Possible explanations of phase transitions in highway traffic

C.F. Daganzo * *, M.J. Cassidy, R.L. Bertini

Department of Civil and Environmental Engineering, and Institute of Transportation Studies, University of California, Berkeley, CA, 94720, USA

Abstract

It is shown that all the phase transitions in and out of freely flowing traffic reported earlier for a German site could be caused by bottlenecks, as are all the transitions observed at two other sites examined here. The evidence suggests that bottlenecks cause these transitions in a predictable way, and does not suggest that stoppages (jams) appear spontaneously in free flow traffic for no apparent reason. It is also shown that many of the complicated instability phenomena observed at all locations can be explained qualitatively in terms of a simple Markovian theory specific to traffic that does not necssarily include spontaneous transitions into the queued state as a feature.

This paper provides additional empirical observations and an alternative explanation of some traffic data from a German highway described previously in Kerner and Rehborn (1997). This work will be referred to as 'PRL,' and figures from this work will be denoted 'Fig. PRL#.'[*][1] Other observations are contained in Kerner and Rehborn (1996a,b). Section 1 describes the explanation and Section 2 presents a possible modeling framework.

1. Interpretation of some empirical evidence

A few studies of North American freeways have also identified related traffic flow properties and the origin of traffic disturbances (Cassidy and Windover, 1995; Cassidy and Bertini, 1997). Reassuringly, the observations on both continents are consistent with each other. However, it will also be shown that neither the German nor the North American data support the PRL conclusion:

that free flowing traffic will spontaneously break down randomly, without obvious reasons, and then remain in that state due to traffic's tendency to self-maintain congestion. Rather, the evidence indicates that traffic *breaks down* (queues form) at locations of freeway inhomogeneities (bottlenecks) due to reproducible exogenous reasons, and that, following breakdown, the bottleneck flow behaves in a predictable way. We see no evidence that queues self-maintain away from a bottleneck.[2] Highways appear to behave like crowds of people going through a series of queues at well defined locations; e.g. fans who wait in line to buy football tickets and then queue through one or more gates to enter the stadium. Some definitions derived from this analogy are introduced below; see e.g. Daganzo (1997a) for additional explanations.

Free flow (or *unqueued)* conditions refer to that state of traffic where small disturbances can only flow forward in space (e.g. as in supersonic gas flow). People walking away from the ticket window, for example, are in free flow because if one of them momentarily slows down just a little, this does not affect the queue upstream of the server. The free flow status of highway traffic can be determined by correlating the curves of cumulative flow at successive detectors with a time lag, as described in Cassidy and Windover (1995) and shown below. Conditions that are not free flow will be called *queued.* In queued traffic, disturbances can and do flow upstream. *Bottlenecks* are those inhomogeneous locations such as a ticket window where queues can form and persist with free flow downstream. A bottleneck in this state will be said to be *active.* Note that a bottleneck may be deactivated by a queue from a downstream bottleneck that blocks it and also when its own queue dissipates. If the flow through an active bottleneck is nearly constant and reproducible this flow will be called the bottleneck *capacity.* The evidence available in German and North American traffic data contains nothing at all to contradict this analogy; i.e., the evidence indicates that queues grow when bottlenecks become active and that instabilities are the result and not the cause of queues.

1.1. Merges

Included here is evidence that traffic instabilities upstream of an active merge bottleneck appear not to influence the bottleneck flow significantly, i.e. one can define a *merge capacity,* that queues grow shortly after the merging flows exceed this capacity, and that queues dissipate predictably. This explains the breakdown described in Figs. PRL1, PRL2 and PRL3, as well as those observed on North American sites.

Fig. PRL2(b) shows that at 7:16 AM a drop in the time series of measured speeds for the passing lanes was recorded at the first detector (D3) downstream of the merge, while the speeds upstream and downstream of it remained higher. This change in speeds was later detected upstream and downstream, suggesting that the bottleneck became active and a queue began to grow near detector D3 at around 7:16 a.m. The sharper changes in speed observed later at upstream locations are what one would expect if the back end of such a queue was growing in the upstream direction; it passed over detector D2 at 7:22 a.m. and detector D1 later. It is further

---

[2] Queues can also grow at unpredictable locations due to observable incidents such as car stalls, and these queues invariably dissipate (they do not self-maintain) when the incident is removed as is repeatedly demonstrated in the field of incident detection (Lin and Daganzo, 1997). Neither inhomogeneities nor incidents were ruled out as causes of the queues described in PRL.

stated in PRL that the 'phase change' (i.e. the onset of queueing) was brought about by an increase in the on-ramp flow and that the phase change was self-maintained for about 2 h.

These observations are consistent with the simplest theories available, e.g. Daganzo (1996): when a queue is generated on the freeway, the bottleneck flow stabilizes at a predictable level and is maintained until the freeway queue dissipates. The PRL data presented suggest that the fluctuations in ramp flow are absorbed by the mainline flow, as predicted by the theory. Incomplete proof of this is contained in Fig. PRL2(e), which presents the rather stable time-series of flow in the left lane at the two detectors immediately downstream of the merge. The figure shows that these average flows did not change appreciably after the onset of queueing, suggesting that the cumulative number of vehicles entering the merge after this event did not change significantly either.

Far from exceptional, the behavior of the German bottleneck was reproduced qualitatively at a North American site analyzed in Cassidy and Bertini (1997). The main results are summarized below. The analysis is based on cumulative flow curves because they are more informative than the PRL time series data of counts and speed. The results are also presented in a form suitable for comparison with the PRL time series data.

The analyzed site is a segment of the Queen Elizabeth Way in metropolitan Toronto, Canada, with the geometry of Fig. 1(a). This freeway is instrumented with detectors that record vehicle counts, occupancies (i.e. the 'dimensionless' measure of density obtained by loop detectors) and
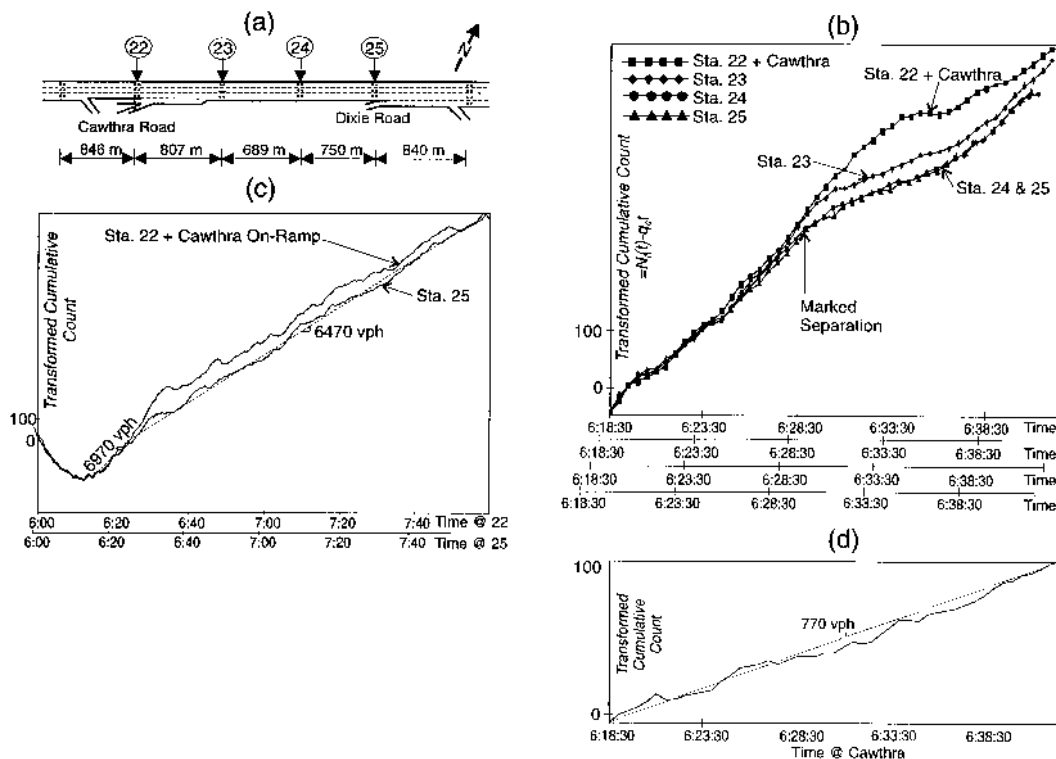

Fig. 1. Queue evolution at a merge bottleneck: (a) segment of Queen Elizabeth Way, Toronto, Canada; (b) cumulative curves showing queue formation at bottleneck; (c) cumulative curves showing features of the active bottleneck; (d) cumulative curve of on-ramp.

time averaged speeds over 30-second intervals. (The detector station labels shown are those used by the Ontario Ministry of Transport.) Typical of several other days that were analyzed at this location, the data shown here were gathered on the morning of 3 May 1995.

Fig. 1(b) presents transformed N-curves of cumulative flow (in all lanes) vs time for the four detector locations of our site during the onset of queueing. Recall that an untransformed N-curve gives the cumulative number of vehicles to have passed detector station i by a certain time, starting the counts (N=0) with the passage of a reference vehicle; see e.g. Newell (1982) and Daganzo (1997a). Thus, horizontal separations between N-curves are trip times and vertical separations the accumulation between detectors. In our figure, the curves (along with their respective time axes) have been shifted to the right by the average free flow trip time between the respective detector and detector 25, so that the vertical separation between curves now represents the excess vehicular accumulation between detectors due to vehicular delays. Such a shift is advantageous because two superimposed curves indicate that traffic in the intervening segment is freely flowing—since every feature of the upstream N-curve is passed to its downstream neighbor later. In addition, the figure only shows the difference between each curve and the line $N=q_o t$, where in this figure t represents the elapsed time from the passage of the reference vehicle at each detector location, because this background flow reduction magnifies details without changing the excess accumulations (Cassidy and Windover, 1995). The superimposed curve portions in this figure indicate that traffic was initially in free flow and remained in free flow between detectors 24 and 25. The marked separation of curves 24/25 from curve 23 from 6:29 a.m. onward (as shown by the arrow) indicates that a bottleneck was activated a little earlier between detectors 24 and 23. The subsequent separation of curve 23 from curve 22 indicates when the queue arrived to detector 23.

Fig. 1(c) presents the transformed N-curves for stations 22 and 25 for the entire rush. [Additional tests explained in Cassidy and Bertini (1997) confirmed that the bottleneck was active during this time.] Note the persistent displacement between these curves (i.e. the queue) while the bottleneck was active. As shown by the dashed line, the queue discharge rate measured at detector 25 varied slowly in time, but N never deviated by more than about 50 vehicles from this trend line with average rate 6470 vph. Also evident in the figure is the maximum flow of 6970 vph that persisted for 12.5 min during the onset of queuing.

Repeated observations at this and at another freeway merge in Toronto indicated that queue formations were always accompanied by brief periods of over-capacity (combined) arrival rates and that the long term average discharge flows were reproducible from day to day (Cassidy and Bertini, 1997), i.e. there was a capacity.[3] Fig. 1(d) shows that the on-ramp flow remained nearly constant during the onset of the queue. Having demonstrated that the Toronto queue formed due to a bottleneck, we next show that the German data are qualitatively similar to Toronto's and thus, the former likely describes the activation of a merge bottleneck as well.

Fig. 2 presents time-series plots of the average vehicle speeds at the Toronto site measured over 1-minute intervals in each travel lane and at each of the four detectors, as was done in PRL. As on the German site, speeds upstream of the bottleneck at detectors 23 and 22 [Fig. 2(c) and (d), respectively] drop markedly during the 'phase change' to the queued state. Also reported in PRL,

---

[3]These observations also indicated that the queue discharge flow fluctuated significantly and that it would return to the initial over-capacity level for periods of several minutes. Whether these over-capacity flows can be sustained for longer periods of time by some form of control is an important research question.

speeds dropped downstream of the merge when the queue formed (as occurred here at station 24 by 6:29 a.m.) with smaller reductions exhibited a little later further downstream (as occurred here at station 25). This is logical; it indicates that vehicles gradually accelerated after discharging from the bottleneck queue at an average rate equal to its capacity, as shown in Fig. 1(c) of this paper and in Fig. PRL3(a).

Fig. 3 presents scatter plots of 1-min occupancies (densities) vs flows in the left-most travel lane of the Toronto site. Unshaded, smaller circles denote observations subsequent to the arrival of
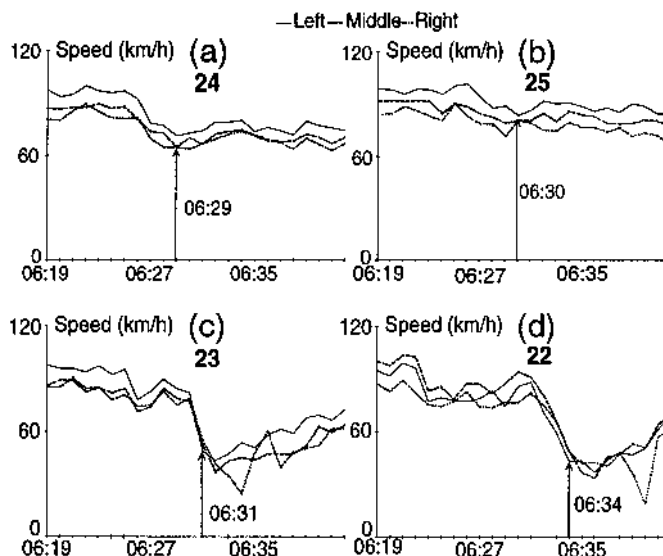

Fig. 2. Time series of vehicle speeds at merge bottleneck: (a) station 24; (b) station 25; (c) station 23; (d) station 22.
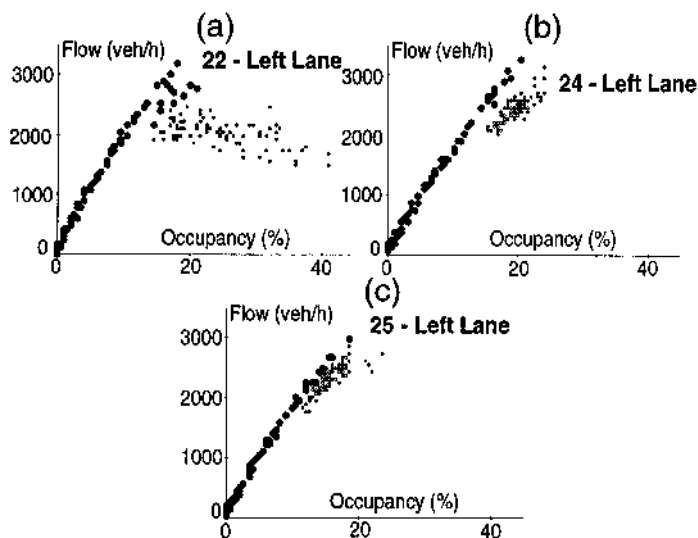

Fig. 3. Occupancy vs flow at merge bottleneck, 6:00 a.m. to 7:54 a.m.: (a) station 22; (b) station 24; (c) station 25.

the queue formation waves at each location. PRL demonstrates that the 1-minute flow-density data fluctuate markedly upstream of the merge, and this can also be seen here at station 22 [Fig. 3(a)]. PRL also contains plots similar to Fig. 3(b) and (c), again indicating that vehicles gradually accelerated downstream of the bottleneck. In fact, Figs 2 and 3 are so similar to Figs. PRL2(a-d) and PRL3(c) there is little doubt that the German data describe a bottleneck formed by merging vehicles.

1.2. Diverges

It is further shown in Fig. PRL4 that a phase transition into the queued state also occurred at a different location and time. This disturbance was characterized as spontaneous and due to a random occurrence because it formed away from a merge. The disturbance began adjacent to an off-ramp with much lower speeds in the shoulder lanes at locations near the exit (see detector D9 in Fig. PRL4) which is a 'signature' of a brief interruption caused by an oversaturated (i.e. queued) diverge.

A diverge on a narrow freeway is not qualitatively different from a single line of cars at an uncontrolled T-junction. Here, a left-turning car may force everyone behind to wait (e.g. because of opposing traffic) even if it is safe for other vehicles to proceed. Likewise, if a freeway off-ramp cannot accept the traffic wishing to exit, a queue entrapping some through vehicles will form on the freeway and reduce the flow, perhaps even blocking all the lanes further upstream. These effects are more pronounced if flows are close to saturation and the freeway is narrow as in PRL.[4]

The onset of queueing next to the off-ramp can be triggered in at least two ways: (1) a queue from the off-ramp spills-over and blocks the freeway traffic; and (2) the off-ramp is unqueued but an increase in the flow of vehicles wishing to exit creates a freeway queue when the off-ramp reaches capacity. According to prevailing theory, a recovery wave like that of Figure PRL4 should be issued from the off-ramp in the latter case if the flow of exiting vehicles approaching the off-ramp drops below the saturation level of the ramp. Note that the effects due to (2) would occur *even if the rate at which vehicles arrived to the queue and the exit ramp flow remained nearly constant;* i.e. they could be interpreted as occurring for no apparent reason. The information in Fig. PRL4 does not include the ramp flows, but from the node conservation law it appears that both the freeway flows (at station D9) and the ramp flows (the difference between the flows of D9 and D10) were close to saturation prior to the genesis of the disturbance. Its short duration suggests a disturbance of type 2, although it could also be due to exiting traffic (perhaps involving trucks) that required some lane changing. The development of this disturbance into a stoppage that propagates upstream is consistent with the theory proposed in Newell (1962) as explained in Section 2.

The situation observed in PRL is not exceptional. Evidence that off-ramp queues can block adjacent freeway lanes can be found even when freeways are wide and upstream flows well below saturation; i.e., when a disturbance would seem even less likely to affect adjacent lanes. As an illustration of this, Fig. 4(a) and (b) describe the evolution of an off-ramp queue on a segment of

---

[4]A mathematical theory of the diverge can be found in Sections 3.3 and 4 of Daganzo (1995), and a preliminary theory for wide freeways where traffic may sort itself by lane depending on destination can be found in Daganzo (1997b).
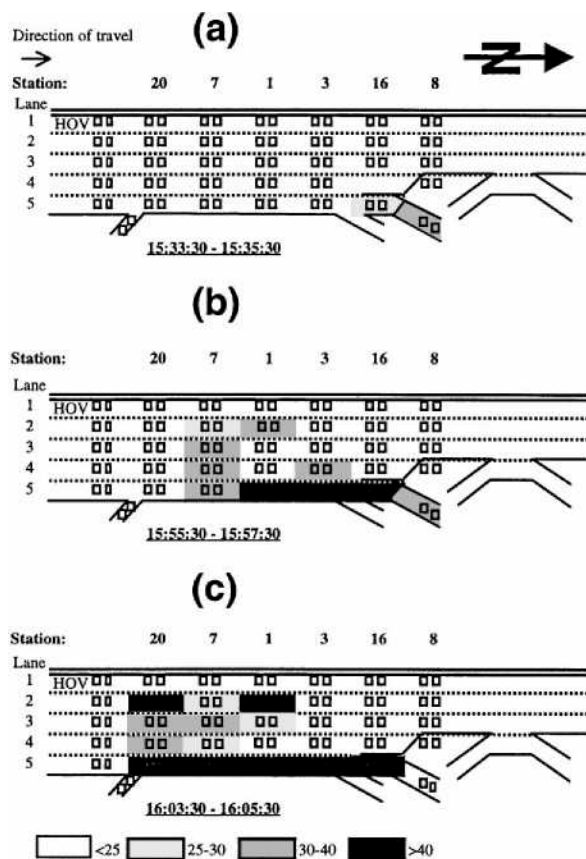
Fig. 4. Queue evolution at a diverge bottleneck: shading intensities indicate occupancy rates.

Interstate 880 in Hayward, CA. At its upstream end, the segment has four regular lanes and a median lane for car pools and buses, numbered as shown. (Also shown in the figure are the detector numbers used by the California Department of Transportation.) The shadings drawn here correspond to the 2-minute occupancies (densities) measured by the detectors at three different times during the afternoon rush on 8 March 1993. An occupancy of about 25% or more corresponds to queued traffic on this freeway (Cassidy, 1998) so that unshaded portions denote free flow. Fig. 4(a) shows that the disturbance originates at the off-ramp while (b) shows that as the queue propagated upstream, it moved in the transverse direction and blocked all regular lanes. The queue eventually passed detector 20; see part (c).

Fig. 5 presents the same data in the form used in PRL; data from the car pool and bus lane are not shown. Fig. 5(a) shows that the flows at the upstream end of the section were always well below saturation in most of the regular lanes. This explains why the queue did not propagate upstream of detector 20 [Fig. 5(h)] or reduce appreciably the output flow measured at detector 8 [Fig. 5(b)]. Fig. 5 also reveals the speed reductions brought by the queue. Three similarities with PRL stand out: (1) larger speed reductions in the shoulder lanes at locations near the exit; (2) upstream propagation of the speed reductions with a delay; and (3) equally low speeds in all lanes at locations sufficiently far upstream of the exit.
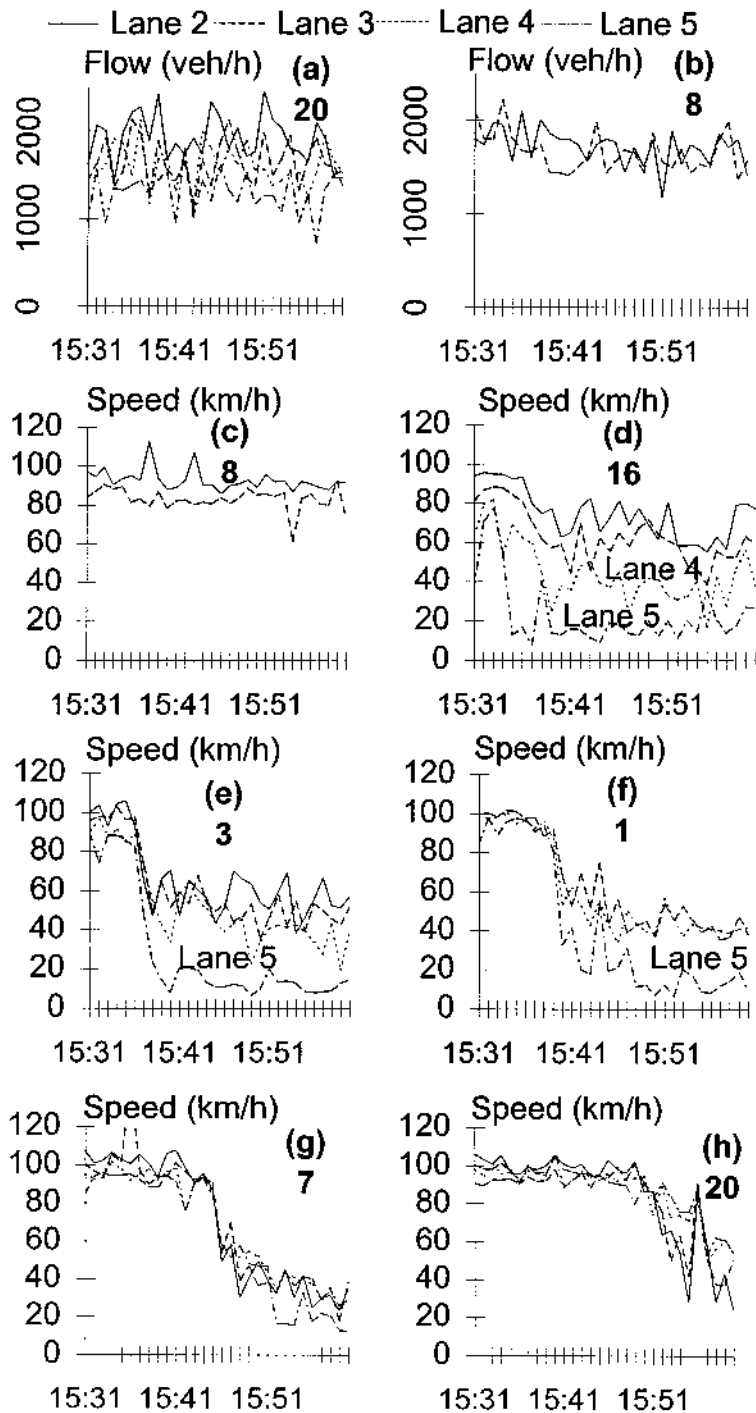
Fig. 5. Time series of flows and vehicle speeds at diverge bottleneck: (a) flow at station 20; (b) flow at station 8; (c) speed at station 8; (d) speed at station 16; (e) speed at station 3; (f) speed at station 1; (g) speed at station 7; (h) speed at station 20.

In summary, the attributes of the California speeds are similar to those from PRL, although the drop in the former persists for over 1 hour. Furthermore, although the flow drop on the California site was not severe, the occupancy shading on the off-ramp leaves no doubt that this ramp spilled over onto the freeway (reason 1) for more than 1 hour. Thus, the flow and speed data of Fig. 5, which might otherwise be even more puzzling than those of the German site, can be explained in terms of a clear cause.

2. Models

It was argued in the previous section that bottlenecks (and other exogenous causes such as incidents) can cause queues. None of the evidence suggests that queues form spontaneously. It also appears that traffic can flow relatively smoothly through a bottleneck and be unstable upstream of it. We ask here: is there a model of traffic behavior that can match these features?

It turns out that there are many models that will match qualitatively the observations in Kerner and Rehborn, (1996a,b, 1997), Cassidy and Bertini (1997) and Treiterer and Myers (1974) without spontaneous changes from free-flow into the queued state. These models do not need to borrow unreasonable features from the theories of materials flow such as waves that overtake vehicles; e.g. as occurs in 'Navier-Stokes-like' models (Del Castillo et al., 1993). Although we remain unable to discriminate among all possible reasonable models based on information available to date, this section shows how the observed phenomena can be linked to *one of the possible* models. Thus, we remain unpersuaded that: (i) queues can arise and self-maintain spontaneously in free-flow, and (ii) that the instability evidence implies a Navier-Stokes-like form for the equations of traffic flow, as suggested in Kerner et al. (1995).

The simplest macroscopic models specific to traffic are Markovian and *local,* in the sense that a platoon of cars, described by a state (q, k) on the flow-density plane, would respond to a (sustained) change in the speed of the lead vehicle always in the same way, independent of its past history and of other distant events. To specify a model of this type for a homogeneous highway it suffices to define two one-parameter families of curves on the (q, k)-plane, A and D. The two curves passing through a (k, q) point corresponding to the current state of our platoon would define the loci of future platoon states after a change in the lead vehicle speed. Family A would correspond to accelerations and family D to decelerations. Fig. 6(a) shows the transition that would be induced by an increase in speed from v_1 to v_2 when the relevant acceleration curve is as shown. Fig. 6(b) shows two possible families; the dotted (convex) lines correspond to accelerations, A, and the solid (concave) lines to decelerations, D. Fig. 6(c) shows another possibility. Diagrams like these completely specify the solution of any well-posed problem with piece-wise constant (k, q) data (e.g. an initial value problem) because they indicate unambiguously how the solution should evolve into the future at every singular point (solution of the Riemann problem). This will be shown below graphically. The solution method can be extended to an inhomogeneous highway where A and D would depend on x. If the slopes of the A and D curves are not allowed to exceed the vehicular speed, as occurs in Fig. 6(b) and (c), then information waves never overtake vehicles in the Markovian model.

The kinematic wave (KW) model is the Markovian special case where the two curve families reduce to the same unique curve. The Markovian family also contains one the earliest models of
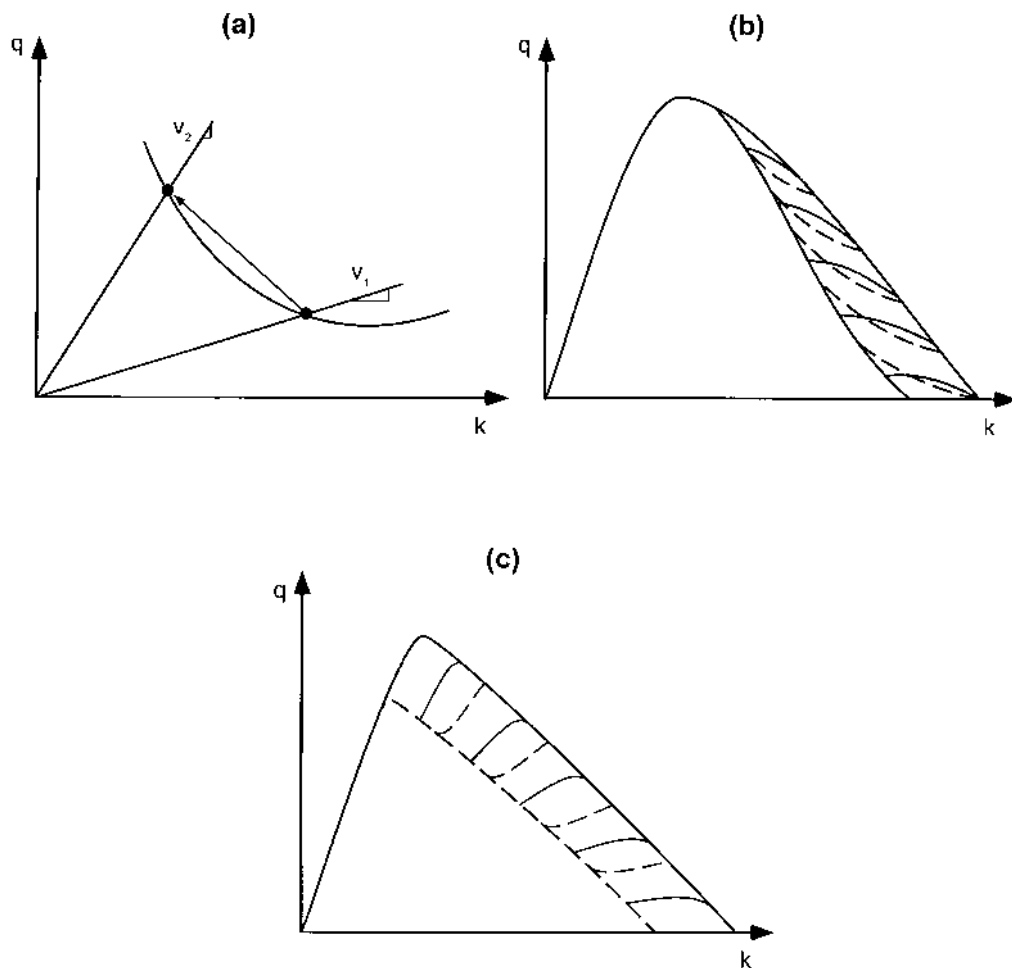
Fig. 6. Example transition curves for the simple Markovian model.

traffic instability (Newell, 1962). This model, which was motivated by Edie and Foote's classic observation of 'stop-and-go' traffic upstream of bottlenecks (Edie and Foote, 1959; Edie, 1963), will be used here to show that much of the observed phenomena in both German and North American highways can be explained qualitatively from a Markovian perspective.

It was noted in Edie (1963) that a decelerating (or just decelerated) platoon of traffic adopted a set of states on the flow-density (q, k) plane that were consistently different from those observed for accelerating platoons. It was speculated in Newell (1962) that these states could be described by two curves such as the D- and A-curves of Fig. 7(a), accelerated traffic always being on the A-curve, decelerated traffic on the D-curve and transitioning traffic in between; i.e. that the world was Markovian with A and D families such as those of Fig. 6(c).

Let us now consider the result of perturbations to a long platoon (queue) of vehicles in a given stationary state, represented by point o' on the q-k plane; see Fig. 7(a). Assume that one of the vehicles in this platoon allows a gap to grow in front of it and later resumes its original speed in a
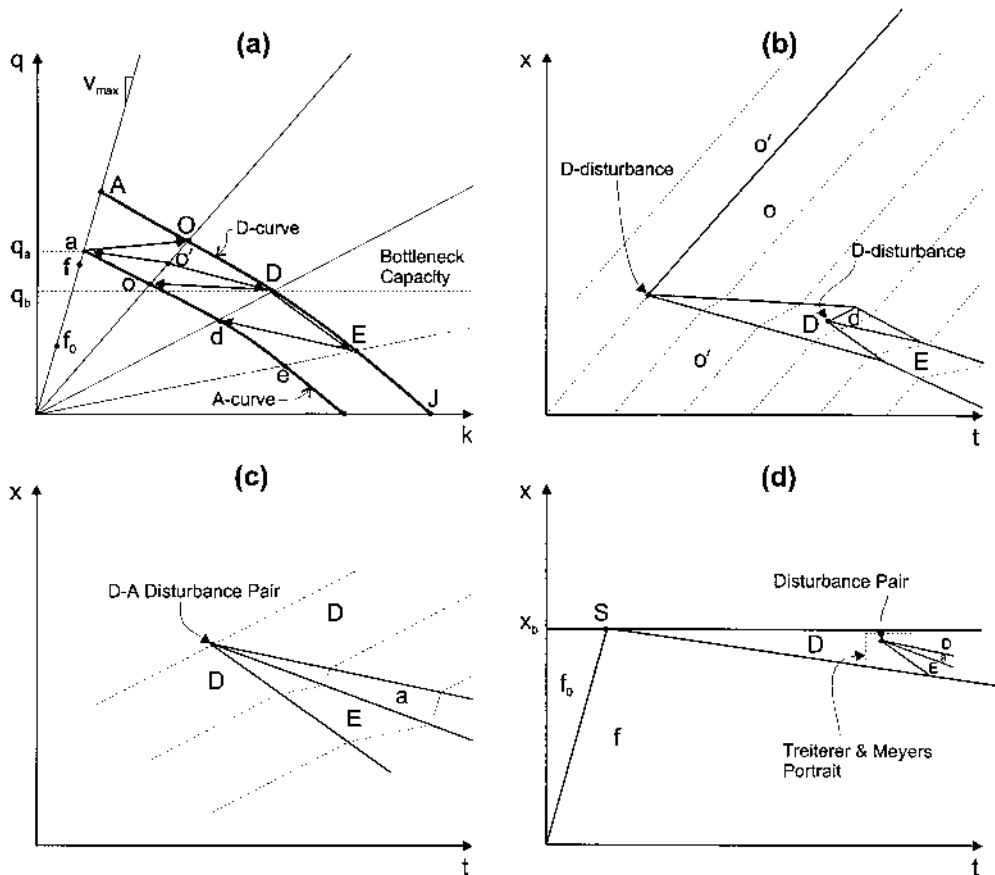
Fig. 7. Dynamical behavior of the simple Markovian model: (a) flow-density transitions (b-d) genesis and growth of various disturbances.

retarded position. This is called a 'deceleration disturbance'. Then, given the A- and D-curves, the response of the following vehicles can be predicted. One would expect the following platoon to decelerate, adopting a state such as D on the D-curve, and later return to its original speed, moving to state o on the A-curve.[5] A following vehicle would then experience state sequence o', D, o, as indicated by the arrows. The result for all vehicles can be conveniently displayed on the time-space (t, x) plane using a scale that includes many vehicles; see Fig. 7(b). The dotted lines of this figure are sample vehicle trajectories and the solid lines are waves that separate regions of homogeneity. Note that states D and o expand as they propagate through the line of vehicles. Observers stationed at fixed locations would see state sequence o', D, o upstream of the disturbance's origin and state sequence o', o downstream.

Consideration shows that an acceleration disturbance, introduced when a vehicle decides to close the gap in front of it, could cause a state sequence such as o', a, O to appear and expand into the solution-also shown by arrows on Fig. 7(a).

---

[5]Lower case letters are used for states on the A-curve and capital letters for the D-curve; the same letter is used to denote states with the same vehicular velocity.

If the initial state is on the D-curve (e.g. o'=0) we hnd that acceleration disturbances do not introduce expanding acceleration states into the solution because the two waves enclosing any such state would travel with the same speed. However, deceleration disturbances would expand (e.g. through a state sequence such as 0, D, o). Thus, state 'D' of Fig. 7(b) is unstable and will decay toward higher density states as additional perturbations occur; e.g., through the sequence D, E, d shown in Fig. 7(b). Eventually a growing 'jam' state, 'J,' would appear in the solution. This decay process is in agreement with the spreading 'jam' shown in Fig. 3 of Kerner and Rehborn (1996a).

The reverse occurs if the initial state is on the A-curve (e.g. o'=o). Then, state 'o' would evolve toward lower densities after an acceleration disturbance, e.g. through a sequence such as o, a, 0, but the same state would not decay toward higher density as a result of a deceleration disturbance.

Thus, we see that only states 'a'[6] and 'J' can persist in a queue; one should not be surprised to hnd a huctuating sequence between states 'J' and 'a' sufficiently far upstream of a 'lead vehicle.' This sequence is no longer disturbed by additional perturbations and would be propagated up the line of cars without change with a velocity equal to the slope of a straight line connecting points a and J of Fig. 7(a). This prediction is consistent with the unchanging propagating stoppages described in Figs 2 and 3 of Kerner and Rehborn (1996b).

This simple theory also implies that (k, q) detector data would tend to form clockwise hysteresis loops on the (k, q) plane, as is frequently observed, with deviations from this pattern due to statistical huctuations caused by driver differences. This can be understood from the hgure by considering the sequence of states seen by an observer of unstable traffic at a hxed location.

Insofar as traffic is not always unstable, one would expect deceleration disturbances of a size signihcant to trigger the effects of Fig. 7(b) to be relatively rare, e.g. when a truck moves into the platoon or a vehicle slows down to change lanes. However, if vehicles show a tendency to 'close the gap' (e.g. to prevent other vehicles from cutting into their line) so that they react to minor acceleration disturbances, then the platoon would behave as if these disturbances were continually present. Thus, A-states would be short-lived and would decay quickly toward 'a.' 0n a macroscopic scale where this process is not shown, each deceleration disturbance could then be modeled as a deceleration-acceleration pair of disturbances as in Fig. 7(c).

A plot on the how-density plane of the stationary states that appear at a hxed location on the (t, x) plane would then trace the D-curve. Indirect evidence of this is given in Fig. 11 of Cassidy (1998), which shows few non-stationiary points falling outside the stationary curve.[7]

This simple model can also shed light on a central question of this paper: whether queue instabilities can cause the how through an active bottleneck to drop signihcantly below the maximum possible and then sustain itself in this state for an extended period. Part (d) of Fig. 7 shows what would happen in this simple theory if at location x_b there was a bottleneck that would preclude hows greater than q_b from crossing it; see Fig. 7(a). As the surge of traffic 'f reaches the

---

[6]State 'a' corresponds to the largest speed that all vehicles in the traffic line can sustain; see Fig. 7(a). With A- and D-curves of different shapes the system may settle into a fluctuation between high and low density states that are not at the extremes of the curves.

[7]Some points fall outside the envelope for very low speeds, and this cannot be explained with the simplified model presented here.

bottleneck this theory would predict that vehicles would slow down and adopt a (decelerated) state D, consistent with the bottleneck service rate. If acceleration states are short-lived, as in Fig. 7(c), and insofar as $< q_a$, we see from the geometry that no waves with positive speed could ever reach the bottleneck; i.e. in this theory, with quick decaying accelerated states, disturbances could not affect the flow through the bottleneck. Disturbance pairs would look as the one shown on Fig. 7(d); they would grow through the queue in the upstream direction and would dissipate on reaching the back of the queue, as would occur in Fig. 7(d) if the drawing had been extended further in time. Not shown, the vehicle trajectories in the part of this figure that is enclosed in a rectangle would be qualitatively similar to the often-quoted data set presented by Treiterer and Myers (1974). The low speed of the lead vehicles in these data, combined with the high density in which they are embedded and the presence of backward-moving disturbances indicate that the disturbance in Treiterer and Myers (1974) was created within a queue as in Fig. 7(d). Thus, one cannot conclude based on Treiterer and Myers (1974) that such disturbances reduce the flow through a bottleneck, and much less that of a freely flowing traffic stream.

Finally, the reader can also verify that the introduction of a temporary bottleneck in a traffic stream that is close to saturation will generate a disturbance with the general characteristics of the one shown in Fig. PRL4, and that a time-dependent bottleneck (e.g. the freeway merge) will generate a series of upstream, 1-minute observations much as the light small circles of Fig. 3(a) and Fig. PRL3(c).

Note that disturbances within a queue could affect the flow through a bottleneck if acceleration states with flow less than 'a' were not short-lived. In this case states such as 'e,' 'd' and 'o' of Fig. 7(a) could propagate forward in space, as occurs for state o in Fig. 7(b), reach the bottleneck before having decayed and starve it for flow. Most researchers and practicing engineers do not believe that this is a significant effect, however, as evidenced by the standard 'capacity manuals' that are currently in use. This is one of the reasons why the assertions made in PRL would be extremely interesting if they could be observed and reproduced. In fact, they would not just be interesting, but also important for control purposes. If disturbances do not affect significantly the rate of bottleneck flow, then the time that a vehicle spends in a system of given length, including the trip time to reach the back of the queue plus the time within it, is independent of where the intermediate changes in speed take place. Control schemes, should then aim to keep the bottleneck flowing at capacity, which can be achieved relatively simply by preventing downstream queues from backing up into the bottleneck (Newman et al., 1969; Daganzo, 1996).[8] However, if queue disturbances could starve the bottleneck for flow, then control schemes would also have to be devised to prevent acceleration disturbances from persisting *upstream* of the bottleneck. This would be much more difficult to achieve.

To be sure, the simple macroscopic theory presented in this section does not explain everything. For example, it does not explain the oscillating transient that is recurrently observed in the data of Cassidy and Bertini (1997) as the merge bottleneck became saturated, which the theory idealizes by a singularity [point 'S' in Fig. 7(d)], and it does not explain some of the high density measurements in Cassidy (1998). But the theory gets enough things right to suggest that an improved model should probably be a modification of it.

---

[8]If it turns out that over-capacity flows can be sustained by metering, then one might try to achieve this by metering the on-ramps upstream of the bottleneck.

The theory is also incomplete because it does not explain the genesis of the disturbances or the rate at which they grow. Unfortunately, and despite the increasing amount of empirical evidence, no empirical studies to date describe the complete evolution of a disturbance through its entire life. Thus, although needed, extensions of the theory would seem premature.

3. Conclusion

A model of freeway traffic flow as a network of (possibly) unstable queues, with bottlenecks and particles that have specific destinations, explains qualitatively much of the observed phenomena including the most puzzling in Kerner and Rehborn, (1996a,b, 1997). Although it may be very difficult to discriminate among possible theories for the behavior of queued traffic, it seems premature to consider theories with spontaneous collapses. In any case, and quite fortuitously, it appears that much of what is important to travelers depends on the status of the bottlenecks, i.e., the delay they cause, and to a lesser extent on the detailed traffic patterns within the queue. Note however that observations of different bottlenecks indicate that the periods of oscillation and growth are quite site-specific. For example, the Holland Tunnel bottleneck of Edie and Foote (1959) exhibited regular oscillation periods on the order of 2 minutes, but observations upstream of other bottlenecks (such as those in Kerner and Rehborn, 1996a,b, 1997) are more sporadic. This should not be surprising, however, given that the vehicular interactions that take place at merge, diverge, weave, lane drop or sag bottlenecks are very different and may not be Markovian in space; they should be studied individually. Thus, insofar as the behavior of bottlenecks in their many possible forms is only partially understood today, additional experimental work seems warranted. These authors believe that bottleneck investigation is the research area where careful experimental observation will yield the greatest payoffs in the near future.

References

Cassidy, M., 1998. Bivariate relations in nearly stationary highway traffic. Transportation Research 32B, 49-59.
Cassidy, M., Bertini, R., 1997. Some traffic features at freeway bottlenecks. University of California at Berkeley Institute of Transportation Studies Research Report, ITS-RR-97-07 (also Transportation Research B, in press).
Cassidy, M., Windover, J., 1995. Methodology for assessing dynamics of freeway traffic flow. Transportation Research Record 1484, 73-79.
Daganzo, C., 1995. The cell transmission model part II: network traffic. Transportation Research 29B, 79-93.
Daganzo, C., 1996. The nature of freeway gridlock and how to prevent it. In: Lesort, J.B. (Ed.), Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Pergamon, Tarrytown, NY, pp. 629-646.
Daganzo, C., 1997a. Fundamentals of transportation and traffic operations. Pergamon, Oxford.
Daganzo, C., 1997b. A simple physical principle for the simulation of freeways with special lanes and priority vehicles. Transportation Research 31B, 105-125.

Del Castillo, J.M., Pintado, P., Benitez, F.G., 1993. A formulation for the reaction time of traffic flow models. In: Daganzo, C. (Ed.), Transportation and Traffic Theory, Proceedings of the 12th International Symposium on Transportation and Traffic Theory. Elsevier, New York, pp. 387-405.

Edie, L.C., Foote, R.S., 1959. Experiments on single-lane flow in tunnels. In: Herman, R. (Ed.), Theory of Traffic Flow, Proceedings of Symposium on the Theory of Traffic Flow. Elsevier, New York, pp. 175-192.

Edie, L.C., 1963. Discussion of traffic stream measurements and definitions. In: Almond, J. (Ed.), Proceedings of the 2nd International Symposium on the Theory of Traffic Flow. OECD, Paris, pp. 139-154.

Kerner, B.S., Rehborn, H., 1996a. Experimental properties of complexity in traffic flow. Phys. Rev. E 53, R4275-R4278.

Kerner, B.S., Rehborn, H., 1996b. Experimental features and characteristics of traffic jams. Phys. Rev. E 53, R1297-R1300.

Kerner, B.S., Rehborn, H., 1997. Experimental properties of phase transitions in traffic flow. Phys. Rev. Let. 79, 4030-4033.

Kerner, B.S., Konhauser, P., Schilke, M., 1995. Deterministic spontaneous appearance of traffic jams in slightly inhomogeneous traffic flow. Phys. Rev. E 51, 6243-6246.

Lin, W., Daganzo, C., 1997. A simple detection scheme for delay-inducing freeway incidents. Transportation Research 31A, 141-155.

Newell, G., 1962. Theories of instability in dense highway traffic. J. Opn. Res. Soc. Japan 5, 9-54.

Newell, G., 1982. Applications of queueing theory. Chapman and Hall, New York.

Newman, L., Dunnet, A., Meirs, J., 1969. Freeway ramp control—what it can and cannot do. Traff. Eng. 19(7), 14-25.

Treiterer, J., Myers, J., 1974. Hysteresis phenomenon in traffic flow. In: Buckley, D.J. (Ed.), Proceedings of the 6th International Symposium on Transportation and Traffic Theory. Elsevier, New York, pp. 13-38.