{\tau'}(V(\cdot) - v(t)),$$

and

$$a(t) = \frac{1}{\tau'}\big(D_0 + D_1 g(t) + D_2 \tilde{v}(t) - v(t)\big).$$

$D_0 = D_1 = 0$ and $D_2 = 1$ yields a linearized version of the model equation (17); both ($C_0 = C_2 = 0, C_1 = 1$) and ($D_0 = D_2 = 0, D_1 = 1$) yield a linearized version of Equation (20).

4.3. Discrete Time and Discrete Space (CA Models)

In the previous section, we described car-following models that were continuous both in time and in space. Intuitively, it seems that these should be the only useful models, or at least the most realistic ones, because reality is continuous both in time and in space. However, computer implementations often are "time stepped" and thus effectively coarse grained in time, for example, in one-second increments. In fact, much progress in recent years has been made by using models that take this approach one step further: they also coarse grain space.[3] These models are often called cellular automata (CA) (von Neumann 1966, Wolfram 1986).

In CA models for traffic, space is typically coarse grained to the length that a car occupies in a jam ($l = 1/\rho_{\text{jam}} \approx 7.5$ m), and timed typically to one second (which can be justified by reaction-time arguments (Krauß 1998)). A side effect of this convention is that space can be measured in "cells" and time in "time steps", and usually these units are assumed implicitly and left out of the equations. A speed of, say, $v = 5$, means that the vehicle travels five cells per time step, or 37.5 m/s, or 135 km/h, or approximately 85 mph. CA models using these values yield plausible results with respect to reality, where plausible means that results are not more than a factor of two away from field data. The premier purpose of CA models is to add to general understanding; the TRANSIMS project (TRANSIMS) is an exception where some effort was spent to make a CA model also useful for a real-world application (Nagel et al. 1997).

CA models are rule based, instead of equation based, such as Equation (17); a simple CA model can, for example, be as follows (Nagel and Schreckenberg 1992):

- **Car-following rule:**

$$v_{t+\frac{1}{2}} = \min\{v_t + 1, g_t, v_{\max}\}. \tag{26}$$

- **Randomization:**

$$v_{t+1} = \begin{cases} \max\{v_{t+\frac{1}{2}} - 1, 0\} & \text{with probability } p_n, \\ v_{t+\frac{1}{2}} & \text{else.} \end{cases} \tag{27}$$

- **Moving:**

$$x_{t+1} = x_t + v_{t+1}.$$

t and $t+1$ refer to the actual time steps of the simulation; $t+1/2$ denotes an intermediate result used during the computation.

The first rule describes deterministic car following: try to accelerate by one velocity unit except when the gap is too small or when the maximum velocity is reached. Gap g is again the distance to the car ahead.

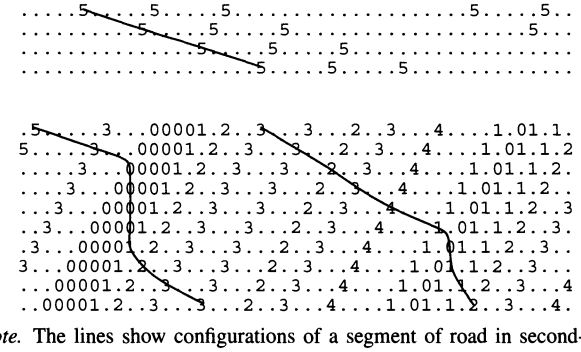
The second rule describes random noise: with probability p_n, a vehicle ends up being slower than calculated deterministically. This parameter simultaneously models effects of (i) speed fluctuations at free driving, (ii) overreactions at braking and car following, and (iii) randomness during acceleration periods.

Despite somewhat unrealistic features on the level of individual vehicles, these models describe aspects of the macroscopic behavior correctly. In addition, for certain simpler CA models it can be shown (Nagel 1999, 1996) that they have fluid-dynamical limits that turn out to be kinematic equations of the type Equation (2).

4.3.1. Deterministic CA 184. A simple deterministic CA model can be obtained when in the above rule set the randomization rule Equation (27) is dropped. Then, all initial conditions eventually lead to one of two regimes, depending on the overall density in the system (see Figure 12):

- **Free-flow regime** (Figure 12, top): Vehicles move with velocity $v_{\max}$, and the gap between vehicles is either $v_{\max}$ or larger. In consequence, the flow in this regime is $\langle q \rangle_L = \rho_L v_{\max}$. This happens when the density in the closed system is smaller than $\rho_c = 1/(v_{\max}+1)$. $\langle \cdot \rangle_L$ refers to averages taken over the whole system of size L; ρ_L is the global density N/L.

Figure 12. Sequence of configurations of CA 184.



Note. The lines show configurations of a segment of road in second-by-second time steps; traffic is from left to right. Integer numbers denote the velocities of the cars. For example, a vehicle with speed "3" will move three sites (dots) forward. Via this mechanism, one can follow the movement of vehicles from left to right, as indicated by some example trajectories. Top: Uncongested traffic. Bottom: Congested traffic. The backwards moving structures are the kinematic waves according to the Lighthill-Whitham (1955) theory.

- **Start-stop regime:** If the systemwide density ρ_L is larger than ρ_c, not all vehicles can move with maximum speed. The system goes to a state where the velocity of a vehicle is always the same as its gap (Figure 12, bottom). In consequence, the average speed is

$$\langle v \rangle_L = \langle g \rangle_L = \frac{1}{\rho_L} - 1 \tag{28}$$

and the average flow is

$$\langle q \rangle_L = \rho_L \langle v \rangle_L = 1 - \rho_L. \tag{29}$$

One will recognize Greenshields' model (Greenshields 1935) $V = (1/\rho) - (1/\rho_{\text{jam}})$ and $Q = 1 - (\rho/\rho_{\text{jam}})$, where $\rho_{\text{jam}} = 1$ due to the cellular structure of the model.

- The two regimes meet at

$$\rho_c = 1/(v_{\max}+1) \quad \text{and} \quad q_c = v_{\max}/(v_{\max}+1); \tag{30}$$

this is also the point of maximum flow.

What all this means is that there is a maximum driving speed, $v = \min\{g, v_{\max}\}$, and that the traffic adjusts in a way that vehicles always drive with this speed, that is, traffic is never slower than $v = \min\{g, v_{\max}\}$. In addition, the vehicles arrange themselves in a way that either all gaps are $\geqslant v_{\max}$ (uncongested regime), or all gaps are $\leqslant v_{\max}$ (congested regime).

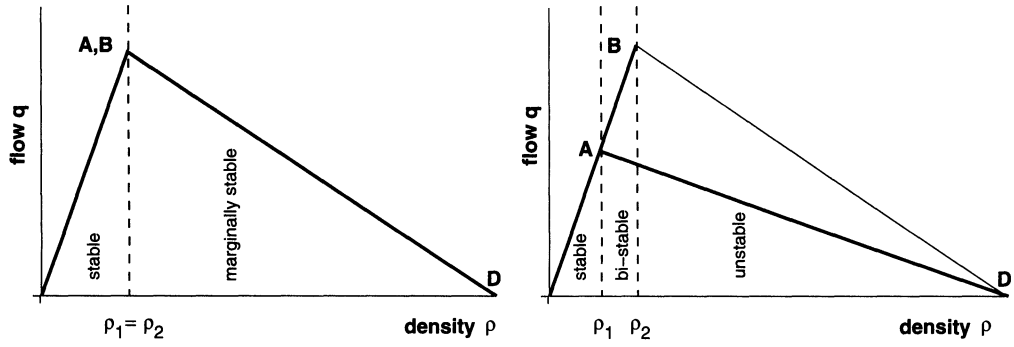
In consequence, the macroscopic traffic flow follows exactly the driving relation in the sense of §4.1.2. With respect to Figure 10, this means that the points A and C and thus the connecting line between them do not exist. Somewhat more precisely, both A and B move up to the maximum of the fundamental diagram (Figure 13, left). And in some sense, the point C becomes spread out over the whole decreasing branch of the fundamental diagram, because in this regime, there are many configurations which are marginally stable, including completely homogeneous configurations and start-stop-wave configurations. Intuitively, marginally stable means that the system neither goes to a homogeneous solution all by itself nor does it cluster all by itself.

Note that this is the piecewise linear model from §3.1. More precisely, the deterministic CA of this section can be analysed in terms of kinematic waves as follows:

- In the laminar regime, we have $dq/d\rho = v_{\max}$. This means that the waves have the same speed as the traffic—that is, they are simply "clusters" or "platoons" of cars (Figure 12, top).
- In the congested regime, $dq/d\rho = -1$. This can be seen in the space-time diagram via the fact that the "patterns" move backwards one cell in each time step (Figure 12, bottom).
- With respect to the introductory problem in §3.1 with the five cars: The jam has density $\rho = 1$ and speed $v = 0$, thus also $q = 0$. Outflow from the jam is eventually at $v = v_{\max}$ and $\rho = 1/(v_{\max}+1)$ (this can be seen by following the dynamics). In consequence,

$$\frac{\Delta q}{\Delta \rho} = \frac{v_{\max}/(v_{\max}+1) - 0}{1/(v_{\max}+1) - 1} = -1. \tag{31}$$

Figure 13. Fundamental diagrams of deterministic CA rules.



Note. Left: Standard CA 184. The fundamental diagram follows exactly the driving relation although the solutions are not necessarily homogeneous. Right: CA 184 with slow-to-start rules. There is a bistable range, where a large enough disturbance will kick homogeneous traffic into the start-stop regime, and an unstable range, where small disturbances will kick homogeneous traffic into the start-stop regime.

Thus, the downstream front of the jam moves backwards with speed $c = -1$. The inflow is somewhere on the "laminar" branch of the fundamental diagram. That means that the slope of the line connecting to $(\rho = 1, q = 0)$ is either -1 or less steep. The inflow front thus moves backwards with speed 1 or less—that is, the jam will eventually vanish except when inflow is exactly equal to maximum flow.

In terms of jam stability, we have the feature that, as long as the traffic is either completely in the laminar or completely in the congested regime, we have $dq/d\rho = \text{const}$. This means that both fronts are exactly marginally stable, meaning that disturbances neither heal out nor trigger further instability. In consequence, this model is a microscopic version of the "special" model of §3.3.

4.3.2. Slow-to-Start Rules for Deterministic CA Models. In §3.1, we mentioned that for jams to be stable, the reaction time and thus the minimum time headway need to be smaller than the escape time of vehicles from the jam. The CA from §4.3.1 violates this rule because both times are exactly the same (which explains why the congested regime is exactly at the margin between stability and instability). A way to obtain models that represent this aspect of the dynamics correctly is the use of so-called slow-to-start rules (Takayasu and Takayasu 1993, Barlovic et al. 1998). One reduces acceleration at low speeds, for example, by writing

$$v_{t+\frac{1}{2}} = \begin{cases} \min\{v_t + 1, v_{\max}, \max\{g_t \underline{-1}, 0\}\} & \text{if } v_t < 1, \\ \min\{v_t + 1, v_{\max}, g_t\} & \text{else,} \end{cases} \tag{32}$$

where the underlined part is the important difference. Here, a vehicle with speed less than one needs to wait until the gap to the vehicle ahead is at least two before it is allowed to move—in a typical outflow situation from a dense jam, this will take two time steps, thus making this "escape time" larger than the minimum headway time.

As a result of this rule modification, moving traffic is unaffected, but once traffic breaks down, the outflow is reduced, which stabilizes the traffic jam. That is, compared to the standard CA 184, ρ_1 and point A have moved to the left and down while ρ_2 and point B remain in place (Figure 13, right). The "driving relation" solution can exist over the whole density range. However, for densities larger than ρ_1, a large enough disturbance will kick traffic into the start-stop solution. Above ρ_2, stopping a car for exactly one time step is sufficient to trigger the instability, while between ρ_1 and ρ_2 a larger disturbance may be necessary. Thus, within the limits of the coarse description, this model displays bistability and instability (Figure 10) exactly as the model in §4.2.2 based on continuous variables.

The result is the typical "reverse λ" shape in the fundamental diagram—a homogeneous branch (the laminar phase) that goes from zero up to $\rho_c = \rho_2$ and $q_c = q_2$ (point B in Figure 10) as defined in Equation (30), and a start-stop branch (coexistence between jammed and laminar phase) that goes from ρ_1 and q_1 (point A) to $\rho_{\text{jam}} = 1$ and $q = 0$ (point D). As explained earlier, ρ_1 is given by the outflow from a macroscopic jam. In the deterministic slow-to-start model, it is thus (Barlovic et al. 1998)

$$\rho_1 = \frac{1}{2v_{\max} + 1}, \tag{33}$$

which is indeed smaller than $\rho_2 = 1/(v_{\max} + 1)$.

4.3.3. "Traditional" Stochastic Traffic CA (STCA) Models. The complete stochastic CA model as introduced at the beginning of §4.3 is not that easy to characterize within our framework. Simulations of that model show jam formation, indicating a 2-phase model. The model does, however, not possess any metastability; that is, traffic jams will either form jams within very few time steps, or essentially never. The minimum gap during free driving is $v_{\max}$, whereas the typical gap for escape from a jam is about three times as big (assuming $p_{\text{noise}} = 0.5$), meaning that the model in principle qualifies as a 2-phase model. There were several discussions about possible critical (=fractal) behavior for this model (Nagel 1994, Nagel and Paczuski 1995, Sasvari and Kertesz 1997, Roters et al. 1999, Chowdhury et al. 2000b). We will get back to this in §4.3.5.

4.3.4. Stochastic CA Models with Slow-to-Start Rule (STCA-s2s). It is obviously possible to add a slow-to-start rule as in §4.3.2 to stochastic CA models (Barlovic et al. 1998, Schadschneider and Schreckenberg 1997). A possible set of rules would be:

- **Car-following rule:**

$$v_{t+\frac{1}{2}} = \begin{cases} \min\{v_t+1, v_{\max}, \max\{g_t-1, 0\}\} & \text{if } v_t < 1, \\ \min\{v_t+1, v_{\max}, g_t\} & \text{else.} \end{cases} \quad (34)$$

- **Randomization:**

$$v_{t+1} = \begin{cases} \max\{v_{t+\frac{1}{2}} - 1, 0\} & \text{with probability } p_{\text{noise}}, \\ v_{t-\frac{1}{2}} & \text{else.} \end{cases} \quad (35)$$

- **Moving:**

$$x_{t+1} = x_t + v_{t+1}. \quad (36)$$

Similar to the deterministic case (§4.3.2), moving traffic is unaffected. But once traffic breaks down, the outflow is reduced when compared to the standard STCA, and this stabilizes the traffic jam. This means that a plot for the breakdown times, which looks as in Figure 7 (right) is unaffected by the slow-to-start rule, while recovery becomes difficult at much lower densities than without the slow-to-start rule. In consequence, it is possible to start STCA-s2s at much lower densities than the standard STCA and still eventually obtain a breakdown, which means that the resulting model is metastable in a meaningful way.

The result is the maybe so far most convincing description of breakdown behavior using CA models (Figure 6). Traffic can remain homogeneous up to a density ρ_2 for long times (which get shorter when approaching ρ_2); once the traffic breaks down, the outflow is at a density ρ_1 significantly smaller than ρ_2.

An alternative version is "velocity-dependent randomization" (VDR), which means that the noise probability p_{noise} is varied depending on what velocity the vehicle has at the beginning of the time step (Barlovic et al. 1998, Schadschneider and Schreckenberg 1997). Often,

$$p_{\text{noise}} = \begin{cases} p_0 = 0.5 & \text{if } v_t < 1, \\ p = 0.99 & \text{if } v_t \geqslant 1. \end{cases} \quad (37)$$

With these particular values, the acceleration from zero to one remains the same as before, but all other accelerations are "faster." The overall behavior is the same.

4.3.5. Phases in Stochastic CA Models. As discussed above briefly, a characterization of the stochastic CA models with respect to the possible existence of phases has proven to be difficult. Based on work by Jost (2002) and Jost and Nagel (2003), we conjecture the following. These results have to be regarded as unconfirmed at this point.

The stochastic slow-to-start model corresponds to the 2-phase description, i.e., to Figure 10. The model has 2 phases, laminar and jammed. For intermediated densities, these phases coexist, and depending on the density, transition from the supercritical laminar to the coexistence regime may take some time, indicating metastability. In other words, this is the first time in this text that we encounter a model that has clear metastability.

When decreasing p in Equation (37) while leaving p_0 at 0.5, then points A and C move closer together and eventually vanish. Our results indicate that this happens below $p = 0.5$, but not far away from it. This would mean that the original STCA of §4.3.3 is technically a 1-phase model, but close to a critical point and thus close to a 2-phase model. This conjecture is consistent with all other observations. In particular, it would predict critical (=fractal) behavior up to a certain length scale, beyond which it would appear homogeneous.

A consequence of the 1-phase behavior is that a 1-phase model technically does not possess an outflow front—any such front will smear out over time. This also means that any argument about escape times is not sufficient; escape time arguments by itself are not enough to predict stable jams (§3.2).

4.3.6. Time-Oriented CA (TOCA). A modification to make the STCA more realistic is the so-called time-oriented CA (TOCA) (Brilon and Wu 1999). The motivation is to introduce a higher amount of elasticity in the car following, that is, vehicles should accelerate and decelerate at larger distances to the vehicle ahead than in the STCA, and resort to emergency braking only if they get too close. The rule set is easier to write in algorithmic notation, where $v := v+1$ means that the variable v is increased by one at this line of the program. For the TOCA velocity update, the following operations need to be done in sequence for each car:

1. If $(g > v \cdot \tau_H)$, then, with probability p_{ac},

$$v := \min\{v+1, v_{\max}\}. \quad (38)$$

2. $v := \min\{v, g\}$.
3. If $(g < v \cdot \tau_H)$, then, with probability p_{dc},

$$v := \max\{v-1, 0\}. \quad (39)$$

Typical values for the free parameters are $(p_{\text{ac}}, p_{\text{dc}}, \tau_H) = (0.9, 0.9, 1.1)$. The TOCA generates more realistic fundamental diagrams than the original STCA, in particular when used in conjunction with lane-changing rules on multilane streets.

4.3.7. Dependence on the Velocity of the Car Ahead. Also in the CA context, there are implementations of the idea to include the velocity of the car ahead. The idea here is that if the car ahead is faster, then this adds to one's effective gap and one may drive faster than without this. This corresponds to C_2 or D_2 larger then zero in §4.2.3. In the CA context, the challenge is to retain a collision-free parallel update. Wolf (1999) achieved this by going through the velocity update twice, where in the second round any major velocity changes of the vehicle ahead were included. Barrett et al. (1996) instead looked at the gap of the vehicle ahead. The idea there is that, if one knows the gap of the

vehicle ahead, and one makes assumptions about the driver behavior of the vehicle ahead, then one can compute bounds on the behavior of the vehicle ahead in the next time step. With respect to breakdown dynamics, both models are less well understood, although Wolf (1999) reports simulation results which imply robust bistability near maximum flow.

4.4. Discrete Time and Continuous Space

This section started with models that are continuous in space and time because they seemed best suited to describe reality. A big disadvantage for those models is that, for being faithful to reality, one needs to introduce the reaction or mechanical time delay τ_v or τ_a. For computer implementations, this means that one needs to keep, within a selected time discretization, the whole history of the system between $t-\tau$ and t. Because for those models one desires fine time discretization, this is a considerable overhead.

We then moved to models that are coarse-grained discrete both in space and in time. Because of their simplicity, those models are well suited for understanding certain phenomena, but they have shortcomings when they need to be calibrated to reality.

There is a third class of models that sits "in between" the other two classes: The class of models that are continuous in space but coarse-grained discrete in time (Gipps 1981, Krauß 1998, Krauß et al. 1997, Yukawa and Kikuchi 1999, Sauermann and Herrmann 1998); sometimes they are called coupled maps. From a practical point of view, they are more suitable for applications than the CA models because they are numerically as efficient as the CA models but easier to calibrate. They are also easier to implement than models with continuous time because they explicitly use coarse-grained time in the model formulation and thus allow to use a corresponding time-stepped update scheme.

Obviously, a multitude of models is possible in this class—as in all of the other classes. We want to concentrate on a single model, a model described by Krauß (Krauß 1998, Krauß et al. 1997). This model is particularly well understood.

The approach starts from a consideration of the braking distances, i.e., from the observation that one's braking distance plus the distance one drives until one reacts should be smaller than the braking distance of the car ahead plus the space in between the two vehicles. This is once more the reaction time argument of §4.1.1. Formally, this yields

$$d(v_n)+v_n\tau \leqslant d(v_{n-1})+g_n, \tag{40}$$

where $d(v_n)$ is the braking distance of the nth car moving with speed v_n, τ is the reaction time, and g_n is the distance to the car ahead.[4]

Under certain conditions, this can be used to derive (see Figure 14) an update scheme for the dynamical state of a car. Dropping once more the car index and denoting the velocity of the leading car by $\tilde{v}$, one obtains

$$v_{\text{safe}} = \tilde{v}(t)+\frac{g(t)-\tilde{v}(t)\tau}{\bar{v}/b+\tau}, \tag{41}$$

$$v_{\text{des}} = \min\{v(t)+ah, v_{\text{safe}}, v_{\max}\}, \tag{42}$$

Figure 14. Derivation of the model by Krauß.

Derivation of the safe velocity

Assuming $v \approx \tilde{v}$, $d(\cdot)$ can be expanded around $\bar{v} := (v+\tilde{v})/2$, because $v = \bar{v}+\delta v$:

$$d(v) = d(\bar{v})+\delta v d'(\bar{v})+\frac{(\delta v)^2}{2}d''(\bar{v})+O((\delta v)^3).$$

As a result, the following is correct up to second order:

$$d'(\bar{v})v+v\tau \leqslant d'(\bar{v})\tilde{v}+g.$$

After noting the kinematic relation

$$d'(\bar{v}) \equiv \frac{d}{dv}d(\bar{v}) = \frac{\bar{v}}{b(\bar{v})}$$

($b(v)$ is the desired deceleration of a car) and rearranging terms, one obtains

$$v \leqslant \tilde{v}+\frac{g-\tilde{v}\tau}{\tau+\bar{v}/b(\bar{v})}.$$

Showing the collision freeness

Assume a time-discrete update, as in Equations (41)–(43). From those equations, it is known:

$$v(t+h) \leqslant \tilde{v}(t)+\frac{g(t)-\tilde{v}(t)\tau}{\tau+\bar{v}/b}. \tag{$*$}$$

Also, in general one has for the gap

$$g(t+h) = g(t)+h\big(\tilde{v}(t+h)-v(t+h)\big).$$

After using Equation ($*$), rearranging terms, and using the notation $\xi(t) := g(t)-h\tilde{v}(t)$ one gets

$$\xi(t+h) \geqslant \xi(t)\left(1-\frac{h}{\tau+\bar{v}/b}\right)+h\tilde{v}\frac{\tau-h}{\tau+\bar{v}/b},$$

a map $\xi(t) \to \xi(t+h)$. Thus, $h \leqslant \tau$ is a sufficient condition to ensure that if $\xi(t) \geqslant 0$, then $\xi(t+h) \geqslant 0$, meaning that $\xi(t) \geqslant 0$ for all t if $\xi(t=0) \geqslant 0$. Because of the definition of ξ, this ensures that $g(t) \geqslant 0$ for all $t \geqslant 0$.

$$v(t+h) = \max\{0, v_{\text{des}}-\epsilon a\eta\}, \tag{43}$$

$$x(t+h) = x(t)+hv(t+h). \tag{44}$$

$\bar{v} = (v+\tilde{v})/2$ is the average velocity of the two cars involved, a is the maximum acceleration of the vehicles, b is their maximum deceleration, ϵ is the noise amplitude, and η is a random number in $[0, 1]$. The terms are similar to what has been seen in the other models:

- The first rule (i.e., Equation (41)) can be rewritten as

$$v_{\text{safe}} = \alpha\frac{g(t)}{\tau}+(1-\alpha)\tilde{v}(t) \tag{45}$$

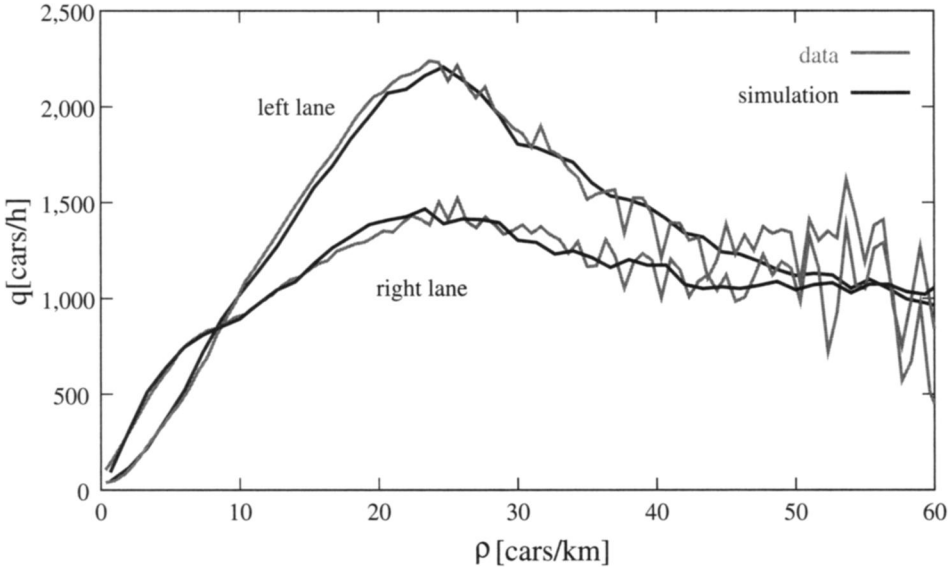
with

$$\alpha = \frac{1}{\bar{v}/(b\tau)+1}. \tag{46}$$

That is, v_{safe} is a weighted average of g/τ and $\tilde{v}$. This is a version of Equation (25) with D_1 and D_2 larger than zero.

- The second rule (i.e., Equation (42)) states that the velocity is limited by the desired acceleration a, by the safe velocity v_{safe} as calculated above, and by the maximum velocity $v_{\max}$.

Figure 15. Comparison between a simulated and a measured fundamental diagram, separated for the two lanes of a two-lane German freeway, using the model of Equations (41)–(44).



• In the third term, noise η is added by randomly making the vehicle slower than so far calculated. η denotes a random variable between zero and one, and ϵ is a noise scaling factor.

• The fourth term denotes the forward movement.

For $h \leqslant \tau$, one can show that this model is free of collisions (Figure 14); normally, one uses $h = \tau = 1$ sec. This model shares some features with Gipps' model (Gipps 1981). However, it is numerically more efficient, and it is better understood than the Gipps model. (See Wilson 2001 for an analysis in a similar spirit.) For both the Gipps and this model it can be shown that they describe empirical fundamental diagrams well, and with the appropriate lane-changing model they describe the lane-usage distributions faithfully (Gipps 1986, Krauß 1998) (Figure 15). Even for multilane traffic the model by Krauß retains the numerical efficiency of the CA models: it can be run with several million of update steps per second.[5]

The importance of the work by Krauß lies in the fact that he pointed out the existence of different regimes depending on the parameters of the model. He varied the parameters a and b and identified the following three classes (Krauß 1998, Brankov et al. 1996, Janz 1998):

• Class "III": For large values of a and b, the model does not display any structure formation, i.e., the space-time plots look homogeneous.

• Class "II": By decreasing a, but keeping large values of b, a regime is entered where the space-time plots show a complicated pattern of small jams within larger jams.

• Class "I": When approaching moderate values of a and b, and large enough density, after long enough times the space-time plot contains few large jams which grow deterministically in size while the number of jams keeps decreasing. The corresponding fundamental diagram displays a bistable behavior (reversed λ-shape).

Recent work indicates that this characterization is incomplete. While the transition between Class I and Class III is well defined, there seems to be no clear quantitative criteria for the other transitions (Janz 1998). Contrary to earlier statements, it is unclear if the STCA (§4.3.3) truly belongs in Class II of the model by Krauß (1998) (Nagel et al. 2003).

Finally, there is some confusion between, on the one hand, the existence of phases, and on the other hand, the characterization of statistical fluctuations that can end a metastable state (Nagel et al. 2003). It turns out that the model by Krauß (1998) with $(a, b, \epsilon) = (0.2, 0.6, 1.0)$ is abnormally stable in the density range where it should be metastable according to all other criteria (Figure 7). It is unclear if this is a quantitative or indeed a qualitative difference. The interested reader is referred to ongoing research (e.g., Nagel et al. 2003, Jost 2002, Jost and Nagel 2003).

4.5. Summary

Microscopic models for traffic follow each car individually. Traditionally, such models are formulated as coupled differential equations. For those, analytical solutions often cannot be obtained, and one resorts to computer simulation. This computer simulation has a time step h, and the simulation approaches the mathematical formulation for $h \to 0$.

For traffic, these coupled differential equations need to be delay equations because of reaction and mechanical delays. Differential delay equations are more difficult than standard ordinary differential equations, both analytically and numerically.

On the other hand, it has been known for quite some time that one can get macroscopic quantities of multiparticle systems right even if one uses coarse space and time

discretizations, and rules instead of equations. The best-known examples are CA for hydrodynamics (e.g., Stauffer 1991). These approaches are even more suitable for traffic, because by using the reaction delay as their natural time step they turn the delay into an advantage.

CA models for traffic are difficult to calibrate to reality. It is, however, possible to use the coarse-time discretization together with a continuous space resolution. This combines the advantages of the two approaches.

All these models can be deterministic or stochastic. Deterministic models display some important features, such as laminar/congested regimes or bistability. Stochastic models, at first glance, can be understood from the deterministic models. Absolute stability of the homogeneous solution in the deterministic model translates into few, quickly dying disturbances in the stochastic model; large-amplitude instability of the homogeneous solution in the deterministic model translates into metastability of the homogeneous solution in the stochastic model; linear instability of the homogeneous solution in the deterministic model translates into plain instability of the homogeneous solution in the stochastic model. Some other aspects, such as the characterization of the fluctuations which will eventually end a metastable situation, are less well understood.

Both for deterministic and for stochastic models, the parameters can be varied such that the model displays a 2-phase behavior or not. With 1-phase behavior, any initial jam will smear out over time; if started in a closed system, then a 1-phase system will eventually have the same density everywhere. This does not preclude the existence of structures at small scales. When the model parameters are close to the transition from a 1-phase to a 2-phase model, then even a 1-phase model shows aspects of a 2-phase model, which makes, without a good understanding of the whole parameter space, characterization of single models difficult. 1-phase models can still display distinct phases in the presence of bottleneck—then the outflow front is kept stable by the bottleneck, while the inflow front is generically stable as explained in §3.2 (Figure 3).

With respect to reality, the question if 1-phase or 2-phase models are better suited to describe reality cannot be answered at this point. While some researchers believe that spontaneous breakdown does not occur, implying a 1-phase model (albeit possibly a "special" one as explained in §3.3), others are suggesting three (Kerner 1999b, Kerner et al. 2002) or even five (Kerner and Rehborn 1996b, Helbing 2001) phases. The correct answers to these questions will significantly impact traffic flow management: If metastability exists, it makes sense to attempt to stabilize it, for example, via ramp metering or via density-dependent speed limits.

5. FLUID-DYNAMICAL TRAFFIC FLOW MODELS

The main thrust of this paper is to point out how different models for traffic flow behave with respect to traffic jam formation. Many models have a density region where the homogeneous solution is unstable, and the stable solution for these densities is one with traffic jam clusters embedded in laminar flow. In §4, those relations were discussed for microscopic models, i.e., models where each vehicle is represented individually. This section will continue from §3 in looking at continuum models. In fact, the notion of stable, metastable, and unstable homogeneous flow was introduced with fluid-dynamical models.

As usual, fluid-dynamical models can be derived heuristically or systematically, for example, via a kinetic theory. In systematic derivations, one obtains a hierarchy of equations, where each hierarchy adds another partial differential equation of higher order. Although such developments are under way for traffic (see §6.4), in this paper we follow a heuristic route.

All stability arguments are discussed with respect to mathematical models because we are not aware of cases where the computer implementation adds new arguments to the stability considerations. An example of a numerical implementation is given in §6.7.

5.1. First-Order Fluid-Dynamical Models

The term "first-order" refers to the fact that only first-order derivatives are used. In traffic, this refers to the Lighthill-Whitham (1955) equation (2):

$$\partial_t \rho + Q'(\rho)\, \partial_x \rho = 0$$

or its integral form in Equation (1). This equation is also obtained as the first approximation from more systematic derivations (§6.4). Important aspects of the first-order theory for traffic were discussed in §3.2. To recall, the two most important ingredients were:

1. The equation of contuinuity (Equation (1) or (2)), and
2. The assumption that q depends on ρ only, that is, $q(x, t) = Q(\rho(x, t))$.

It was subsequently pointed out that $Q(\rho)$ is related to the driving relation $V(\rho)$ via $Q(\rho) = \rho V(\rho)$. That is, the concept of a desired velocity from car-following models finds its equivalent in $Q(\rho)$ or $V(\rho)$ in the kinematic theory.

5.2. Second-Order Fluid-Dynamical Models I

The first-order equations exhibit shock waves, i.e., discontinuities, so that one has to be careful with numerical integration (Ansorge 1990, Lebacque 1995). For that and other reasons, a variety of proposals has been made to add a second-order diffusion-like term ∂_x^2 to the equations to smooth the shock waves appropriately; see in particular Whitham (1974), Payne (1971, 1979), Kühne (1984), and Kerner and Konhäuser (1993).[6]

One of the simplest models including such a diffusion term is (see Whitham 1974) to replace $v(x, t) = V(\rho(x, t))$ by

$$v = V(\rho) - \frac{D}{\rho}\, \partial_x \rho, \tag{47}$$

where V denotes the equilibrium value of v. Heuristically, such a term means that velocity does not only depend on the density around oneself, but also on the gradient of the density—all things being equal, one drives more slowly when density increases in the driving direction. This is sometimes called the anticipation term.

It can also be justified as follows: In $V(\rho(x,t))$ we are interested in the desired velocity as a function of the density at x. Car-following models, however, use the distance to the car ahead as input, that is, they use the density at $x+\Delta x/2$ as input. This yields

$$V\left(\rho\left(x+\frac{\Delta x}{2}\right)\right) \approx V(\rho(x)) + \frac{\Delta x}{2}\frac{dV}{d\rho}\,\partial_x\rho \approx V(\rho(x)) - \frac{\text{const}}{\rho}\,\partial_x\rho,$$

which is of the same mathematical form as Equation (47). The first approximation is a Taylor expansion, and the second approximation uses $\Delta x \approx 1/\rho$ and $dV/d\rho \approx -\text{const}$. Overall, this yields

$$\partial_t\rho + \left[V(\rho) + \rho\frac{dV}{d\rho}\right]\partial_x\rho = D\,\partial_x^2\rho. \tag{48}$$

5.3. Second-Order Fluid-Dynamical Models II

As a next step, the equation $v = V(\rho)$ or Equation (47) is replaced by a time-dependent equation including an appropriate time constant of velocity adaptation τ, for example

$$\frac{d_v}{dt}v = \frac{1}{\tau}[V(\rho) - v]. \tag{49}$$

Here, we use the so-called total or substantial derivative $d_v/dt = \partial_t + v\,\partial_x$. What this means is that one replaces the assumption that v, via $v = V(\rho)$, *instantaneously* adapts to the surrounding density by the assumption that this adaptation is delayed. The meaning of the substantial derivative is that the equation is valid if one follows the particle. In consequence, the above equation says that the acceleration $d_v v/dt$ of a vehicle equals $1/\tau$ times the term in brackets. The term in brackets, in turn, is $V(\rho) - v$, as we have seen before in car-following models (Equation (20)). Thus, Equation (49) is a translation of the car-following model Equation (20) into fluid dynamics. If one adds the density gradient as in Equation (47), then one obtains

$$\frac{d_v}{dt}v = \frac{1}{\tau}\left[V(\rho) - \frac{D}{\rho}\,\partial_x\rho - v\right]. \tag{50}$$

Now we have two partial differential equations, one for ρ and one for v. To give an overview over a large number of macroscopic models one can rewrite the equations in the following general way (Helbing 1997a, 2001):

$$\frac{d_v}{dt}\rho = -\rho\,\partial_x v + D(\rho)\,\partial_x^2\rho + \xi_1(x,t) \tag{51}$$

for ρ, and for v

$$\frac{d_v}{dt}v = -\frac{1}{\rho}\,\partial_x P + \nu(\rho)\,\partial_x^2 v + \frac{1}{\tau}(V(\rho) - v) + \xi_2(x,t), \tag{52}$$

where P fulfills

$$\partial_x P(\rho) = \frac{dP(\rho)}{d\rho}\,\partial_x\rho. \tag{53}$$

Here, ξ_1 and ξ_2 are noise terms, P is the so-called traffic pressure emerging from Equation (50), the function $V(\rho)$ is the desired velocity as function of density, and $\nu(\rho)$ (Kühne 1984) is a viscosity-like term. Note, however, that these equations themselves cannot be taken literally, because (51) in general would violate the continuity equation.

The equations usually yield one of the numerous different traffic models when setting some of the constants to zero. For example, the Kerner-Konhäuser model is the model with $D = \xi_1 = \xi_2 = 0$, $P = \theta_0\rho$, where θ_0 is a positive constant, and $\nu = \eta/\rho$ (Kerner and Konhäuser 1993, 1994). These equations are in analogy to the Navier-Stokes equations of a compressible fluid, therefore the name fluid-dynamical models. A special case is the limit $\tau \to 0$, resulting in $v = V(\rho)$ *instead* of Equation (52), that is, this limit leads back to the standard theory of kinematic waves (Equation (3)). It is important to state this, because second-order theories are often seen as a replacement of kinematic theories. Instead, as one sees here, they are an extension.

A complete analysis of these models in terms of linear and nonlinear stability is considerably more work than for the first-order models. As was already stated in §3.3, the result is exactly the scheme given in Figure 10, with regions of stable, metastable, and instable homogeneous flow. It was also stated that these regions only exist for certain choices of parameters.

There are critical comments on second- and higher-order models (Cremer and May 1985, Michalopoulos et al. 1987, Leo and Pretty 1992, Daganzo 1995a). One point of criticism is that the introduction of a diffusion term such as in Equation (48) will make cars at the end of a jam drive backwards, which is clearly not realistic. A more general criticism is that second-order models are not any better in making predictions than first-order models. Some of these problems can be solved by doing the derivation of the equations carefully (Aw and Rascle 2000, Klar and Wegener 2000, Aw et al. 2002).

6. OPEN QUESTIONS, OTHER APPROACHES, AND DISCUSSION

6.1. Synchronized Traffic

As pointed out in §2, the fundamental diagram can display strong fluctuations and scatter on the congested branch. This is sometimes called "synchronized traffic." Traffic breakdown by itself, as described as a possibility in earlier sections of this paper, in particular in §3.3, does not explain the data scatter; some mechanism which limits the amplitude of the disturbance would have to be added.

There are, however, many other possibilities which may explain the data scatter. For example, there may be not only one but many stable states (Tomer et al. 2000, Kerner and Osipov 2002), drivers might be sloppy in their car

following (Newell 1962, Wiedemann 1974, Kerner and Rehborn 1996b, Kerner 1999b, Kerner et al. 2002), or braking anticipation partly caused by braking lights of the leading car can lead to data scatter (Knospe et al. 1999a).

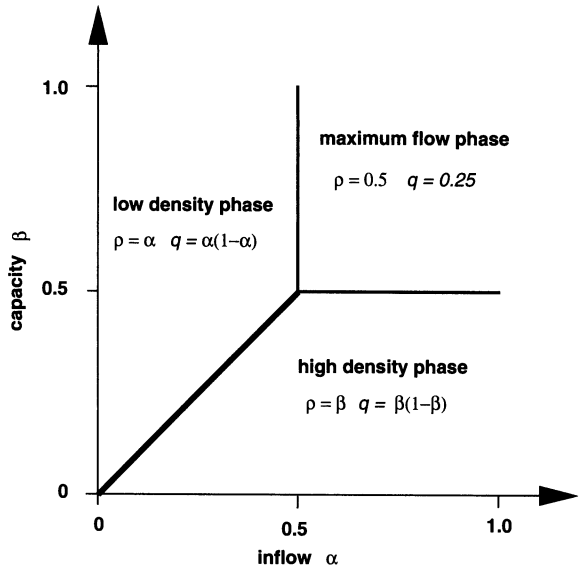
Another possibility is that the data scatter is caused by a downstream bottleneck with relatively high capacity. An example for such a bottleneck is an on-ramp where additional traffic squeezes agressively into the freeway, thus reducing the capacity for through traffic (Lee et al. 1998, Helbing and Treiber 1998, Windover and Cassidy 2001).

Finally, the scatter could be caused by the inhomogeneous driver population or lane-changing phenomena (Daganzo et al. 1999, Daganzo and Cassidy 2001). In fact, the existence of different types of vehicles (cars and trucks) can lead to such a behavior even for single lane traffic (Treiber et al. 1999b). If, in addition, drivers change their behavior in reaction to congested conditions, then the transition to the synchronized regime could be self-reinforcing. For example, if drivers do not accelerate as strongly as usual for some seconds after a slow down this can lead to scatter in the data (Krauß 1998): the idea is that drivers in congested flow give up some space deliberately. This may be captured by an order parameter which accounts for internal correlations in the vehicle motion caused by lane changing (Lubashevsky and Mahnke 1999, 2000). This large number of different possible explanations shows that synchronized flow is an interesting phenomenon that currently is not well understood.

6.2. Open vs. Closed Boundaries

Traffic flow models are often tested in a closed system, i.e., on a loop. The situation is different when open systems are considered (Cheybani et al. 2001, Mitarai and Nakanishi 2000, Popkov and Schütz 1999, Santen and Appert 2001). Different from thermodynamics, where surface effects can be neglected in most cases, an open traffic flow system in one space dimension is controlled sensitively by the conditions on its boundaries. For the so-called totally asymmetric simple exclusion process (TASEP), which serves as a simple substitute for real traffic, some analytically rigorous results are possible. The TASEP is defined, as the CA model above, on a grid. Particles move in one direction and they can jump only to the next place on the grid, provided it is not already occupied. The jumps are stochastic and the next jump is tried after a random waiting time. At one end of the system, say to the left, particles are injected into the system from a reservoir of density α. Particles leave the system at the other end which is coupled to a reservoir of particles that is held constant at some density $1-\beta$. Therefore, α is the (demand for) inflow into the system, while β is the capacity of the outflow end. This system reveals interesting dynamics and serves as one of the rare examples where the microscopic model can be translated completely into its macroscopic analogue (Popkov and Schütz 1999, and references therein).

Figure 16. The phase diagram for the TASEP with open boundaries.



Note. Shown are the different phases, the phase transition lines, and the flows and densities of the phases. The broken lines define another transition in this system which is no phase transition: the type of the shock wave changes at these lines.

The average fundamental diagram of the related *closed* system can be computed exactly:

$$\bar{q}(\bar{\rho}) = C\bar{\rho}(1-\bar{\rho}), \tag{54}$$

where the overbar denotes averaging. This system displays the formation of shock waves, which can be described macroscopically by the nonlinear Burgers equation (11) with noise added.

In case of an *open* system, the two borders are modeled by two cells that are occupied with a certain probability α (modeling the inflow of particles into the system) and β (modeling the outflow). The phase diagram of this system is then controlled exclusively by α and β (Figure 16). If the inflow is much larger than the outflow ($\alpha > \beta$), a high density phase is observed, and a low density phase if $\alpha < \beta$. If both are larger than the maximum possible flow, $q_{\max}$, which is determined by the fundamental diagram of the closed system (Equation (54)), then the system sticks at this maximum flow (the maximum flow phase). These three phases are real phases in the thermodynamic sense; they are connected by nonequilibrium phase transitions. The transition from low density to high density is of first order, while the respective transitions to the maximum flow phase are of second order.

More complicated models with internal degrees of freedom (e.g., real traffic flow models with speed, acceleration, etc.) may display different details (Cheybani et al. 2001, Mitarai and Nakanishi 2000, Santen and Appert 2001). However, the overall picture seems to remain largely unchanged (Cheybani et al. 2001, Namazi et al. 2002).

An interesting point with respect to the discussion of synchronized flow is the following. Suppose traffic would behave like the TASEP above. Then, one can construct

a situation with time dependent $\alpha(t)$ and $\beta(t)$ where the breakdown happens either by decreasing the capacity (β) at the bottleneck downstream, or by first increasing the demand (α) and only then decreasing β. The former leads to a first-order phase transition, while in the latter case two second-order phase transitions would occur. Second-order phase transitions are very hard to observe at all, because all variables change continuously.

6.3. Modeling Headways

When looking at the distribution of headways (i.e., gaps) between vehicles, one finds that the simulation models of §4 generate distributions which are too narrow (Krbalek et al. 2001). Therefore, it make sense to take a closer look at the mechanisms of vehicle interaction. Traffic on a one-lane road can be considered as a gas of interacting fermions, for which electrons are an example: It is not allowed to put two of them at the same location, and exchanging them is forbidden, too. When confined to a ring and interacting via the well-known Coulomb-force, this system arranges itself into a configuration where the headway distribution between electrons is the distribution of the so-called Gaussian Unitary Ensemble (GUE), which follows from the Random Matrix Theory (RMT) (Mehta 1991). The headway distribution of this system, $P_\beta(g)$, can be computed analytically and is approximately given by

$$P_\beta(g) \approx a_\beta g^\beta \exp(-c_\beta g^2), \tag{55}$$

where the constants a_β, c_β are fitting parameters and β is related to a "temperature" T of this system, $\beta = 1/(kT)$. The original RMT allows $\beta = 0, 1, 2, 4$ only. The corresponding dynamical system, the so-called Dyson gas (Dyson 1962), can be described as follows:

$$\dot{v}_i = \frac{\partial}{\partial x_i} \sum_{j \neq i} \log |x_i - x_j|. \tag{56}$$

This equation models a repulsive acceleration that is inversely proportional to the distance. It bears some similarity to car-following equations in the sense that it models a particle-particle interaction which decays with distance.

In Equation (56), a given electron interacts with any other electron in the system. Restricting the interaction to neighboring pairs only, one obtains (Bogomolny et al. 1999)

$$P_\beta(g) = \frac{(\beta+1)^{\beta+1}}{\Gamma(\beta+1)} g^\beta \exp(-(\beta+1)g), \tag{57}$$

which is a rigorous analytical result. In traffic research, this is known as a Pearson type III distribution (see Troutbeck and Brilon 1997, and references therein)

$$P(g) = \frac{1}{b\Gamma(k)} \left(\frac{g-a}{b}\right)^{k-1} \exp\left(-\frac{g-a}{b}\right), \tag{58}$$

which is, for $a = 0$ and $k = 1/b = \beta + 1$, identical to the distribution Equation (57).

Real-world traffic data can be fitted well with these distributions, while, as said before, the microscopic traffic models of §4 deviate significantly towards a much narrower distribution. That is, interaction rules for electrons may have to say something about interaction between vehicles which is not captured by the car-following equations currently in use. A possible explanation is that many car-following models have, besides the collision-preventing repulsion, an effective or actual long-range attraction; in consequence, followers follow leaders at a typical distance which is the same for everybody.

6.4. Micro-Macro Link and Kinetic Theory

It would be nice to have a justifiable connection between microscopic and macroscopic models. Kinetic theory may be able to provide that bridge. Starting from a certain microscopic model that describes traffic as a set of cars driving along a road, one can ask about their corresponding phase space density $f(x, v, t)$. This phase space density is the distribution function to find a car in the interval dx, dv, and dt around (x, v, t). Density and speed can be recovered via the following integrations:

$$\rho(x, t) = \int_0^\infty f(x, v, t)\, dv$$

and

$$v(x, t) = \int_0^\infty v f(x, v, t)\, \mathrm{d}v.$$

One can write down a continuity equation for f, which takes into account the so-called scattering processes that are capable of changing this density:

$$\frac{d_v}{dt} f = \left(\frac{\delta f}{\delta t}\right)_{\text{gain}} - \left(\frac{\delta f}{\delta t}\right)_{\text{loss}}. \tag{59}$$

The scatter processes are approximations of what is really happening microscopically, and, therefore, the resulting equations may be blind for time intervals smaller than a given threshold value. For example, in the original Prigogine approach (Prigogine and Herman 1971), the braking maneuver is condensed into a discontinous change of speed.

From here on, one typically attempts to obtain dynamical equations for $\rho(x, t)$ and $v(x, t)$ in which f does not occur any more. This is usually only possible via further approximations. One eventually obtains a systematic hierarchy of partial differential equations, which means that one obtains models such as Equations (2) and (9), or Equations (51) and (52) systematically as opposed to heuristically. For further information, see Prigogine and Herman (1971), Helbing (1996), Helbing (1996, 1997b), Nelson (1997), and Klar and Wegener (1997).

6.5. More Realistic Single-Lane Models

It was already mentioned earlier in this paper (see also Figure 9) that describing traffic microscopically by car-following equations (§4) may not be sufficient. Some

30 years ago, following a theoretical line of reasoning of so-called psycho-physical modeling, Wiedemann (1974) derived a model that had some fundamental extensions beyond the OVM theory of §4.2.2. Although this model is rather complex, its main idea is easy to describe. First, if a car has a certain distance to the car ahead, and a certain velocity difference Δv, then one can in principle compute what acceleration/deceleration this car has to employ to follow in a certain prescribed distance with $\Delta v = 0$. So far, this is similar to the OVM, albeit on a kinematic basis (i.e., ignoring acceleration and deceleration limits) and therefore crash free. In addition, decisions to change the acceleration are made from time to time only, which is similar to the models in §4.4. The twist now is that there is a certain region in the $(g, \Delta v)$-phase space where the acceleration stays at what it has been when the car entered this region. In other words, the model assumes that there is a "don't care" region in phase space where the driver simply sticks with the decision she/he has last made. This normally means that one gets oscillations in speed and distance when following another car, even if it drives with constant speed: A follower who is too far away will accelerate, will then accelerate through that "don't care" zone, ending up too close to the car ahead, then changing to slowing down, again slowing down at the same rate through that zone, ending up too far away, accelerating again, etc. It is relatively clear that, with this setup, there is no chance to reach the desired headway exactly.

There are other approaches that serve a similar purpose, (e.g., Newell 1962, Treiber et al. 1999a, Tomer et al. 2000, Kerner et al. 2002). These models are capable of producing a similar car-following behavior as the Wiedemann model; the model in Tomer et al. (2000) has additionally the feature that it produces a large number of spatially oscillating states that yield a rather complex fundamental diagram.

6.6. Multilane Traffic

Up to now, we have talked about single-lane traffic. Multilane traffic is of concern, because real traffic consists of nonhomogeneous fleets—and the faster cars want to pass the slower ones.

From a theoretical viewpoint, this makes it considerably more complicated, because many more interactions need to be considered. In addition, it is more difficult to interpret empirical data because there is considerably more scatter. This may be the reason why much of the modeling work concentrates on single-lane traffic. However, there exists a number of investigations of the modeling of multilane traffic.

A number of microscopic rule sets that control the behavior of the car in the simulation have been proposed (Sparmann 1978, Gipps 1986, Wagner et al. 1997, Nagel et al. 1998, Helbing and Huberman 1998, Kerner and Osipov 2002). The line of reasoning behind these models is similar; it divides the lane-changing process into two steps:

- Reason to change lanes (car ahead is too slow, I have to turn off somewhere, I have to change back to my preferred lane after overtaking), and
- Safety criterion (can I do it without causing a crash?).

In general, lane changing destabilizes laminar flow, that is, as a tendency it decreases times to breakdown in Figure 7. This effect can be reduced when drivers make a strong effort to not disturb vehicles behind them when they change lanes (Rickert et al. 1996).

Moreover, the possibility of lane changing greatly enhances the mixing of different types of vehicles, in particular, of vehicles with different desired velocities (Ben-Naim and Krapivsky 1997). Lanes are used asymmetrically, with faster cars on the median lanes. In some countries with a very strong asymmetry, this leads to density inversions near maximum flow, meaning that the density on the median lane is higher than that on the shoulder lane. This means that traffic breakdown is triggered by the left lane (Sparmann 1978, Wagner et al. 1997), possibly followed by a rearrangement of the densities after the breakdown. Although this is speculation, this would help explaining why in such countries flows on the median lane can be reduced by as much as 1/3 when traffic breaks down. The combination of all these effects leads to a large variety of possible dynamical states (Daganzo 1999), which adds to the challenge of correctly interpreting measurements.

6.7. Even Simpler Traffic Models and Regional Traffic Simulations

Real traffic occurs on a network of links, which are connected via intersections. Therefore, additional concepts have to be used to perform simulations of entire networks. For applications such as traffic system management or ITS (Intelligent Transportation Systems), these simulations have to be dynamic to be useful. Even for long-term planning purposes, modeling realistic daily traffic dynamics is probably useful.

A basic requirement for a network simulation is knowledge about the turns that cars take at intersections. This can be achieved either by providing the network with time-dependent turn counts on all intersections, or by using origin-destination information together with a dynamic traffic assignment algorithm leading to a set of routes. In the following, we will assume that we have the routes at our disposition, and will focus on models for traffic dynamics on the links. Gawron (1999) and Bottom (2000) provide further information on algorithms to obtain the routes for dynamic user equilibrium.

Of course, any simulation model which processes routes and generates time-dependent link travel times can be used for the dynamics on the links. In fact, given some moderate computing power, it is possible to use the CA model (TRANSIMS, Rickert and Nagel 2001) from §4.3.3. There are, however, even simpler models for this task, for example, a so-called queue model (Gawron 1998, Bottom 2000) or mixed microscopic/macroscopic approaches

(Wiedemann and Schwerdtfeger 1987, Kates et al. 1998, Mahmassani et al. 1995, DYNAMIT 1999).

Both classes have the same basic ansatz: To approximate the microscopic interactions between the cars by something more simple to increase numerical efficiency by more than one order of magnitude compared to a genuine microscopic model, even a fast one. Despite the fact that this practice of ignoring the microscopic details seems rather crude, a lot of sensible information can be extracted from such models, for example, travel times and even emissions (Eissfeldt and Schrader 2001).

Let us focus first on the queue model. There, any link of the network is regarded as a waiting queue. Cars entering this queue have to stay there until their minimum travel time has elapsed. After that, only a limited numbers of cars can leave this link, bounded by the capacity of the link, and they can leave only if the next link in their route plan has enough free space. Models of this type can simulate traffic in a big city, e.g., the roughly $3 \cdot 10^6$ daily trips in the German city of Cologne, in about 30 minutes computing time, even on moderate hardware.

By dividing long links into smaller pieces, one approaches the realm of the cell-transmission model (Daganzo 1994). Its basic unit is the so-called cell, a piece of road about 100 m long (more precisely, the length is the maximum speed on the link times the update time step), which can hold at maximum N cars. Given the flow capacity of a cell as Q, the dynamical variable is the actual number of cars in a cell i at time t, which is updated as

$$n_i(t+h) = n_i(t) + \delta_i(t) - \delta_{i+1}(t) \tag{60}$$

where

$$\delta_i(t) = \min\{n_{i-1}(t), Q_i, \alpha(N - n_i(t))\}. \tag{61}$$

$\alpha = w/v$ is the quotient between the free-branch slope (v) and the jammed branch slope (w) of the fundamental diagram. The cell-transmission model converges for $h \to 0$ to the macroscopic Lighthill-Whitham equation (2), therefore, it can be regarded as a discretization of the Lighthill-Whitham equation.

An alternative method for the integration of the Lighthill-Whitham equation is the *Godunov Scheme* (Godunov 1959, Lebacque 1995). In fact, any partial differential equation (PDE) suitable for traffic (§5) can be discretized. Because discretization can be done in a number of ways (e.g., Chronopolous and Michalopoulos 1991), many possible traffic flow simulation models are obtained by this procedure.

Discretized fluid-dynamical models do not contain particles (=vehicles) whose trajectories are needed for many problems of current interest. It is, however, possible to move actual particles via the dynamics of the cell-transmission model or any other discretized fluid-dynamical model (Daganzo 1995b). The mathematically correct way to do this is the following: Any fluid-dynamical equations can be regarded either in their Eulerian or their Lagrangian form. The former means the PDE, while the latter allows the substitution of the PDEs by a system of ordinary differential equations (ODEs) for test particles flowing with the flow. This is the trick how to have microscopic particles within macroscopic equations: they follow the field that is generated by themselves. Basically, the left-hand side of the PDE $\partial_t v(x,t) + v\,\partial_x v(x,t)$ transforms back into $\dot{v}$ while the partial derivatives on the right-hand side are exchanged for corresponding averages over neighbouring particles (Lee et al. 2001, Rosswog and Wagner 2002). This leads to ODEs which are sometimes hard to solve. Replacing these ODEs by some approximations can lead to a numerically very efficient scheme. This is the approach used, e.g., in DYNEMO (Wiedemann and Schwerdtfeger 1987) and other simulators (Kates et al. 1998, DYNAMIT 1999, Mahmassani et al. 1995).

7. SUMMARY

This paper reviews traffic simulation models with a special emphasis on the understanding of traffic jam formation. The text goes from phenomenological observations via general insight into traffic jam formation to microscopic and fluid-dynamical traffic models and, finally, to open questions.

In the last ten years, some new attempts have been undertaken to understand how traffic flows. These new methods include cellular automata models, nonlinear methods for fluid-dynamical equations, and a second view on the concept of phase transitions in the context of traffic. As was pointed out in this paper, many models currently in use display similar dynamics of traffic flow breakdown, i.e., the emergence of traffic jams. The emergence of traffic jams in these models is either deterministic, which implies an external disturbance, or stochastic, which means that they can emerge spontaneously. Field measurements and their interpretations are nonunequivocal with respect to the question of the existence of spontaneous breakdown. In our view, the current data situation is not sufficient to resolve the issue. In particular, more experimental work needs to be done which includes the whole surrounding situation, and at the same time it needs to be clarified which aspects of a possible phase transition could be confirmed or disproved in real-world situations. The search for "spontaneous" traffic breakdown is not enough.

In addition, there are further empirical findings, sometimes labeled "synchronized" traffic. No consensus about their origin has been reached. Furthermore, there are many additional open questions, ranging from multilane traffic to good models for regional traffic simulations.

Let us close with a few recommendations about what could further improve the modeling of traffic flow. First, good and publicly available datasets are needed. For instance, having the complete car trajectories in a freeway intersection with a time resolution of 10 Hz would give the opportunity to test models more thoroughly and to understand the interactions between cars in more detail. There are a number of endeavors approaching this world wide (e.g., Coifman et al. 2000, Ruhe 2002, DLR 2002). Second, what

is as important as good datasets are clear and good ideas for testing the models. Such testing suites are currently being developed (e.g., DYNAMIT 1999, DLR/VF 2001). Finally, more research is needed to better understand the influence of driving strategies on driving behavior. These strategies may change according to road, weather, and traffic conditions according to the vigilance of the driver and can be different for different drivers. Here, an interdisciplinary ansatz involving disciplines such as traffic engineering, operations research, physics, mathematics, and psychology is needed.

ENDNOTES

1. We normally cite recent work. For a historical perspective see, e.g., Daganzo (1999).
2. The traffic version of the Burgers equation is obtained from the standard version of the Burgers equation by a mirror transformation and a transformation onto the free speed v_f of the vehicles.
3. We use the term "coarse-grained discrete" to distinguish from the "fine-grained discreteness" that floating point numbers in computers have.
4. Note, however, that although this formulation looks fairly general, it includes an assumption about driving behavior. Namely, the driver does not accelerate during the reaction time. More precisely, one should write $d(v_n(t+\tau))+x_n(t+\tau)+l \leqslant d(v_{n-1})+x_{n-1}$ (Mahut 1999).
5. According to simulations performed by the authors, the deterministic version runs in 12.5 million vehicle updates per second in C++ with gcc; the stochastic version is about 9.7 million vehicle updates per second, respectively, on a 1.1 GHz PIII machine.
6. In fluid dynamics, viscosity describes the friction that causes the macroscopic flow to dissolve into microscopic chaos. There are also other justifications for diffusion-like terms, such as kinematic viscosity or numerical viscosity. In traffic, the diffusion term has originally been introduced to avoid the slopes of the fronts to become infinitely steep. This can, for example, be justified via the fact that the acceleration of the vehicles is limited.

ACKNOWLEDGMENTS

The Landschaftsverband Rheinland, the Northrhine-Westfalia Ministry for Economy and Transport, and the city of Cologne provided data that helped to improve our understanding of the relationship between models and reality. We also thank Nils Eissfeldt, Dirk Helbing, Claudia Hertfelder, Georg Hertkorn, Boris Kerner, Reinhart Kühne, Andreas Schadschneider, and Michael Schreckenberg for discussions and input. Special thanks go to Mike Cassidy and Carlos Daganzo, who read and commented on several versions of the manuscript. Although or perhaps because they do not share all of our views, they considerably helped to make the paper more balanced and more well-grounded both in theory and with respect to field data. Their helpful and consistent comments improved the quality of this paper. The responsibility for all errors remains, as usual, with the authors.

REFERENCES

Acha-Daza, J. A., F. L. Hall. 1993. Graphical comparison of predictions for speed given by catastrophe theory and some classic models. *Transportation Res. Record* **1398** 119–124.

Ansorge, R. 1990. What does the entropy condition mean in traffic flow theory? *Transportation Res. B* **24** 133–143.

Aw, A., M. Rascle. 2000. Resurrection of second order models of traffic flow. *SIAM J. Appl. Math.* **60** 916–938.

——, A. Klar, T. Materne, M. Rascle. 2002. Derivation of continuum traffic flow models from microscopic follow-the-leader models. *SIAM J. Appl. Math.* **63**(1) 259–278.

Bachmann, E. 1938. Leistungsfähigkeit von Strassen und Plätzen. *Strasse und Verkehr* **24** 237–246. In German.

Bando, M., K. Hasebe, A. Nakayama, A. Shibata, Y. Sugiyama. 1994. Structure stability of congestion in traffic dynamics. *Japan J. Indust. Appl. Math.* **11**(2) 203–223.

——, ——, ——, ——, ——. 1995. Dynamical model of traffic congestion and numerical simulation. *Physical Rev. E* **51**(2) 1035–1042.

Barlovic, R., L. Santen, A. Schadschneider, M. Schreckenberg. 1998. Metastable states in CA models for traffic flow. *Eur. Physical J. B* **5**(3) 793–800.

Barrett, C. L., M. Wolinsky, M. W. Olesen. 1996. Emergent local control properties in particle hopping traffic simulations. D. E. Wolf, M. Schreckenberg, A. Bachem, eds. *Traffic and Granular Flow.* World Scientific, Singapore, 169–173.

Ben-Naim, E., P. L. Krapivsky. 1997. Stationary velocity distributions in traffic flows. *Physical Rev. E* **56**(6) 6680–6686.

Biham, O., A. Middleton, D. Levine. 1992. Self-organization and a dynamical transition in traffic-flow models. *Physical Rev. A* **46** R6124–R6127.

Bogomolny, E. B., U. Gerland, C. Schmit. 1999. Models of intermediate spectral statistics. *Physical Rev. E* **59** R1315–R1318.

Bottom, J. A. 2000. Consistent anticipatory route guidance. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Brankov, J. G., V. B. Priezzhev, A. Schadschneider, M. Schreckenberg. 1996. The Kasteleyn model and a cellular automaton approach to traffic flow. *J. Physics A Math. General* **29** L229–L235.

Brilon, W., N. Wu. 1999. Evaluation of cellular automata for traffic flow simulation on freeway and urban streets. W. Brilon, F. Huber, M. Schreckenberg, H. Wallentowitz, eds. *Traffic and Mobility: Simulation–Economics–Environment.* Springer, Heidelberg, Germany, 163–180.

Burgers, J. M. 1974. *The Nonlinear Diffusion Equation.* Reidel, Dordrecht, The Netherlands.

Cassidy, M. J. 1998. Bivariate relations in nearly stationary highway traffic. *Transportation Res. B* **32** 49–59.

Ceder, A. 1977. A time-sequence analysis of a two-regime traffic flow model. *Proc. 7th Internat. Sympos. Transportation Traffic Theory.* Kyoto, Japan, 145–165.

Cheybani, S., J. Kertesz, M. Schreckenberg. 2001. Nondeterministic Nagel-Schreckenberg traffic model with open boundary conditions. *Physical Rev. E* **63** 1–9.

Chowdhury, D., L. Santen, A. Schadschneider. 2000a. Statistical physics of vehicular traffic and some related systems. *Physics Rep.* **329**(4–6) 199–329.

——, J. Kertész, K. Nagel, L. Santen, M. Schadschneider. 2000b. Comment on: "Critical behavior of a traffic flow model." *Physical Rev. E* **61**(3) 3270–3271.

Chronopolous, A., P. Michalopoulos. 1991. Traffic flow simulation through parallel processing. Final research report. Technical report, Center for Transportation Studies, University of Minnesota, Minneapolis, MN.

Coifman, B., D. Lyddy, A. Skabardonis. 2000. The Berkeley Highway Laboratory—Building on the I-880 field experiment. *2000 IEEE Intelligent Transportation Systems Proc.*, Dearborn, MI, 5–10.

Cremer, M., J. Ludwig. 1986. A fast simulation model for traffic flow on the basis of Boolean operations. *Math. Comput. in Simulation* **28** 297–303.

——, A. D. May. 1985. An extended traffic model for freeway control. Technical research report UCB-IST-RR-85-7, Institute of Transportation Studies, University of California, Berkeley, CA.

Daganzo, C. F. 1994. The cell transmission model: A simple dynamical representation of highway traffic consistent with the hydrodynamic theory. *Transportation Res. B* **28**(4) 269–287.

——. 1995a. Requiem for second-order fluid approximations of traffic flow. *Transportation Res. B* **29** 277–286.

——. 1995b. The cell transmission model. Part II: Network traffic. *Transportation Res. B* **29** 79–93.

——. 1999. A behavioral theory of multi-lane traffic flow. I: Long homogeneous freeway sections. Technical research report UCB-IST-RR-99-6, Institute of Transportation Studies, University of California, Berkeley, CA.

——, M. Cassidy. 2001. Personal communication.

——, ——, R. L. Bertini. 1999. Possible explanations of phase transitions in highway traffic. *Transportation Res. A* **33** 365–379.

DLR. 2002. Multisat Web Services. www.multisat.dlr.de.

DLR/VF. 2001. Institute for Transportation Research. ivf.dlr.de/clearing.

DYNAMIT. 1999. Massachusetts Institute of Technology, Cambridge, MA. its.mit.edu.

Dyson, F. J. 1962. Statistical theory of energy levels of complex systems, parts I–III. *J. Math. Physics* **3** 140–175.

Edie, L. C. 1961. Car-following and steady-state theory for noncongested traffic. *Oper. Res.* **9** 66–77.

Eissfeldt, N., R. Schrader. 2001. Calculation of street traffic emissions with a queuing model. *J. Comput. Tech.* Submitted.

Gawron, C. 1998. An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. *Internat. J. Modern Physics C* **9**(3) 393–407.

Gawron, Ch. 1999. Simulation-based traffic assignment. Ph.D. thesis, University of Cologne, Cologne, Germany.

Gazis, D. C., R. Herman, E. W. Montroll, R. W. Rothery. 1961. Nonlinear follow-the-leader models of traffic flow. *Oper. Res.* **9** 545–560.

Gerlough, D. L. 1956. Simulation of freeway traffic by an electronic computer. F. Burggraf, E. M. Ward, eds. *Proc. 35th Annual Meeting*. Highway Research Board, National Research Council, Washington, DC, 543–547.

——, M. J. Huber. 1975. Traffic flow theory. Special report 165. Transportation Research Board, National Research Council, Washington, DC.

Gipps, P. G. 1981. A behavioural car following model for computer simulation. *Transportation Res. B* **15** 105–111.

——. 1986. A model for the structure of lane-changing decisions. *Transportation Res. B* **20B**(5) 403–414.

Godunov, S. K. 1959. A difference scheme for numerical computation of discontinuous solutions of equations of fluid dynamics. *Math. Sbornik* **47** 271–306. In Russian.

Gray, L., D. Griffeath. 2001. The ergodic theory of traffic jams. *J. Statist. Physics* **105**(3–4) 413–452.

Greenshields, B. D. 1935. A study of traffic capacity. *Proc. Highway Res. Board*, vol. 14. Washington, DC, 448–477.

Haberman, R. 1977. *Mathematical Models in Mechanical Vibrations, Population Dynamics, and Traffic Flow*. Prentice-Hall, Englewood Cliffs, NJ.

Haight, F. A. 1963. *Mathematical Theories of Traffic Flow*. Academic Press, New York.

Hall, F. L., K. Agyemang-Duah. 1991. Freeway capacity drop and the definition of capacity. *Transportation Res. Record* **1320** 91–98.

——, V. F. Hurdle, J. H. Banks. 1992. A synthesis of recent work on the nature of speed-flow and flow-occupancy (or density) relationships on freeways. *Transportation Res. Record* **1365** 12–18.

Helbing, D. 1996. Gas-kinetic derivation of Navier-Stokes-like traffic equations. *Physical Rev. E* **53**(3) 253–282.

——. 1997. *Verkehrsdynamik*. Springer, Berlin, Germany. In German.

——. 2001. Traffic and related self-driven many-particle systems. *Rev. Modern Physics* **73** 1067–1141.

——, B. A. Huberman. 1998. Coherent moving states in highway traffic. *Nature* **396** 738–740.

——, B. Tilch. 1998. Generalized force model of traffic dynamics. *Physical Rev. E* **58** 133–138.

——, M. Treiber. 1998. Gas-kinetic-based traffic model explaining observed hysteretic phase transition. *Physical Rev. Lett.* **81**(14) 3042–3045.

Herman, R., E. W. Montroll, R. B. Potts, R. W. Rothery. 1959. Traffic dynamics: Analysis of stability in car following. *Oper. Res.* **7** 86–106.

Janz, S. 1998. Mikroskopische minimalmodelle des strassenverkehrs. Master's thesis, University of Cologne, Cologne, Germany. In German.

Jost, D. 2002. Breakdown and recovery in traffic flow models. Master's thesis, ETH Zurich, Switzerland. e-collection. ethbib.ethz.ch.

——, K. Nagel. 2003. Probabilistic traffic flow breakdown in stochastic car following models. Paper 03-4266, Transportation Research Board Annual Meeting, Washington, DC. sim.inf.ethz.ch/papers. Shorter version to be published in *Transportation Res. Record*.

Kates, R., K. Bogenberger, M. Hoops. 1998. Mesoscopic simulation with ANIMAL. M. Schreckenberg, D. E. Wolf, eds. *Traffic and Granular Flow '97*. Springer, Berlin, Germany, 453–467.

Kerner, B. S. 1998. Experimental features of self-organization in traffic flow. *Physical Rev. Lett.* **81** 3797–3800.

——. 1999a. Congested traffic flow: Observations and theory. *Transportation Res. Record* **1678** 160–167.

——. 1999b. Phase transitions in traffic flow. D. Helbing, H. J. Hermann, M. Schreckenberg, D. E. Wolf, eds. *Traffic and Granular Flow '99*. Springer, Berlin, Germany, 253–284.

——. 1999c. The physics of traffic. *Physics World* **8** 25–30.

——, S. L. Klenov. 2002. *J. Physics A: Math. General* **35** L31–L43.

——, P. Konhäuser. 1993. Cluster effect in initially homogeneous traffic flow. *Physical Rev. E* **48**(4) R2335–R2338.

——, ——. 1994. Structure and parameters of clusters in traffic flow. *Physical Rev. E* **50**(1) 54–83.

——, X. Osipov. 2002. Cluster effect in initially homogeneous traffic flow. *J. Physics A: Math. General* **35** L31–L43.

——, H. Rehborn. 1996a. Experimental features and characteristics of traffic jams. *Physical Rev. E* **53**(2) R1297–R1300.

——, ——. 1996b. Experimental properties of complexity in traffic flow. *Physical Rev. E* **53**(5) R4275–R4278.

——, ——. 1997. Experimental properties of phase transitions in traffic flow. *Physical Rev. Lett.* **79**(20) 4030–4033.

——, S. L. Klenov, D. E. Wolf. 2002. Cellular automata approach to three-phase traffic theory. *J. Physics A: Math. General* **35**(47) 9971–10013.

Kikuchi, M., et al. 2001. Video presented at Traffic and Granular Flow (TGF)'01, Nagoya, Japan.

Klar, A., R. Wegener. 1997. Enskog-like kinetic models for vehicular traffic. *J. Statist. Physics* **87**(1–2) 91–114.

——, ——. 2000. Kinetic derivation of macroscopic anticipation models for vehicular traffic. *SIAM J. Appl. Math.* **60**(5) 1749–1766.

Knospe, W., L. Santen, A. Schadschneider, M. Schreckenberg. 1999a. CA models for traffic flow: Comparison with empirical single-vehicle data. D. Helbing, H. J. Hermann, M. Schreckenberg, D. E. Wolf, eds. *Traffic and Granular Flow '99*. Springer, Berlin, Germany, 366–376.

——, ——, ——, ——. 1999b. Single-vehicle data of highway traffic. *Physical Rev. E* **60** 6480–6490.

——, ——, ——, ——. 2000. Towards a realistic microscopic description of highway traffic. *J. Physics A Math. General* **33** L477–L485.

——, ——, ——, ——. 2002. Single-vehicle data of highway traffic: Microscopic description of traffic phases. *Physical Rev. E* **65**(056133) 1–16.

Koshi, M., M. Iwasaki, I. Ohkura. 1983. Some findings and an overview of vehicular flow characteristics. *Proc. 8th Internat. Sympos. Transportation Traffic Theory.* Elsevier, Amsterdam, The Netherlands, and Toronto, Ontario, Canada, 403–426.

Krauß, S. 1998. Microscopic modelling of traffic flow: Investigation of collision free vehicle dynamics. Ph.D. thesis, University of Cologne, Cologne, Germany.

——, P. Wagner, C. Gawron. 1997. Metastable states in a microscopic model of traffic flow. *Physical Rev. E* **55** 5597–5605.

Krbalek, M., P. Seba, P. Wagner. 2001. Headways in traffic flow—Remarks from a physical perspective. *Physical Rev. E* **64**(066119) 1–7.

Kühne, R. D. 1984. Macroscopic freeway model for dense traffic stop-start waves and incident detection. J. Volmuller, R. Hamerslag, eds. *Proc. 9th Internat. Sympos. Transportation Traffic Theory*. VNU Science Press, Utrecht, The Netherlands, 21–42.

Lax. P. D. 1973. Hyperbolic systems of conservation laws and the mathematical theory of shock waves, no. 11. *Regional Conference Series Appl. Math.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Lebacque, J. P. 1995. The Godunov scheme and what it means for first order traffic flow models. Technical report 95-48, Centre de'Enseignement et de Recherche en Mathmatiques, Informatique et Calcul Scientifique (CERMICS), Marne la Vallée, France. cermics.enpc.fr/reports/.

——, J. B. Lesort. 1999. Macroscopic traffic flow models: A question of order. A. Ceder, ed. *Transportation and Traffic Theory.* Pergamon Press, Amsterdam, The Netherlands, 3–25.

Lee, H. Y., H.-W. Lee, D. Kim. 1998. Origin of synchronized traffic flow on highways and its dynamic phase transitions. *Physical Rev. Lett.* **81**(5) 1130–1133.

——, ——, ——. 2001. Macroscopic traffic models from microscopic car-following models. *Physical Rev. E* **64**(056126) 1–12.

Leo, C. J., R. L. Pretty. 1992. Numerical simulation of macroscopic continuum traffic models. *Transportation Res. B* **23** 79–94.

Leutzbach, W. 1988. *Introduction to the Theory of Traffic.* Springer, Berlin, Germany.

Lighthill, M. J., J. B. Whitham. 1955. On kinematic waves. I: Flow movement in long rivers. II: A theory of traffic flow on long crowded roads. *Proc. Royal Soc. A* **229** 281–345.

Lubashevsky, I. A., R. Mahnke. 1999. Order parameter as an additional state variable of unstable traffic flow. D. Helbing, H. J. Hermann, M. Schreckenberg, D. E. Wolf, eds. *Traffic and Granular Flow '99*. Springer, Berlin, Germany, 377–382.

——, ——. 2000. Order-parameter model for unstable multilane traffic flow. *Physical Rev. E* **62**(5) 6082–6093.

Mahmassani, H. S., T. Hu, R. Jayakrishnan. 1995. Dynamic traffic assignment and simulation for advanced network informatics (DYNASMART). N. H. Gartner, G. Improta, eds. *Urban Traffic Networks: Dynamic Flow Modeling and Control.* Springer, Berlin/New York.

Mahut, M. 1999. Speed-maximizing car following models and macroscopic curves based on save stopping rules. Preprint number 990351, Transportation Research Board Annual Meeting, Washington, DC.

Mehta, M. L. 1991. *Random Matrices*, 2nd ed. Academic Press, San Diego, CA.

Michalopoulos, P. G., J. Lin, D. E. Beskos. 1987. Integrated modelling and numerical treatment of traffic flow. *Appl. Math. Modelling* **2** 447–457.

Mika, H. S., J. B. Kreer, L. S. Yuan. 1969. Dual-mode behaviour of freeway traffic. *Highway Res. Records* **279** 1–13.

Mitarai, N., H. Nakanishi. 2000. Spatio-temporal structure of traffic flow in a system with an open boundary. *Physical Rev. Lett.* **85** 1766–1769.

Muñoz, J. C., C. F. Daganzo. 2003. Structure of the transition zone behind freeway queues. *Transportation Sci.* **37**(3) 312–329.

Nagel, K. 1994. Life-times of simulated traffic jams. *Internat. J. Modern Physics C* **5**(3) 567–580.

——. 1996. Particle hopping models and traffic flow theory. *Physical Rev. E* **53**(5) 4655–4672.

——. 1999. From particle hopping models to traffic flow theory. *Transportation Res. Records* **1644** 1–9.

——, M. Paczuski. 1995. Emergent traffic jams. *Physical Rev. E* **51** 2909–2918.

——, M. Schreckenberg. 1992. A cellular automaton model for freeway traffic. *J. Physique I France* **2** 2221–2229.

——, C. Kayatz, P. Wagner. 2003. Breakdown and recovery in traffic flow models. Y. Sugiyama et al., eds. *Traffic and Granular Flow '01*. Springer, Heidelberg, Germany. In press.

——, D. E. Wolf, P. Wagner, P. Simon. 1998. Two-lane traffic rules for cellular automata: A systematic approach. *Physical Rev. E* **58**(2) 1425–1437.

——, P. Stretz, M. Pieck, S. Leckey, R. Donnelly, C. L. Barrett. 1997. TRANSIMS traffic flow characteristics. Los Alamos unclassified report (LA-UR) 97-3530. Los Alamos National Laboratory, Los Alamos, NM. transims.tsasa.lanl.gov.

Namazi, A., N. Eissfeldt, P. Wagner, A. Schadschneider. 2002. Boundary-induced phase transitions in a space-continuous traffic model with non-unique flow-density relation. *Eur. Physical J. B* **30**(4) 559–570.

Nelson, P. 1997. Kinetic theories. N. H. Gartner, C. J. Messer, A. Rathi, eds. *Monograph on Traffic Flow Theory*. www.tfhrc.gov/its/tft/tft.htm.

——. 2000. Synchronized traffic flow from a modified Lighthill-Whitham model. *Physical Rev. E* **62**(2) R6052–R6055.

Newell, G. F. 1955. Mathematical models of freely moving traffic. *J. Oper. Res. Soc. Amer.* **3** 176–188.

——. 1961. Nonlinear effects in the dynamics of car following. *Oper. Res.* **9**(2) 209–229.

——. 1962. Theories of instability in dense highway traffic. *J. Oper. Res. Soc. Japan* **5** 9–54.

——. 1963. Instability in dense highway traffic, a review. *Second Sympos. Theory Traffic Flow.* OECD, London, U.K., 73–83.

——. 2002. A simplified car-following theory: A lower order model. *Transportation Res. B* **36** 195–205.

Owren, B., M. Zennaro. 1992. Derivation of efficient, continuous, explicit Runge-Kutta methods. *SIAM J. Sci. Statist. Comput.* **13** 1488–1501.

Payne, H. J. 1971. Models of freeway traffic and control. G. A. Bekey, ed. *Mathematical Models of Public Systems*, no. 28. *Simulation Council Proc.* Simulation Council, La Jolla, CA, 51–61.

——. 1979. A critical review of macroscopic freeway model. W. S. Levine, ed. *Research Directions in Computer Control of Urban Traffic Systems*. American Society of Civil Engineers, New York, 251–265.

Persaud, B. N., F. L. Hall. 1989. Catastrophe theory and patterns in 30-second freeway traffic data—Implications for incident detection. *Transportation Res. A* **23A**(2) 103–113.

Popkov, V., G. M. Schütz. 1999. Steady-state selection in driven diffusive systems with open boundaries. *Europhysics Lett.* **48** 257–263.

Prigogine, I. 1961. A Boltzmann-like approach to the statistical theory of traffic flow. *Theory of Traffic Flow*. R. Herman, ed. Elsevier, Amsterdam, The Netherlands, 158–164.

——, R. Herman. 1971. *Kinetic Theory of Vehicular Traffic*. Elsevier, New York.

Rajewsky, N., M. Schreckenberg. 1997. Exact results for one-dimensional cellular automata with different types of updates. *Physica A* **245**(1–2) 139–144.

Reiss, H., A. D. Hammerich, E. W. Montroll. 1986. Thermodynamic treatment of nonphysical systems: Formalism and an example (single-lane traffic). *J. Statist. Physics* **42**(3/4) 647–687.

Reuschel, A. 1950. Fahrzeugbewegung in der Kolonne bei gleichförmig beschleunigtem oder verzögertem Leitfahrzeug. *Zeitschrift des österreichischen Ingenieur und Architekten-vereins.* 7/8, 9/10; 59–62, 73–77. Vienna, Austria. In German.

Richards, P. I. 1956. Shockwaves on the highway. *Oper. Res.* **4** 42–51.

Rickert, M., K. Nagel. 2001. Dynamic traffic assignment on parallel computers in TRANSIMS. *Future Generation Comput. Systems* **17**(5) 637–648.

——, ——, M. Schreckenberg, A. Latour. 1996. Two lane traffic simulations using cellular automata. *Physica A* **231**(4) 534–550.

Rosswog, S., P. Wagner. 2002. Towards a macroscopic modelling of the complexity in traffic flow. *Physical Rev. E* **65**(036106) 1–9.

Roters, L., S. Lübeck, K. D. Usadel. 1999. Critical behavior of a traffic flow model. *Physical Rev. E* **59**(3) 2672–2676.

Ruhe, M. 2002. *Luftgestütztes Verkehrsmonitoring*. ivf.dlr.de/berlin/projekte/luftgest_verkmonitor.htm.

Santen, L., C. Appert. 2001. Boundary induced phase transitions in driven lattice gases with metastable states. *Physical Rev. Lett.* **86** 2498–2501.

Sasvari, M., J. Kertesz. 1997. Cellular automata models of single lane traffic. *Physical Rev. E* **56**(4) 4104–4110.

Sauermann, G., H. J. Herrmann. 1998. A 1d traffic model with threshold parameters. D. E. Wolf, M. Schreckenberg, eds. *Traffic Granular Flow '97*. Springer, Heidelberg, Germany, 481–486.

Schadschneider, A., M. Schreckenberg. 1997. Traffic flow models with "slow-to-start" rules. *Ann. Physics* **6**(7) 541–551.

Sheffi, Y. 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, NJ.

SMARTEST. Simulation modelling applied to road transport European scheme tests. www.its.leeds.ac.uk/projects/smartest/index.html.

Sparmann, U. 1978. *Spurwechselvorgänge auf zweispurigen BAB-Richtungsfahrbahnen*, no. 263. Forschung Straßenbau und Straßenverkehrstechnik. Bundesminister für Verkehr, Bonn-Bad Godesberg, Germany.

Stauffer, D. 1985. *Introduction to Percolation Theory*. Taylor & Francis, London, U.K., and Philadelphia, PA.

——. 1991. Computer simulations of cellular automata. *J. Physics A* **24** 909–927.

Takayasu, M., H. Takayasu. 1993. Phase transition and $1/f$ type noise in one dimensional asymmetric particle dynamics. *Fractals* **1** 860–866.

Tomer, E., L. Safonov, S. Havlin. 2000. Presence of many stable non-homogeneous states in an inertial car-following traffic model. *Physical Rev. Lett.* **84** 382–385.

TRANSIMS. TRansportation ANalysis and SIMulation System. transims.tsasa.lanl.gov.

Treiber, M., A. Hennecke, D. Helbing. 1999a. Derivation, properties, and simulation of a gas-kinetic-based, non-local traffic model. *Physical Rev. E* **59** 239–253.

——, ——, ——. 1999b. Microscopic simulation of congested traffic. D. Helbing, H. J. Hermann, M. Schreckenberg, D. E. Wolf, eds. *Traffic and Granular Flow '99*. Springer, Berlin, Germany, 366–376.

Treiterer, J., J. A. Myers. 1974. The hysteresis phenomenon in traffic flow. D. J. Buckley, ed. *Proc. 6th ISTT*, Sydney, Australia. Elsevier, New York, 13–38.

Troutbeck, R. J., W. Brilon. 1997. Unsignalized intersection theory. N. H. Gartner, C. J. Messer, A. Rathi, eds. *Monograph on Traffic Flow Theory*. TFHRC. www.tfhrc.gov/its/tft/tft.htm.

von Neumann, J. 1966. *Theory of Self-Reproducing Automata*. Edited and completed by A. W. Burks. University of Illinois Press, Urbana, IL.

Wagner, P., K. Nagel, D. E. Wolf. 1997. Realistic multi-lane traffic rules for cellular automata. *Physica A* **234** 687–698.

Wehner, B. 1939. *Die Leistungsfähigkeit von Strassen, Forschungsarbeiten aus dem Strassenwesen*. Volk und Reich, Berlin, Germany. In German.

Whitham, G. B. 1974. *Linear and Nonlinear Waves*. Wiley, New York.

Wiedemann, R. 1974. Simulation des Straßenverkehrsflußes. Technical report, Institut für Verkehrswesen, Universität Karlsruhe, Karlsruhe, Germany.

——, T. Schwerdtfeger. 1987. Makroskopisches simulationsmodell für schnellstrassennetze mit berücksichtigung von einzelfahrzeugen. Technical report, Bundesministerium für Verkehr, Abt. Strassenbau, Bonn, Germany. In German.

Wilson, R. E. 2001. An analysis of Gipps' model of highway traffic. *IMA J. Appl. Math.* **66**(5) 509–537.

Windover, J. R., M. J. Cassidy. 2001. Some observed details for freeway traffic evolution. *Transportation Res. A* **A35** 881–894.

Wolf, D. 1999. Cellular automata for traffic simulations. *Physica A* **263** 438–451.

Wolfram, S. 1986. *Theory and Applications of Cellular Automata*. World Scientific, Singapore.

Wu, N. 2000. Verkehr auf schnellstraßen im fundamentaldiagramm—ein neues modell und seine anwendungen. *Straßenverkehrstechnik* **44** 7–12. In German.

Yukawa, S., M. Kikuchi. 1999. Coupled-map modeling of one-dimensional traffic flow. D. Helbing, H. J. Hermann, M. Schreckenberg, D. E. Wolf, eds. *Traffic and Granular Flow '99*. Springer, Berlin, Germany, 319–334.

Zhang, X., D. F. Jarrett. 1997. Stability analysis of the classical car-following model. *Transportation Res. B.* **31** 443–462.