INSERT CHAPTER NUMBER

A CRITICAL COMPARISON OF THE KINEMATIC-WAVE MODEL WITH OBSERVATIONAL DATA

Kai Nagel

Institute for Land and Sea Transport Systems

Technical University of Berlin

Sek SG12, Salzufer 17-19

10587 Berlin, Germany

Paul Nelson

Department of Computer Science

Texas A&M University

College Station, Texas 77843-3133

U.S.A.

ABSTRACT

There is considerable disagreement to which extent different versions of the kinematic wave model (KWM) can explain/predict certain effects of traffic flow dynamics. In this paper, first the traditional KWM, with a strictly convex fundamental diagram (FD), is explored. It is found that, after careful examination, such a model can explain most but not all observations of traffic flow dynamics. KWMs with piecewise linear segments are closer to the observations, but do still leave open questions, most notably the spontaneous emergence of jams. KWMs with concave segments are less well investigated in this context, but do not seem to offer a significantly better alternative.

An important conclusion of our study is that many existing data sets are not well enough documented to be of help to answer critical questions. For example, there seems to be only one empirical observation of spontaneous traffic breakdown, and it is neither published nor widely disseminated. Similarly, there are very few investigations that include the spatial dimension, without which many questions cannot be answered.

In: Transportation and Traffic Theory -- Flow, Dynamics, and Human Interaction (Proc. 16th ISTTT), H. Mahmassani (ed.), Elsevier, 2005, p. 145-164.

1. INTRODUCTION

The objectives of this work are to:

- critically summarize the predictive capabilities and predictions of the classical kinematic-wave model (KWM) of Lighthill and Whitham (1955) and Richards (1956), where by "classical" is intended the case of a strictly convex fundamental diagram (traffic stream model);
- critically summarize the extant data that seem to represent phenomena that respectively are and are not reproduced by those predictions, especially with a view toward identifying those disagreements between observed phenomena and the classical KWM that seem most firmly established;
- briefly discuss prospect for reproducing some of the phenomena deemed most inconsistent with the classical KWM via use of a fundamental diagram (FD) that is not strictly convex.

It is important to note that whether an allegedly observed phenomenon is ``firmly established'' is decidedly a subjective judgment. An element of firmness certainly is reproduction by multiple independent observers; however, that is not sufficient, because it is possible for multiple observers to build upon the same mistaken assumption or technique. It is therefore possible for two reasonable people to disagree as to whether existence of an alleged phenomenon has been firmly established, and indeed the two authors themselves do not completely agree in all instances. Our approach in cases where either of us has doubts is simply to summarize, as objectively as we can manage, both sides of the issue.

Somewhat dually, on the theoretical side of agreement there arise issues associated with interpretation of the predictive variables produced by a theory. This issue certainly arises for continuum models, of which any version of the KWM is an instance, because such models have continuous functions (e.g., mean speed) as their predictive variables, whereas any particular instance of traffic flow manifestly consists of discrete vehicles having a specific location at any given instance of time. Tying theoretical continuum results to observations is perhaps particularly difficult because much of the extant data is obtained by aggregating over time at somewhat sparsely spaced detectors, yet there seems a relative paucity of studies of the effect of either aggregation interval or detector spacing.

The contents of this work are as follows. In Section 2 we summarize the predictions of the classical KWM, and associated issues, especially those involving computational employment of the KWM. Section 3 contains a discussion of instances of agreement of those predictions with observations (qualitative features, cumulative flows), along with a critique of such observations, including especially areas where further observational studies seem warranted.
Section 4 is devoted to a critical discussion of alleged observations that seem to disagree with the classical KWM, or in some instances with any version of the KWM. The specific alleged discrepancies discussed there include (Subsection 4.1) unstable flow and related phenomena (spontaneous breakdown, the two-capacity phenomenon), (Subsection 4.2) what we term

empirically jammed flow (particularly "wide jams"), and (Subsection 4.3) internal structure in congested flow.

Section 5 is devoted to a brief discussion of the prospects for variants (e.g., nonconvex FDs) of the KWM to reproduce some observations that seem inconsistent with the classical KWM. Our concluding Section 6 contains conclusions and recommendations for further related work.

2. THE CLASSICAL KWM

The classical KWM for traffic consists of the equation of continuity,

$$\partial_t \rho(x,t) + \partial_x q(x,t) = 0,$$

plus the assumption that flow ($q=q(x,t)$) can be written as a function (fundamental diagram, FD) of density ($\rho=\rho(x,t)$) and possibly explicitly of longitudinal position (x) and time (t),say

$$q(x,t) = Q(\rho(x,t),x,t).$$

Here Q is strictly convex in ρ, i.e. $\partial_\rho^2 Q(\rho,x,t) > 0$, and vanishes at $\rho=0$ and at some jam density $\rho=\rho_{jam}$ (see Fig. 1).

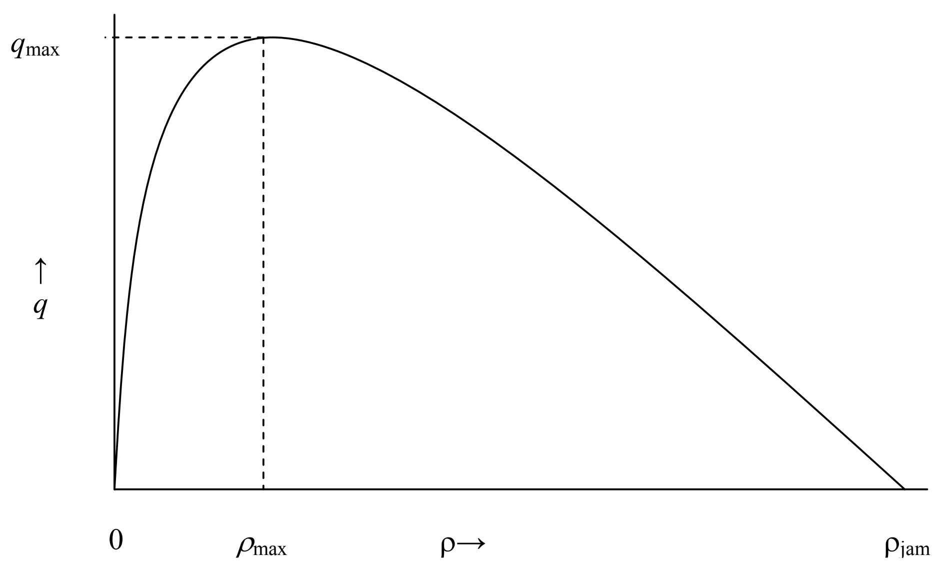

Figure 1 - A strictly convex FD, as typically conceptualized in regard to the KWM.

In the special case that the FD is spatially homogeneous (independent of x), it follows that density is a solution of the nonlinear first-order conservation law

$$\partial_t \rho(x,t) + \partial_\rho Q(\rho,t) \partial_x \rho(x,t) = 0. \tag{1}$$

In practice the FD commonly is assumed to be piecewise constant in x, so that Eq. (1) holds except at points of discontinuity. (See Sec. 2 for further brief discussion of the effect of such points.) The FD further is commonly assumed *stationary* (i.e., not to depend explicitly upon time), although nonstationary effects could be included (e.g., to model the effect of sudden changes in weather).

In the homogeneous stationary case the conservation law Eq. (1) implies that density is constant along the "characteristic curves" in the (t,x)-plane that are defined by the ordinary differential equations

$$\frac{dx}{dt} = c(x,t) := \partial_\rho Q(\rho(x,t)). \tag{2}$$

Because density is constant along a characteristic, these curves are in fact lines in the (t,x)-plane, with slope equal to the (constant) wave velocity $c=c(\rho)$ along that line.

Let us momentarily consider, for simplicity, an infinitely long one-dimensional system with initial data given at $t=0$, say $\rho(x,0)=\rho_0(x)$. If through (t,x), with $t>0$, there passed exactly one characteristic intersecting the x-axis (i.e., $t=0$), then the density at that point would be uniquely determined as that initially specified at that point of intersection with the initial line. Although there exist situations in which this stipulation holds, in general there are two ways in which it can fail.

First consider a situation (e.g., near the end of a rush period) in which the density is increasing as one moves downstream (increasing x) along the initial line. Because of the strict convexity the wave speed is (algebraically) decreasing along that interval, and therefore any two characteristics emanating from points along this interval will eventually intersect. At such a point the ``constant density along characteristics'' specification of the preceding paragraph therefore fails to specify a unique value of the density.

This ambiguity is traditionally resolved via a *shock wave*, which is to say a curve $x=x_s(t)$ in the (t,x)-plane that is the locus of the initial points of intersection of characteristics, and along which the density (and hence flow) are permitted to be discontinuous. The *shock condition*

$$\frac{dx_s}{dt} = \frac{\Delta q}{\Delta \rho} := \frac{Q(\rho_u(t)) - Q(\rho_d(t))}{\rho_u(t) - \rho_d(t)} \tag{3}$$

then follows from conservation of vehicles. Here $\rho_u(t)=\rho(x_s(t)-,t)$ ($\rho_d(t)=\rho(x_s(t)+,t)$) is the density immediately upstream (downstream) of the shock wave at time t. If one then applies the "constant density along characteristics" specification only so long as the characteristic does not cross a shock, then the solution again is uniquely specified everywhere. In particular, now Eq. (3) is an ordinary differential equation that serves to determine the trajectory of the shock wave.

In the alternative case that the initial density is decreasing as one moves downstream no corresponding difficulty arises; the characteristics diverge ("fan out"") with increasing time, but one and only one characteristic passes through each point (t,x), $t>0$. However, in the limiting case in which at some location x_0 there is a jump discontinuity in the initial density, with the downstream density smaller (e.g., at a traffic signal that has just turned green), there arises a situation in which there are points (t,x), $t>0$, through which *no* characteristic passes. In such a case one mathematical (weak) solution is a shock, with densities downstream (upstream), up to the downstream (upstream) characteristic passing through $(t,x)=(0,x_0)$, equal to $\rho(x_0+,0)$ ($\rho(x_0-,0)$), and trajectory otherwise determined by Eq. (3), just as in the preceding paragraph. However a second solution that fills the gap between the upstream characteristic through $(t,x)=(0, x_0)$ and the downstream characteristic through the same point is given by the *similarity solution*

$$\rho(x,t) = c^{-1}\left(\frac{x-x_0}{t}\right), \tag{4}$$

where c^{-1} is the functional inverse (i.e. the inverse function) of the wave velocity, which exists because of the strict convexity of Q.

In the case of such initial discontinuities one can in fact create infinitely many (weak) solutions by piecing together segments of this similarity solution with pieces of the preceding shock solution. However it is generally deemed that the similarity solution is that best representing the true state of traffic. The simplest mathematical argument supporting that choice is "stability": if one considers a sequence of initial densities having linear segments about $x=x_0$ that converge to the discontinuous initial density, and otherwise are equal to that initial density, then the corresponding (now uniquely determined, as previously) solutions resemble and converge to the similarity solution. That is, the similarity solution is self-healing, in that small perturbations from it return to it, whereas (in the present case) small perturbations of the shock solutions diverge from that shock solution.

The simplest argument supporting the choice of similarity and involving driver behavior is that any shock segment represents drivers traveling more slowly than they safely could, according to the prevailing FD. The choice of the similarity solution also relates mathematically to the so-called "entropy condition," which in turn also has ties to driver behavior. In its application to the classical KWM the entropy condition essentially asserts (Ansorge, 1990) that drivers accelerate as soon as safely possible and wait as long as safely possible to decelerate, where safety is denominated by adherence to the given FD. But the entropy condition is relevant mostly because it can be given a quantitative formulation (e.g., Leveque, 1992) that can be applied to computational approximations. This is extremely important, because early within the computational development of traffic flow theory some investigators mistakenly concluded that the classical KWM was fundamentally flawed because they inappropriately employed simple and reasonable computational approximations *that fail to satisfy the entropy condition*, and therefore converge, for an initial discontinuity as discussed in the preceding two paragraphs, to the unstable shock solution, rather than to the stable similarity solution; cf. Ross (1988), Newell (1989), Ross (1989). In fact this mistaken conclusion seems to have provided much of the impetus for development of so-called higher-order methods (Payne, 1971), some of which have recently been shown (Aw and Rascle, 2000) to be fatally flawed, in the sense that they can develop negative (i.e., upstream!) flows, even though all flows are initially downstream (and all initial densities nonnegative and less than jam density).

The STRADA code (Buisson, 1997), which has been widely applied in France, is based on computational solution of classical KWMs via the Godunov method (Lebacque, 1996), which has a discretization error that is first order in both space and time. Computational methods having higher-order computational approximations exist (Bale *et al,* (2002), but seem not to have been explored in traffic flow, perhaps because lower-order methods in fact have some advantages for the real-time demands of traffic control. Of course it is necessary, for any computational implementation, to consider a system of only some finite length, and therefore to incorporate appropriate boundary conditions. See Lebacque (to appear) for a discussion of boundary conditions, and Nelson and Kumar (2004) for the extension of these to interfaces (jump discontinuities of the FD in x) and point constrictions (removable singularities, in x, of the FD). In these cases consideration of only the similarity (i.e., stable) solution, for downward jumps in density as one moves downstream, is again conventionally adopted to obtain uniqueness.

3. EMPIRICAL AGREEMENT WITH THE CLASSICAL KWM

In this section we review empirical results that are predominantly considered to reflect agreement between observations and predictions from the classical KWM. The next section is then devoted to a discussion of empirical observations that are conversely generally considered to reflect disagreement between observational data and the KWM. In many instances there are widely held but contrary opinions regarding these matters, in the sense that procedures for both the analysis and interpretation of data and the mapping of continuum predictions onto discrete data are decidedly subject to differences of opinion. To that extent the classification of a particular allegedly observed phenomenon as reflecting agreement or disagreement with the KWM is somewhat arbitrary.

Qualitative consistency between observations and the predictions of the KWM, particularly the shock waves and acceleration waves delineated in the preceding section, comprise perhaps the most widely accepted type of agreement. This agreement is not easily or widely documented, perhaps precisely because it is so widely accepted. For example, much of the technology underlying the Highway Capacity Manual, which is a semi-official U.S. standard for the analysis and design of roadway facilities, is based on supply-demand analysis, which is the steady-state version of the KWM. The situation is also complicated by the fact that many of the empirical studies that might provide verification have been reported in the context of the literature devoted to a particular type of facility (e.g., freeway or signalized intersection), as opposed to the traffic modeling literature, and therefore not necessarily considered in the light of consistency between data and the KWM.

3.1 Shock waves

Shock waves have been primarily observed in the context of freeway bottlenecks, which is to say freeway sections at which the capacity (maximum) flow in the FD of Fig. 1 is lower than immediately upstream. This is presumably because (inductive-loop) data from such locations are widely available, and because an understanding of the manner in which congestion at such locations forms and behaves is widely considered important to mitigation of the undesirable increase in travel times that is associated with congestion. According to the classical KWM, whenever the demand (flow) just upstream of the bottleneck entrance exceeds the capacity of the bottleneck, a shock wave will form at the entrance the bottleneck and propagate upstream, leaving behind (i.e., bounded upstream by the moving shock discontinuity and downstream by the fixed bottleneck entrance) a "queue" of vehicles in which flow is equal to bottleneck capacity. However, in that queue the density will be the *higher* of the two densities corresponding to that flow on the FD, and therefore mean vehicle speed (= flow divided by density) will be (typically much) lower than that either upstream of the shock wave or downstream of the bottleneck entrance. Further, the speed of the shock should be predictable from the shock condition (3).

Agreements between these KWM predictions and observations have been documented, perhaps most notably in a series of data-oriented papers (Cassidy and Mauch, 2001; Windover and Cassidy, 2001) and companion modeling-oriented work (Newell, 2002) emanating from the transportation group at the University of California at Berkeley; see also Banks (1999). This collective body of work even argues somewhat persuasively that a "zeroth order" (Newell, 2002) theory of highway traffic, which relies on an entirely triangular FD, is still able to predict accurately the most important effects of highway traffic, especially cumulative flows.

Note however that cumulative flows are integral quantities, in the sense that at any location x they are the time integral of the differential flow $q(x,t)$, and while integral quantities are inherently easier to both measure and predict than the corresponding differential quantity, they may also be of inherently less interest; e.g., it appears that numerical differentiation, with the associated inherent loss of accuracy, will be required in order to obtain speeds and hence travel times from cumulative flows. Of course the KWM characterization of a deceleration region as a "shock wave" of zero thickness is a manifest idealization; nonetheless the actual structure of such transition zones upstream of queues has only recently been investigated empirically, by Muñoz and Daganzo (2003), who provide detailed instances of shocks that are about 1 km wide, corresponding to a relatively gradual deceleration of about 1/3 m/sec^2.

Obviously this imposes some spatial (and presumably temporal) limitation on the scale of validity of the KWM, but these workers further argue that even with such a wide shock the KWM can predict many relevant quantities quite accurately. As an example they cite vehicle positions to within 160 m, or about five vehicle spacings; however this bears further elaboration, because neither the KWM not any other continuum model predicts vehicle spacing *per se*.

3.2. Acceleration waves

The focus in validation of the KWM has tended toward shock waves, perhaps because of the ready availability of data from “point” (e.g., inductive-loop) detectors on freeways, and the practical interest in congestion (and means for its alleviation) on freeways. On the other hand there seems to be a relative paucity of observational data capable of exemplifying the companion KWM prediction of acceleration waves. This is presumably because freeway detectors tend to be located *upstream* of potential bottlenecks, and therefore not in position to acquire data representative of the “queue discharge” that is the manifestation of acceleration waves downstream of a bottleneck. On the other hand acceleration waves are extremely important to the behavior of traffic at signalized intersections; we believe the literature related to such intersections is an important source of potential data relative to validation of the KWM that has been too much neglected.

A full realization of the potential use of such data would require spatially extensive measurements of flow during a green cycle, both upstream and downstream of the signal. With more widespread use of camera-controlled signals, such data seems destined to become more widely available. Even now there is some data available, in the form of measurements of the “saturation flow' at a signal during the green phase; this parameter is extremely important to the performance of signalized intersections, and has been extensively discussed in the literature on signalized intersections. The prediction of the KWM is that (unopposed) queue discharges at such an intersection will be equal to the minimum of the capacity (maximum) flows immediately upstream of the intersection, at the intersection itself, or immediately downstream of the intersection, but in any event that it will remain constant during the green phase of the signal. In regard to this prediction, Lin, Tseng and Su (2004) state that

> “actual queue discharge patterns, however, often do not display an identifiable steady maximum rate.”

On the other hand, the variations in observed flow at the intersection that are reported by these workers are only of the order of 10-15% (see also Bonneson, 1991). Thus here the KWM seems to be categorically neither a success nor a failure, but rather the appropriate view is dependent upon the level of accuracy required for the application at hand. We note that Lebacque (2002) has considered replacing the entropy condition for the KWM by what he terms as a “bounded acceleration” condition, precisely because of doubts about sufficiency of the accuracy of the entropy solutions for modeling ramp metering. It should, however, also be noted that a KWM with a strictly convex FD does not even predict a constant flow rate out of the traffic light; at best, it predicts maximum flow *exactly at* the position of the light.

3.3. Boundary-induced Breakdown

Section 2 mentions the FD as a hypothetical representation of flow as a function of density. Let us call the FD so related to the KWM theory *KWM-FD*. One can also *measure* flows and densities in the field, leading to an *empirical FD*. Classically such observations lead to empirical FDs that resemble the KWM-FD well in the lower-density free-flow regime, but at higher densities display a tail extrapolating the free-flow data lying above a "cloud" of widely scattered data points (Drake, Schofer and May, 1967, Koshi, Iwasaki and Ohkura, 1981; Kerner, 2002, esp. Figs. 14 and 16); cf. Fig. 2 for a typical instance of such an "inverted-lambda" empirical FD. Often it is better to plot density ρ, flow q and velocity v as functions of time, as in Fig. 3. These time-series data tend to show that a typical transition is from the free-flow regime to a regime where flow is only somewhat diminished but densities are much higher, meaning much lower velocities (Mika, Keer and Yuan, 1969; Kerner and Rehborn, 1996a; Kerner, 1998, 1999a); such a transition often is termed "breakdown."

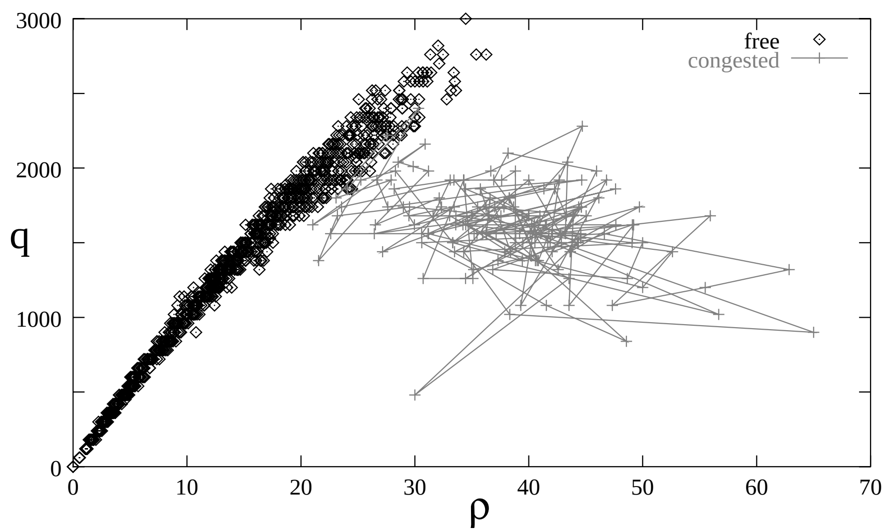

Figure 2 - Empirical fundamental diagram, as recorded on the German freeway A 43. Source: Nagel, Wagner, and Woesler, 2003.

Daganzo, Cassidy and Bertini (1999) note that a bottleneck downstream of a measurement location can easily generate such a plot, in the following way:

- The system starts with low flow at low densities.
- Both flow and density keep increasing, along the "free flow" branch of the fundamental diagram.

- This flow can be larger than what can flow through the bottleneck. Then, a queue begins to form upstream of the bottleneck, but that does not immediately influence the measurement.
- 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 now drop to bottleneck capacity.

This is exactly what the KWM predicts, in that there is a transition from free flow to congested flow that accompanies a shock wave propagating upstream. Thus this must be marked as a success of the classical KWM.

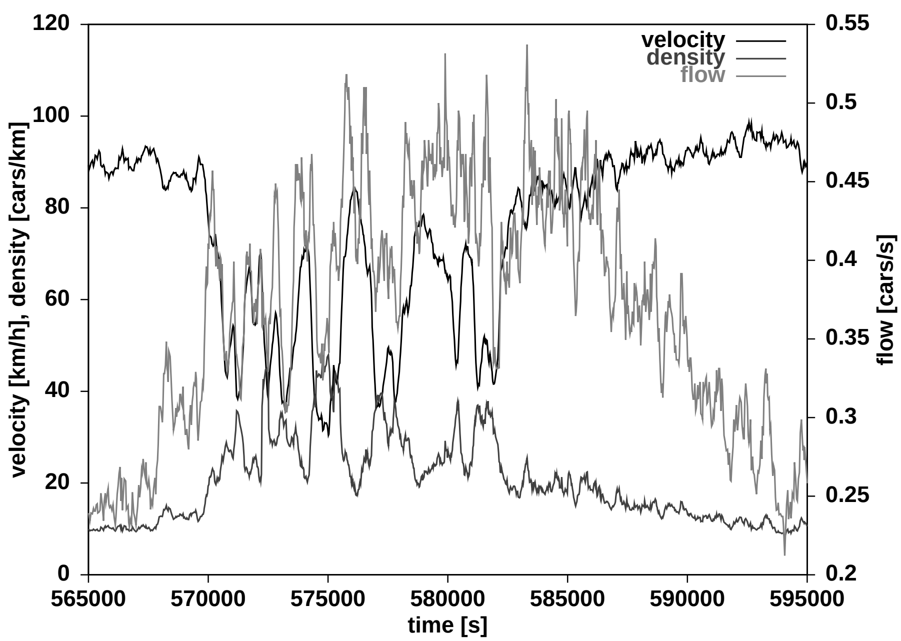

Figure 3 - Time series of the three fundamental variables of traffic flow. The data have been recorded on the German freeway A1 near an intersection with German freeway A59, in June 1996. Source: Nagel, Wagner, and Woesler, 2003.

Unfortunately, there is too much flow-density data in the literature where the associated geometrical constraints are not well enough documented. Such data is relatively useless for many of the points raised in the current scientific discussion. At the same time alternative descriptions, such as that involving three phases, free flow, congested or synchronized and jammed "synchronized flow," do not yet seem to have evolved to the point that they can provide quantitative predictions even capable of being tested. (But see Schönhof and Helbing, submitted.)

This KWM interpretation of empirical observations of breakdown has recently given rise to questions regarding the data analysis techniques traditionally employed for empirical FDs. Specifically, the empirical FD is obtained from averaging, for example over fixed time intervals. If one assumes that the KWM describes traffic, then these averages can extend over several regimes. This means that linear combinations of points on the KWM-FD can become points of the empirical FD. For a convex KWM-FD this means that points *within* the KWM-FD can become an empirical data point.

Cassidy and Bertini (1999) pursue this line of reasoning in some detail. They also describe how averages can be restricted to time intervals of (nearly) stationary traffic, ensuring that the empirical data points lie *on* a possibly existing KWM-FD as much as possible. Unfortunately, these results are not conclusive as regards the possibility of obtaining a KWM-FD via filtering empirical data for stationarity, because data so filtered are sparse both near capacity flow and in the congested flow regime. On the other hand, such sparsity is exactly what would be predicted by the KWM, in the region upstream of a bottleneck, which is where data customarily are taken (and indeed where the KWM predicts data *must* be taken, in order to observe *any* congested data). That is, the KWM predicts that once flow at an observation point upstream of a bottleneck exceeds the capacity of that bottleneck, it will do so for only a limited time, after which the queue formed at the bottleneck will spill back to the observation point and flow at that point subsequently will be equal to the bottleneck capacity (i.e., will correspond to a single density-flow point). Indeed, the chief inconsistency between the filtered near-stationary data and the KWM seems to be the existence of multiple data points in the congested regime (cf. Fig. 11 of Cassidy and Bertini, 1999). We believe much remains to be done in the area of data analysis related to empirical FDs.

4. EMPIRICAL DISAGREEMENT WITH THE CLASSICAL KWM

In this section we review empirical results that are predominantly considered to reflect disagreement between observations and predictions from the classical KWM. As in the preceding section, "agreement" and "disagreement" often are a matter of degree, and to that extent it is somewhat subjective as to whether one considers the KWM consistent or inconsistent with particular observations. Further, the disposition of an individual or group toward perceiving consistency or inconsistency between the KWM and a particular set of observations often is remarkably reflective of the degree of comfort and familiarity with the KWM or with some alternative theory. We make a significant effort to factor out such inherent bias, but perfection in that direction is remarkably elusive.

4.1 Unstable flow, spontaneous breakdown and the two-capacity phenomenon

For the classical strictly convex KWM-FD, as exemplified in Fig. 1, one typically conceives of two flow regimes, a free flow and congested flow, separated roughly by the density value ρ_{max} at which the flow achieves its maximum value, $q=q_{max}=Q(\rho_{max})$. These two regimes are thus qualitatively distinguished by the sign of the associated wave speed, $c=Q'$, which is to say by the direction of propagation along the roadway of small perturbations in (e.g.) the density. Unfortunately, this clean separation into qualitatively different regimes does not empirically appear so distinctly (cf. Fig. 2).

Basically, empirical FDs are linear for small densities, but become more complicated for higher densities, presumably because interactions become vehicles become increasingly more important as densities increase. However, in addition to the *empirical free-flow* regime, one sees underneath its higher-density extension (“tail”), and overlapping it for some range of densities (Hall and Agyemang-Duah, 1991; Kerner and Rehborn, 1996a), a regime consisting of widely scattered data points. In the preceding section we have already offered one hypothesis that might somewhat reconcile such data with the KWM-FD, but if one takes such data at face value - as has been done throughout much of the history of traffic flow theory - then it certainly seems doubtful they can be represented adequately by any single continuous function. This observation is classical in traffic flow theory, and is behind (e.g.) the well-known “two-regime” theory (Ceder, 1967), in which flow is represented as a *discontinuous* function of density.

The interpretation of this regime is considerably controversial, and even the terms one chooses to employ in referring to it can evoke some conscious or subconscious bias toward one or another interpretation, and perhaps associated theory. In a perhaps tedious effort to avoid any such bias, we shall henceforth employ the following terminology:

- By *empirical congested flow* we shall intend any data clearly not falling within the empirical free flow regime, as defined above.
- By *empirically uncorrelated flow* we shall intend any data set within which there is considerable scatter in the density-flow plane, as illustrated in Figure 2. (Thus empirically uncorrelated flow is a subclass of empirical congested flow.)
- By *empirical jammed flow* we shall intend yet another subclass of empirical congested flow that has been hypothesized by some investigators, as discussed in more detail in the following subsection.

In using these terms the word “empirical” and its derivatives will be omitted whenever the context makes it clear that observational data is the topic of discussion.

The traditional acceptance at face value of these density-flow measurements in the (empirically) congested regime in and of itself raises serious doubts regarding the validity of the KWM. This alone has been persuasive to many, and has considerably fueled both substantial skepticism directed toward the KWM, and a search for alternative data interpretations and associated explanatory theories. On the other hand, the successes of the KWM discussed in the preceding section have made others considerably reluctant to accept any suggestion that this theory has no validity, even at higher densities.

In the preceding section we have already indicated a recently suggested possibility for reconciling kinematic-wave theory with empirically uncorrelated flow. The fundamental hypothesis underlying that possibility is that breakdown is *exclusively* what nonlinear dynamicists would term a “boundary-driven phenomenon,” which is to say that it derives from shocks propagating upstream from a bottleneck, as summarized from Daganzo, Cassidy and Bertini (1999) in the preceding section. In our opinion such bottleneck-driven breakdowns certainly can and do occur, and empirical reports should include sufficient information to permit determination of the extent to which boundary effects contribute to data reflecting breakdown. This is a nontrivial challenge, because it requires spatially distributed data, or at least a very careful description of the roadway downstream of the point of observation. An extreme version of this suggestion is that empirical results should not be published at all, unless the data on which they are based are made publicly available, so that alternative hypotheses can also be tested against those data.

Nonetheless, when one takes into all account the possibility of bottleneck-driven breakdown, there remain observed phenomena that are difficult to explain absent spontaneous ("emergent") breakdown. The objective of this subsection is to summarize the observational data supporting such spontaneous breakdown.

More generally, bottleneck-driven breakdown is an instance of boundary-driven phenomenon, which in turn is instance of an "external" phenomenon that stems from circumstances outside the system being modeled, while spontaneous breakdown is conceived as an emergent phenomenon that arises from phenomena internal to the system being modeled. However, to a very large extent this distinction is subjective, in that it depends upon what one is willing to assume is known *a priori* (i.e., an external variable), as opposed to being a quantity that the model is required to predict (i.e., an internal variable). As a very concrete instance of this dichotomy, the presence of a slow truck in the right lane certainly can be a significant factor in determining traffic flow. But should this presence be regarded as an external factor, and therefore known *a priori*, or as a stochastic fluctuation whose potential existence must be represented among the statistical distribution giving rise to the mean values that a continuum model should provide?

In fact there exist circumstances under which either of these views is reasonable, depending upon the purpose of the model, and therefore what one is willing to assume is known *a priori*. For example, if the objective of the model is to better understand the effect of a slow truck (e.g., Daganzo and Laval, 2005), then it is perfectly appropriate to take the presence (and characteristics) of such a truck as an known external variable. However, if the purpose is to predict mean traffic behavior where it cannot be (continuously) measured - and otherwise there is ultimately little reason to attempt predictions - then the presence or absence, characteristics, and effect of such a truck must be regarded as an stochastic internal variable to be appropriately reflected in the model predictions.

The simplest way to determine internal effects for a *model* of traffic flow is to implement it on a homogeneous closed loop, which completely eliminates external (boundary) effects. Such "ring roads" are rare in practice, but Sugiyama *et al.* (to appear) report a remarkable experiment conducted on a test circuit. Briefly, they *experimentall*y show that under appropriate circumstances (as suggested by simulations based on the microscopic "optimal velocity model" of Bando *et al.*, 1994), spatially homogeneous flow is unstable for sufficiently high densities, but rather tends "spontaneously" to transition, after a few minutes, to “jammed flow.” In this context the characteristic of jammed flow is a single jam, propagating opposite to the direction of traffic flow, within which vehicles come to a stop, or very nearly so. Quantitative details of the observed speed-headway distribution over the entire jam flow are not reported, but supporting simulations suggest they comprise a characteristic (“universal”) limit cycle (“hysteresis loop”).

This remarkable result certainly suggests that spontaneous breakdown can occur in traffic flow. However, in and of itself it signifies neither failure of the KWM nor nonexistence of a KWM-FD, in the sense that flows and densities *averaged over the characteristic limit* cycles could still comprise a perfectly well defined function in the density-flow plane. This possibility is exemplified by the simple cellular automata model CA-184a of Nelson (submitted), which produces results similar to the empirical FD of Fig. 2, in simulations of traffic stream observations upstream of a bottleneck, but that nonetheless fall - for densities up to jam density - upon a perfectly well-defined KWM-FD in the density flow plane, with multiple underlying limit cycles, when density and flow data are obtained by averaging over an entire ring road. However, at a minimum the results of Sugiyama *et al.* (to appear) suggest existence of a lower limit on the spatial and temporal scales for validity of the KWM; existence of some such scale limitations is to be expected for any continuum model; cf. Lesort *et al.* (to appear) for further discussion of scale issues in traffic flow. In addition, Nagel and Jost (2003) argue that *if* the spontaneous breakdown is related to a true (thermodynamical) phase transition, then the jam features will coarsen until they are visible on all spatial scales, even the largest. The mechanism for coarsening is that jams coagulate until there is eventually only one large jam left in the system. However, they cannot answer if their modeling results apply to real-world traffic or not.

Observations of what we term as “two-capacity flow” have been persistently reported in the literature. This alleged phenomenon is relevant here because it also is strongly suggestive of spontaneous breakdown, and certainly is inconsistent with the classical KWM. Here by two-capacity flow we mean observations and related analyses that suggest bottlenecks can operate at two distinct maximum flows, typically a higher value prior to and early in the formation of a queue upstream of the bottleneck and a somewhat lower value subsequent to development of such a queue. This contradicts the classical KWM, because the (entropy solution of the) latter predicts flow at an enqueued bottleneck will be a constant, specifically the capacity in the bottleneck (Nelson and Kumar, 2004).

Observations of two-capacity flow have been reported, among others, by Edie and Foote (1960), by Agyemang-Duah and Hall (1991) and by Cassidy and Bertini (1999), while other studies (e.g., Persaud and Hurdle, 1991) have reported lack of conclusive evidence supporting a reduction in flow coincident with formation of a queue. From an empirical perspective, Banks (1991; cf. also Banks, 1990 and Banks, 1991a) indicates that

> “the hypothesis that flow decreases when it breaks down is confirmed, provided the hypothesis applies to individual lanes,”

but

> "when averaged across all lanes there was no significant change," and concludes that alleged two-capacity flow "is unlikely to provide a basis for metering … ."

Note that if one fully subscribes to the KWM, then there is no rational basis for ramp metering, absent some societal decision to favor mainline traffic over that entering freeways.

When differences between the levels of flow at a bottleneck, before and after queue formation, are reported, their magnitude tends to be only about 5-10%. Given that relatively small alleged difference, it is difficult to know whether the different conclusions reached in different studies are attributable to differences in methodologies for obtaining and analyzing data, or to actual differences in traffic behavior at distinct sites, or possibly some combination of these factors. This inherent difficulty in determining existence of the two-capacity phenomenon has been discussed in particularly cogent fashion by Persaud and Hurdle (1991). One can certainly question whether we have yet advanced significantly beyond the status observed by Wattleworth (1963):

> "The question of whether or not the flow downstream of a freeway bottleneck decreases when congestion sets in is currently the subject of much discussion in engineering circles. Research findings support both the yes and no answers to this question. Several studies ... suggested that perhaps the question did not have a simple yes or no answer."

Yet, on balance we tend to agree with Cassidy and Bertini (1999)

> "the average rate vehicles discharge from a queue can be ... lower than the flow measured prior to the queue's formation."

See Elefteriadou, Roess and McShane (1994), and Lorenz and Elefteriadou (2001) for a suggestion that the time of occurrence of this type of breakdown can usefully be viewed as stochastic in nature.

4.2 Empirical jammed flow

Some workers perceive an additional subclass of empirical congested flow, which is the empirical jammed flow already mentioned above. The characteristic signature of empirical jammed flow is so-called "wide jams,'" which are described by Knospe *et al.* (2002) as follows:

> "Wide (moving) jams are regions with a very high density and negligible average velocity and flow. The width of these structures is much larger than its fronts at the upstream and downstream ends where the speed of the

> vehicle changes sharply. … wide jams can propagate undisturbed through either free flow or synchronized traffic without impact on these states, which thus allows their co-existence."

(See the following subsection for more on "synchronized flow.")

The predominant evidence for wide jams is time-series data such as that shown in Fig. 23 of Knospe *et al.* (2002); i.e., data from a stationary (loop) detector that shows at early times relatively large flows and speeds with small densities, followed by an intermediate period of small flows and speeds and high densities, finally followed by yet a third time period qualitatively resembling the first. The hypothesis commonly invoked is that such data represent a structure ("wide jam") traveling upstream, within which vehicles are closely spaced, and in which both upstream and downstream fronts remain sharply defined as the structure propagates. If this interpretation is correct, then it provides a severe challenge to the classical KWM, because (see Section 2) that model predicts that an initially sharp front between a high-density upstream region and a low-density downstream region (such as the downstream front of the hypothesized wide jam) will dissipate (spread) into an acceleration fan (similarity solution).

However, absent additional data from other spatial points there is an interpretation of these data that reflects a boundary-driven phenomenon, and is consistent with the classical KWM. Once more, consider a detector located upstream of a bottleneck, with initial flow and downstream demand both below bottleneck capacity, then increasing to a value above bottleneck capacity, and finally decreasing to a value again below bottleneck capacity. As already discussed in the preceding section, the KWM then predicts flow at the bottleneck will be the lagged value of input demand, until the tail of the queue generated at the bottleneck by the excess of demand over bottleneck capacity arrives at the detector, in the form of a shock wave. Thereafter flow at the detector will remain at the bottleneck capacity, with density equal to the corresponding point on the congested branch of the FD, and the tail of the queue will continue to move upstream as a shock wave, all until such time as the upstream demand falls below the bottleneck capacity. Once that decreased demand reaches the tail of the queue the tail will remain a shock wave, and thus be a stable structure within the confines of the classical KWM, but now it will propagate downstream. It will eventually reach the detector, where it will be registered as a sharp transition from a high-density low-flow/speed regime to low-density high-flow/speed. As seen at the hypothesized single detector the traffic pattern will thus be exactly the observed signature of an alleged wide jam, as described in the preceding paragraph.

Thus the question of whether the data supporting existence of wide jams are consistent with the classical KWM comes down to the question of whether the second observed sharp front is moving upstream or downstream. This seems difficult, if not impossible, to answer without data from more than one spatial location. At this time the totality of the data of the type described above that has been reported in the literature in support of the wide-jam hypothesis seems to be Kerner and Rehborn (1996a, 1996b), Knospe *et al.* (2002) and Schönhof and Helbing (submitted). Of these three works, only in Figures 3(a)-(c) of Kerner and Rehborn (1996a) and Figure 2 of Kerner and Rehborn (1996b) does one find data from detectors at distinct positions along the roadway. These data definitely support existence of wide jams. Certainly it would be desirable if they were replicated by other investigators. Nonetheless, in the following sections we conditionally accept the hypothesis of existence of wide jams, and hence empirical jammed flow, and explore the potential for various models to reproduce that phenomenon.

Some empirical properties of wide jams that have been reported (Kerner and Rehborn, 1996a, 1997) include the following:

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.

4.3 Further internal structure in empirically uncorrelated flow

The "wide jams" characteristic of the empirical jammed flow discussed in the preceding subsection constitute one instance of internal structure in empirically uncongested flow. However, other types of internal structure have been reported. In fact, "stop-and-go" flow under congested conditions is anecdotally familiar to virtually all drivers. Notwithstanding that this phenomenon was initially studied scientifically nearly forty years past (Edie and Baverez, 1967), there remains a relative paucity of related quantitatively reliable scientific data. This very possibly is because such data require short-term and short-distance temporally and spatially distributed data, which necessarily will be subject to a high degree of stochastic uncertainty (stemming from, e.g., variations between individual drivers). Nonetheless there are persistent reports of "something else" (e.g., Kerner, Klenov and Wolf, 2002), and there is considerable discussion (Daganzo, Cassidy and Bertini, 1999) as to whether these observations are or are not entirely explainable by KWM theory. This subsection is devoted to a brief review of these issues.

Much of the discussion focuses on what Kerner and collaborators (Kerner and Rehborn, 1996b, 1997) term as "synchronized flow." One of the barriers to an objective discussion of synchronized flow is the lack of consensus as to its characteristic signature(s). An approximately common (i.e., "synchronized") speed across multiple lanes, and the existence of embedded very narrow (compared to the wide jams of the preceding subsection) jams are two signatures that have been mentioned by workers who have considered the matter. Yet another barrier is lack of public availability of much of the data on which many of the alleged observations of synchronized flow are based, and the consequent inability of the community at large to reproduce the underlying data analysis, or suggest alternate analyses. Even among those who are disposed favorably toward existence of synchronized flow there is not agreement on the question of whether the associated strong scatter of the data has a dynamical origin, for example stemming from an on-ramp, or a statistical origin, such as being caused by a mixture of cars and trucks that display different driving characteristics.

Sugiyama *et al.* (to appear) provide simulation results, based on the optimal-velocity model (Bando *et al*., 1995) that suggest a hypothesis regarding possible internal structure of congested flow. Briefly, their simulations reveal the following structure upstream of a bottleneck: immediately upstream a region of "laminar flow," within which vehicles are spaced rather regularly, followed by a region within which the flow consists of very narrow "jams," finally followed by a region even further downstream within which wide jams appear. These workers hypothesize the laminar region is an inherently metastable region, the region of wide jams is inherently stable, and the region of narrow jams is a connecting unstable region (as dictated by Hopf bifurcation) that is somehow stabilized by presence of the bottleneck. Note that the latter interpretation also is consistent with considerable difficulty in observing the narrow jams, which in turn would be somewhat consistent with difficulties in observing synchronized flow, if one hypothesized an identification between synchronized flow and narrow jams, as do Sugiyama and Nakayama (2003).

The above description, in particular that narrow jams coagulate to form wide jams, is also consistent with work by Jost and Nagel (2003). Jost and Nagel do not investigate bottlenecks, use a closed homogeneous system, i.e. a homogeneous loop. Once of their initial conditions is a completely homogeneous distribution of cars, i.e. all vehicles with the same space headway, as given by the density. Also in that system, one obtains, for certain densities, initially "laminar flow,", then "narrow jams," and then wide jams. As pointed out earlier, the wide jams will further coagulate until there is only one jam left in the system, and that statement holds for arbitrarily large systems. Further simulations of our own show that the same code, after the introduction of a bottleneck, displays the same behavior as the model of Sugiyama *et al.* (to appear). An interpretation is that, either by a bottleneck or by having a closed system, traffic can be "pinched" at a certain (average) density, at which the system displays the above-described behavior. However, even here it must be said that there are very few empirical observations that include the spatial picture, and at least one of them (Windover and Cassidy, 2001) displays little or no coarsening with increasing distance from the bottleneck.

The principal point of interest here is the extent to which structure internal to empirically congested flow disagrees with the KWM. Within the community of those interested in modeling traffic flow a commonly encountered, if implicit, assumption is that the KWM, and presumably other continuum models, predict (e.g.) *actual* vehicular speed, and therefore is contradicted by any local variation in empirical vehicular speeds. While such a "first-in first-out" interpretation of continuum models might be useful for certain applications of continuum models, it is not inherent to continuum models *per se*. That is, continuum models of traffic flow make predictions regarding density, mean speed and flow, all of which are continuum quantities that can be interpreted as mean values over some underlying distribution of vehicular speeds. But continuum models themselves are silent on the form of the associated speed distributions, and indeed even regarding the very nature of these distributions (i.e., over time and space, in a particular instance, or over an "ensemble" of instances'). From this perspective structure within empirically congested flow perhaps provides information (e.g., the appropriate nature of the associated speed distribution) beyond that available from the KWM itself, but that information is not necessarily in contradiction to the KWM. See Nelson (submitted) for an example of a simple model of traffic flow that has such an internal structure within its congested regime, but nonetheless can be modeled reasonably well via the KWM, with an appropriately constituted FD. However, once more also note that in some models the structures can coarsen beyond any arbitrarily large but fixed length scale.

5. NONCLASSICAL KINEMATIC-WAVE MODELS

Because a strictly convex FD will not lead to wide jams, but such jams have been observed in the field (Subsection 4.2), it is necessary to explore other forms of $Q(\rho)$. This section is devoted to such considerations. Specifically, in subsection 5.1 we briefly review known results for FDs containing linear segments. Subsections 5.2-5.4 are devoted to considerations related to a FD having a concave "tail" at higher densities, and satisfying other conditions described in more detail in Subsection 5.2. More precisely, in Subsection 5.2 we construct the Riemann solutions for such a FD, Subsection 5.3 is devoted to possible use of these Riemann solutions to construct wide jams, and Subsection 5.4 is a brief discussion of possible driver behavior underlying some of the novel aspects of solutions of the KWM for FDs containing a concave segment.

5.1 FDs with linear segments

Lin and Lo (2003) noted that once we allow strictly linear pieces in $Q(\rho)$, then some sort of stability of both fronts can be achieved as long as both the density within the jam and the surrounding densities are within the same linear segment. This is because the slope of $Q(\rho)$ denotes the phase velocity of the wave features, and if all densities in the range of interest are within the same linear segment, then their wave features move with the same velocity. Thus at least some of the elements of wide jams seemingly can be reproduced within the context of the simple triangular FD suggested by Newell (2002).

At either the upstream or downstream front of such a wide jam the phase speeds on either side of the front are equal to the speed of the front itself; i.e., in the (x,t)-plane the characteristics are *parallel* to the trajectory of the front. Such a discontinuity is known as a "contact discontinuity." A contact discontinuity is only weakly stable, in the sense that if the initial densities are slightly "smeared," where "slight" means they still remain in the aforementioned linear segment of the FD, then the ensuing density profile retains this smeared form, rather than reorganizing into a sharp front (i.e., displaying "self healing"). At this time, there seems to be no general agreement if this type of "weak" jam stability is sufficient to explain what is observed.

5.2 An FD with a concave tail

This leaves the case of a FD $Q(\rho)$ with piecewise concave regions. This situation is understood in the mathematical literature (e.g., Liu, 1981; Li, 2003), but the detailed implications for the KWM – particularly possible interpretations regarding the driver behavior underlying any novel aspects of the corresponding solutions - do not seem to have been worked out. Here we restrict ourselves, for the sake of illustration, to an FD that is strictly convex on say the density interval $0 \le \rho < \rho_{\text{inf}}$, strictly concave but nonetheless with $\partial_\rho Q(\rho) < 0$ on $\rho_{\text{inf}} < \rho \le \rho_{\text{jam}}$ (so that ρ_{inf} is an inflection point, $\partial_\rho^2 Q(\rho_{\text{inf}}) = 0$), and further $\rho_{\max} < \rho_{\text{inf}}$ (see Fig. 4). Note that these assumptions imply mean vehicular speed, $v(\rho) = Q(\rho)/\rho$, is a decreasing function of ρ.

We note parenthetically this there is little experimental evidence for (or against) such an FD, which is perhaps why the details have not been previously worked out. However, the homogeneous solutions of the optimal-velocity model (Bando *et al.*, 1995) give rise to such an FD, as shown in Fig. 4. Of course it remains to determine how this relates to the jammed flow solutions that appear to be the stable form of the solutions of the optimal-velocity model under somewhat congested conditions. The hallmark parametric values in Fig. 4 agree reasonably well with those generally accepted within the North American transportation community, although both the value of maximum flow ($q_{\max} \approx 2772$ vehicles per hour per lane) and the density at which this flow occurs ($\rho_{\max} \approx 28.95$ vehicles per kilometer) seem a bit high.

In this subsection our objective is to work out the Riemann solutions for such an FD, which is to say the solutions of the KWM on an open (infinitely long) section of roadway, with given spatially homogenous initial values on either side of some initial discontinuity. If both the upstream and downstream initial densities lie in the convex region (i.e., are $\le \rho_{\text{inf}}$), then there is no difference from the classical situation previously discussed. However, if the upstream and downstream densities are both initially in the strictly concave region (i.e., are $\ge \rho_{inf}$), then matters change somewhat. In the first instance, suppose $\rho_u < \rho_d$, where ρ_u (ρ_d) denotes the initial upstream (downstream) density. In the classical case this would correspond to a situation in which characteristics intersect at the initial discontinuity, and therefore a shock wave forms there, corresponding to a region within which incoming vehicles decelerate. Now however the characteristics initially *diverge* at the discontinuity, which leads to an entropy solution having the form of Eq. (4), where now the inverse function c^{-1} must be interpreted as the inverse function of the wave speed for densities restricted to the concave region. The initial discontinuity therefore develops into a region within which the characteristics "fan out," so that the densities more-or-less (more later, less initially) gradually increase from the initial upstream value to the initial downstream value, as one moves from upstream to downstream. However, mean vehicular speeds decrease across such a region, so that in this case the entropy solution of the Riemann problem perhaps should be termed a deceleration wave (or deceleration fan).

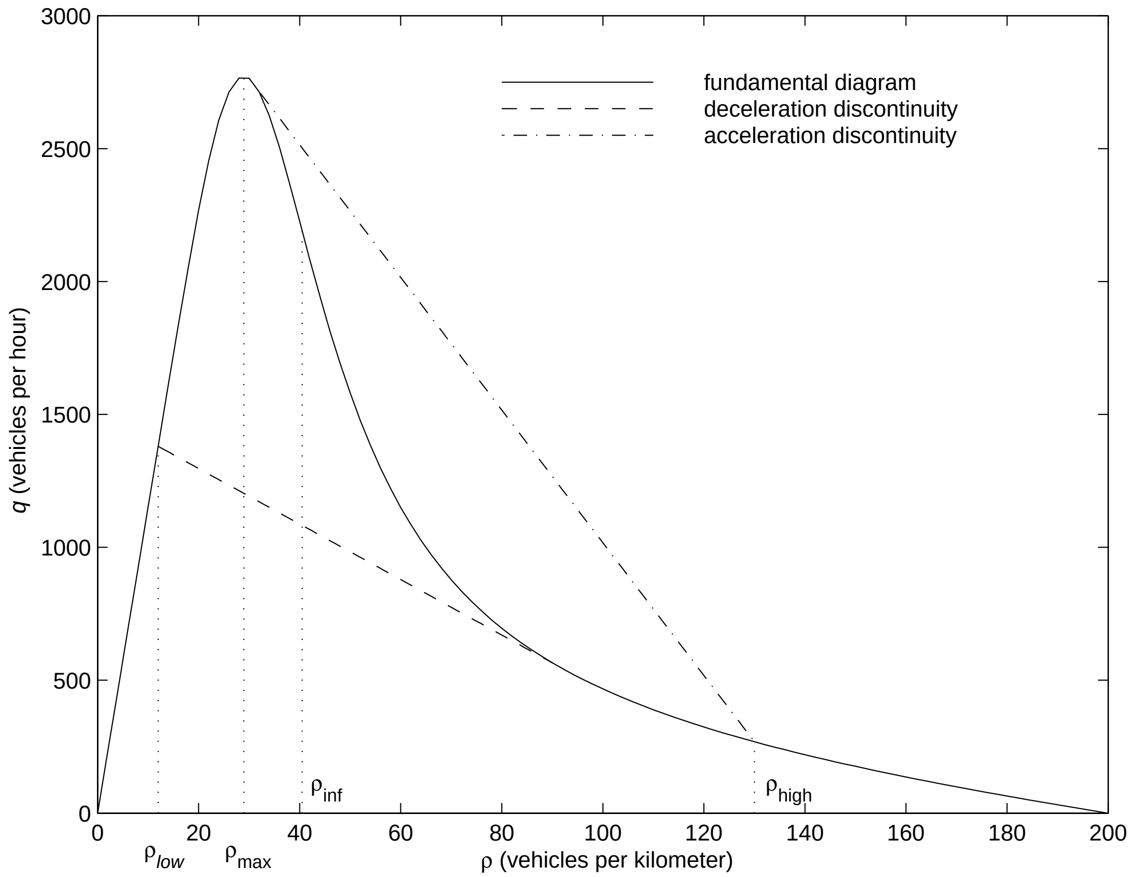

Figure 4 – The fundamental diagram corresponding to the homogeneous solution of the optimal-velocity model. Parametric values for the optimal-velocity model were selected following Sugiyama *et al.* (to appear). Other indicated parameters and associated graphics intended to support illustration of the wide jam of Subsection 5.3.

Similarly, if the initial upstream density is greater than that initially downstream, then the characteristics initially intersect at the discontinuity. As in the corresponding classical case a shock wave therefore forms, and moves according to the shock condition (3) (and therefore necessarily upstream, in the case presently considered). However, in contrast to the case of a classical shock wave, vehicle speeds increase as one moves across the shock. That is, the entropy solution in this case is an *acceleration shock,* in contrast to the classical case of a deceleration shock.

It remains to describe the Riemann solutions for the case that one of the initial densities is in the concave region and the other in the convex region. The general principle underlying this construction is that any discontinuity, say connecting regions of respective immediately upstream and downstream densities ρ_u and ρ_d, and therefore propagating at speed specified by the shock condition (3), must satisfy the entropy condition (Li, 2003), i.e. that the inequality

$$\frac{Q(\rho_u)-Q(\rho_d)}{\rho_u-\rho_d} \leq \frac{Q(\rho_u)-Q(\rho)}{\rho_u-\rho}$$

must hold for any density ρ lying between ρ_u and ρ_d.

In the first case, suppose the initial upstream density is the smaller of the two initial densities, then it must lie in the free flow regime, and therefore necessarily the convex region (i.e., $\rho_u \leq \rho_{\max} \leq \rho_{\inf}$), and the initial downstream density, ρ_d, must lie in the concave region. Let $\rho^*(\rho_d)$ be the density corresponding to the (unique) point in the convex region where the tangent line to the FD at the point $(\rho_d, Q(\rho_d))$ intersects the FD. In the subcase that $\rho_u \leq \rho^*(\rho_d)$ the Riemann solution simply consists of an upstream region of density ρ_u and a downstream region of density ρ_d, connected by a shock propagating as specified by the shock condition. This is a deceleration shock, more-or-less as in the classical case, because vehicular speeds are slower downstream than upstream of the shock.

In the subcase that $\rho^*(\rho_d) < \rho_u$ ($< \rho_{\inf}$) the entropy solution consists of four regions, from upstream to downstream as follows:

i) A far upstream region of density ρ_u.

ii) A discontinuity connecting the far upstream region of density ρ_u to an immediately downstream density $\rho^t(\rho_u)$, where the latter is the density associated to the unique point in the concave region where the secant line from $(\rho_u, Q(\rho_u))$ is tangent to (i.e., touches) the FD. Again the mean vehicular speed decreases across this discontinuity, and it is a true shock (incident characteristics) as seen from immediately upstream, but now it is a contact discontinuity (parallel characteristics) as seen from immediately downstream. This *deceleration discontinuity* propagates according to the shock condition (3), of course with ρ_d replaced by $\rho^t(\rho_u)$. The deceleration discontinuity corresponding to a specific situation to be described in the following subsection is indicated in Fig. 4.

iii) A deceleration wave, as described above, but now connecting the region of density $\rho^t(\rho_u)$ that is immediately downstream of the deceleration discontinuity just described to a far downstream region of density ρ_d. (Note that necessarily $\rho^t(\rho_u) \leq \rho_d$.) The density within this deceleration wave is again given by (4), with c the wave speed restricted to the concave region.

iv) The far downstream region of density ρ_d just described.

Similarly, but briefly, if the initial upstream density is the larger of the two initial densities, then the unique entropy solution is as follows. Let $\rho^t(\rho_u)$ be defined exactly as above, but note now that $\rho^t(\rho_u)$ lies in the convex region, and the secant line defining it lies entirely (on or) above the FD. If $\rho_d \geq \rho^t(\rho_u)$ then the solution is an acceleration shock, precisely as in the case that both initial densities lie in the concave region. If $\rho_d < \rho^t(\rho_u)$, then the four-region solution consists of a far upstream region of density ρ_u, followed by an acceleration discontinuity connecting that region to a downstream density $\rho^t(\rho_u)$, followed by an acceleration wave connecting upstream density $\rho^t(\rho_u)$ to a far downstream region of density

ρ_d. The discontinuity travels again at speed given by the shock condition (3), with ρ_d replaced by $\rho^t(\rho_u)$, and the density profile in the acceleration wave is again given by (4), but now with c the wave speed in the *convex* region. Note that the acceleration discontinuity is a shock (contact discontinuity) as seen from immediately upstream (downstream). The acceleration discontinuity corresponding to a specific situation to be described in the following subsection is indicated in Fig. 4

5.3 Wide jams from an FD with concave tail?

Consider a FD as in Fig. 4, and associated initial conditions consisting of a density ρ_{high} lying in the associated concave region over some section of roadway, say $x_0 \le x \le x_1$, and density ρ_{low} lying in the convex region otherwise. In Fig. 4 ρ_{high} is illustrated as 130 vehicles per kilometer, and ρ_{low} as 12 vehicles per kilometer. For these values for the initial conditions, and the optimal-velocity FD of Fig. 4, $\rho^t(\rho_{\text{low}}) \approx 90$ vehicles per kilometer and $\rho^t(\rho_{\text{high}}) \approx 30$ vehicles per kilometer are respectively the densities corresponding to the points at which the secant lines to the points on the FD corresponding to respectively the low density and high density initial values just touch the FD.

According to the Riemann solutions of the preceding subsection these initial conditions will tend to evolve toward a structure within which, in order from upstream to downstream, the upstream front consists of a deceleration discontinuity followed by a deceleration wave, while the downstream front consists of an acceleration discontinuity followed by an acceleration wave. The two discontinuities are defined by the two secant lines mentioned in the preceding paragraph, and are illustrated diagrammatically in Fig. 4. The acceleration (deceleration) wave is the portion of the FD lying between the acceleration (deceleration) discontinuity and the density ρ_{low} (ρ_{high}).

The question of relevance here is whether it is possible for such a structure to evolve into a stable structure that has the elements of the wide jams discussed above, as opposed to dissipating into the background density ρ_{low}. Although we do not have a mathematical proof, it appears that the answer is "no." The essence of what seems to happen is that the upstream face of the downstream front of this pulse eventually catches and "absorbs" the entire upstream front. Alternatively, the downstream face of the upstream front will eventually reduce the pulse density to at most ρ_{inf}. At this point, all of the pulse lies in the convex region of the FD and will therefore behave (and dissolve) as discussed earlier. Space does not permit further discussion of this matter here.

5.4 Possible driver behaviour associated with acceleration shocks and deceleration waves

In Section 2 we justified the entropy condition for classical FDs as stemming from the tendency of drivers to accelerate as soon as safely possible as soon as safely possible, thereby creating acceleration waves, and to wait as long as safely possible to decelerate, thereby creating deceleration shocks. Is there a similar explanation for the appearance of the dual deceleration waves and acceleration shocks in a region of concavity for an FD? The details of any such explanation necessarily are intertwined with the observed or supposed driver behavior that leads to this concavity. However one possible generic explanation is that such a concave region represents driver behavior within a regime that drivers are on high alert, and therefore relatively quick to brake and slow to accelerate.

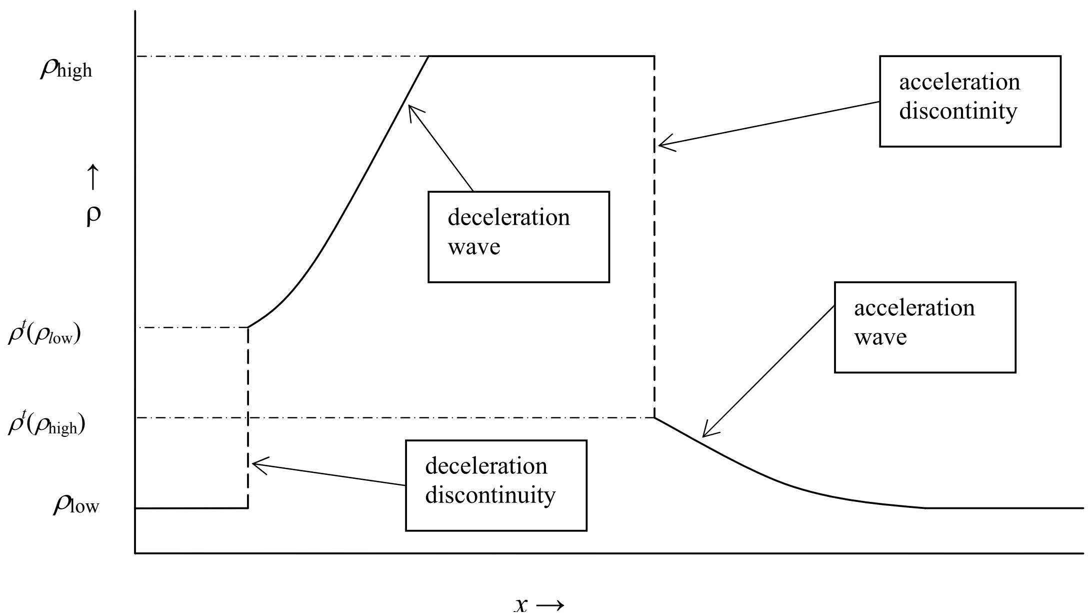

Figure 5 – Elements of the traffic-flow structure that evolves from an initial square pulse in density. (The solid lines are density profiles.)

6. CONCLUSIONS

This paper was written with the intent to first reach a firm basis of agreement between the two authors, and then to continue from there with the description and analysis of simulations that describe some of the more subtle effects of traffic flow, such as structure formation in queues. Somewhat surprisingly, just reaching the firm basis has exhausted the limits given for this paper. This seems to be caused by the following: (1) Both authors have their own intuitions about the dynamics of traffic flow. (2) Both authors agree that the other author's intuition is valid, although considerable explanation and thoughtfulness was necessary in order to have that agreement based on true mutual understanding. (3) Both authors' intuitions are consistent with the data sets that they were aware of; indeed, only very few data sets are able to answer at least some of the critical questions. The perhaps most significant example is the breakdown experiment of Sugiyama *et al.* (to appear), which, although it has been reported in 2001, has neither been published not been widely disseminated. Nevertheless, it is the *only unequivocal* example of spontaneous breakdown that we are aware of. Similar statements hold for the availability of data that is both temporal and spatial.

The maybe most important results of our effort are the following:

- The kinematic wave model (KWM) explains and predicts many if not most effects of real world traffic dynamics. As pointed out, e.g., by Daganzo, Cassidy, and Bertini (1999), much of the single detector data supposedly in support of spontaneous breakdown can also be explained by queue dynamics in conjunction with a geometrical constraint.
- On the other hand, there are some data sets that are, in both authors' views, very difficult to reconcile with the traditional KWM. This holds certainly for KWMs with strictly convex fundamental diagrams (FDs), which do not explain stable jams, spontaneous breakdown out of nothing, or structure formation in queues.
- Also KWMs with piecewise linear segments do not seem to explain all of these effects, although the decision is not so clear-cut: Such models predict quasi-stable jams which could, although they are not "self-healing," be consistent with the few instances of wide jams for which we were able to find unequivocal data. Structure formation in queues could be argued as being beyond the spatial scale that KWMs claim to deal with, although a "true" phase transition interpretation of traffic flow would predict that these structures coagulate into larger and larger structures with increasing distance from the bottleneck. Finally, there seems to be only one unequivocal observation of spontaneous breakdown.
- Finally, KWMs with FDs with concave segments are less well understood, but do not seem to offer a better alternative.
- As said before, more traffic data that includes the spatial aspect is needed in order to make further progress.

Let us close noting that traffic flow is not alone as an instance of an area in which there is some divergence of opinion between traditional practitioners and those seeking to explore the possibility that some of the more intriguing aspects of the possible behavior of low-dimensional dynamical systems can arise outside the laboratory. In this respect we note the following quotation (Schreiber, 1999), regarding the difficulty of convincing identification of possibly interesting low-dimensional dynamical effects in inherently noisy systems:

> "The most direct link between chaos theory and the real world is the analysis of time series data in terms of nonlinear dynamics. Most of the fundamental properties of nonlinear dynamical systems have by now been observed in the laboratory. However, the usefulness of chaos theory in cases where the system is not manifestly deterministic is much more controversial. In particular, evidence for chaotic behaviour in field measurements has been claimed and disputed in many areas of science, including biology, physiology, and medicine; geo- and astrophysics, as well as the social sciences and finance."

REFERENCES

Agyemang-Duah, K. and F. L. Hall (1991). Some issues regarding the numerical value of capacity. In: *Procs. Int. Symp. Highway Capacity and Level of Service* (U. Brannolte, ed.), pp. 1-15, A. A. Balkema Press, Rotterdam.

Ansorge, R. (1990). What does the entropy condition mean in traffic flow theory? *Transportation Research B*, **24B**, 133-144.

Aw, A. and M. Rascle (2000). Resurrection of "Second Order'" Models of Traffic Flow. *SIAM J. Applied Mathematics*, **60,** 916-938.

Bando, M. *et al*. (1995). Dynamical model of traffic congestion and numerical simulation. *Physical Review E*, **51,** 1035-1042.

Banks, J. H. (1990). Flow processes at a freeway bottleneck. *Transportation Research Record*, **1287**, 20-28.

Banks, J. H. (1991). Two-capacity phenomenon at freeway bottlenecks: A basis for ramp metering? *Transportation Research Record*, **1320**, 83-90.

Banks, J. H. (1991a). The two-capacity phenomenon: Some theoretical issues. *Transportation Research Record*, **1320**, 234-241.

Banks, J. H. (1999). Investigation of some characteristics of congested flow. *Transportation Research Record*, **1678**, 128-134.

Bale, D. S. *et al.* (2002). A wave propagation method for conservation laws and balance laws with spatially varying flux functions. *SIAM J. Scientific Computing*, **24**, 955-978.

Bonneson, J. A. (1991)]. Modeling queued driver behavior at signalized junctions. *Transportation Research Record*, **1365**, 99-107.

Buisson, C. (1997). Découpage objets de Strada. Internal report LICIT, ENTRE-INRETS, France.

Cassidy, M. J. (1998). Bivariate relations in nearly stationary highway traffic. *Transportation Research B*, **B32***,* 49-59.

Cassidy, M. J. and R. Bertini (1999). Some traffic features at freeway bottlenecks. *Transportation Research B*, **B33** 25-42.

Cassidy, M. J. and M. Mauch (2001). An observed traffic pattern in long freway queues. *Transportation Research A*, **A35** 143-156.

Ceder, A. (1967). A deterministic traffic flow model for the two-regime approach. *Transportation Research Record*, **567**, 16-30.

Daganzo, C. F., M. J. Cassidy, and R. Bertini (1999). Possible explanations of phase transitions in highway traffic. *Transportation Research A*, **33**, 365-379.

Daganzo, C. F. and J. A. Laval (2005). On the numerical treatment of moving bottlenecks. *Transportation Research B,* **39**, 31-46.

Drake, J. S., J. L. Schofer and A. D. May (1967). A statistical analysis of speed density hypotheses. *Highway Research Record*, **154**, 53-87.

Edie, L. C. and E. Baverez (1967). Generation and propagation of stop-start traffic waves. In: *Vehicular Traffic science: Proceedings of the Third International Symposium on the Theory of Traffic Flow* (L. C. Edie, R. Herman and R. Rothery, eds.) pp. 26-37, American Elsevier, New York.

Edie, L. C. and R. S. Foote (1960). Effect of shock waves on tunnel traffic flow. *Proceedings of the Highway Research Board*, **37**, 492-505.

Elefteriadou, L., R. P. Roess and W. R. McShane (1994). Probabilistic Nature of Breakdown at Freeway Merge Junctions. *Transportation Research Record*, **1484**, 80-89.

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

Jost, D. and K. Nagel (2003). Probabilistic traffic flow breakdown in stochastic car following models. *Transportation Research Record,* **1852**, 152-158.

Kerner, B. S. (1998). Experimental features of self-organization in traffic flow. *Physical Review Letters*, **81**, 3797-3800.

Kerner, B. S. (1999a). Phase transitions in traffic flow. In: *Traffic and Granular Flow '99* (D. Helbing, H. J. Hermann, M. Schreckenberg, and D. E. Wolf, eds.), pp. 253-284. Springer, Berlin.

Kerner, B. S. (2002). Empirical macroscopic features of spatial-temporal traffic patterns at highway bottlenecks. *Physical Review E*, **65**, Paper No. 046138.

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

Kerner, B. S. and H. Rehborn (1996a). Experimental properties of complexity in traffic flow. *Physical Review E*, **53**, R4275-R4278.

Kerner, B. S. and H. Rehborn (1996b). Experimental features and characteristics of traffic jams. *Physical Review E*, **53**, R1297--R1300.

Kerner, B. S. and H. Rehborn (1997). Experimental properties of phase transitions in traffic flow. *Physical Review Letters*, **79**, 4030-4033.

Knospe, W. *et al.* (2002). Single-vehicle data of highway traffic: Microscopic description of traffic phases. *Physical Review E*, **65**, 1-16.

Koshi, M., M. Iwasaki and I. Ohkura (1981). Some findings and an overview on vehicular flow characteristics. In: *Transportation and Traffic Theory: Proceedings of the 8th ISTTT* (V. F. Hurdle, E. Hauer and G. N. Steuart, eds.), pp. 403-426, University of Toronto Press, Toronto.

Lebacque, J. P. (1996). The Godunov Scheme and what it means for first order traffic flow models. In: *Transportation and Traffic Theory: Proceedings of the 13th ISTTT* (J. B. Lesort, ed.) pp. 647-677, Pergamon, Oxford.

Lebacque, J. P. (2002). A two-phase extension of the LWR model based on the boundedness of traffic aceleration. In: *Transportation and Traffic Theory in the 21st Century: Proceedings of the 15th ISTTT* (M. A. P. Taylor, ed.) pp. 697-718, Pergamon, Amsterdam.

Lebacque, J. P. (to appear). Intersection modeling, application to macroscopic network traffic flow models and traffic management. In: *Proceedings of the Workshop on Traffic and Granular Flow 2003*.

Leveque, R. (1992). *Numerical Methods for Conservation Laws*. Birkhäuser Verlag, Basel.

Li, T. (2003). Global solutions of nonconcave hyperbolic laws with relaxation arising from traffic flow. *Journal of Differential Equations,* **190**, 131-149.

Lighthill, M. J. and G. B. Whitham (1955). On kinematic waves II - a theory of traffic flow on long crowded roads. *Proc. Royal Society, London*, **A229**, 317-345.

Lin, F. B., P. Y. Tseng and C. W. Su (2004). Variations in queue discharge patterns and their implications in analysis of signalized intersections. In : *Procs. 83rd Annual Meeting Transportation Research Board*, Wash. D.C., Jan. 11-15 (published in CD format).

Liu, T.-P. (1981). *Admissible solutions of hyperbolic conservation laws*. American Mathematical Society (Memoirs Series, Vol. 240), Providence.

Lorenz, M. R. and L. Elefteriadou (2001). Defining freeway capacity as function of breakdown probability. *Transportation Research Record*, **1776**, 43-51.

Mika, H. S., J. B. Keer and L. S. Yuan (1969). Dual mode behavior of freeway traffic. *Highway Research Record,* **279**, 1-13.

Muñoz, J. C. and C. F. Daganzo (2003). Structure of the transition zone behind freeway queues. *Transportation Science*, **37**, 312-329.

Nagel, K., P. Wagner, and R. Woesler (2003), Still flowing: Approaches to traffic flow and traffic jam modeling. Operations Research, **51**, 681-710.

Nelson, P. (submitted). On two-regime flow, fundamental diagrams and kinematic-wave theory. Submitted to *Transportation Science*, March 2004.

Nelson, P. and N. Kumar (2004). Point-constriction, interface and boundary conditions for the kinematic-wave model.: In *Procs. 83rd Annual Meeting Transportation Research Board*, Wash. D. C.,Jan. 11-15 (published in CD format).

Newell, G. F. (1989). Comments on traffic dynamics. *Transportation Research B*, **23B**, 386-389.

Newell, G. F. (2002). A simplified car-following model: A lower order model. *Transportation Research B*, **36B** 195-205.

Payne, H. J. (1971). Models of freeway traffic and control. *Simulation Council Procs.,* **1**, Simulation Council, Inc., La Jolla, CA, 51-61.

Persaud, B. N. and V. F. Hurdle (1991). Freeway capacity: definition and measurement issues. In: *Procs. Int. Symp. Highway Capacity and Level of Service* (U. Brannolte, ed.), pp. 289-307, A. A. Balkema Press, Rotterdam.

Richards, P. I. (1956). Shockwaves on the highway. *Operations Research*, **4**, 42-51.

Ross, P. (1988). Traffic dynamics. *Transportation Research B*, **22B**, 421-435.

Ross, P. (1989). Response to Newell. *Transportation Research B*, **23B**, 390-391.

Schreiber, T. (1999). Interdisciplinary application of nonlinear time series methods. *Physics Reports*, **308**, 1-64.

Schönhof, M. and D. Helbing (submitted). Empirical features of congested traffic states and their implications for traffic modeling. *Transportation Science,* submitted. See www.helbing.org.

Sugiyama, Y. *et al.* (to appear). Observation, theory and experiment for freeway traffic as physics of many-body system. In: *Proceedings of the Workshop on Traffic and Granular Flow 2003.*

Sugiyama, S. and A. Nakayama (2003). Understanding "synchronized flow" by optimal velocity model. In: *Modeling of Complex Systems: Seventh Granada Lectures* (P. L. Garrido and J. Marro, eds.), pp. 111-115, American Institute of Physics.

Wattleworth, J. A. (1963). Some aspects of macroscopic freeway traffic flow theory. *Traffic Engineering*, **A35**, 15-20.

Windover, J. R. and M. J. Cassidy (2001). Some observed details for freeway traffic evolution. *Transportation Research A*, **A35**, 881-894.