

Institute of Transportation Studies
University of California at Berkeley

**A Behavioral Theory of Multi-Lane Traffic Flow Part I:
Long Homogeneous Freeway Sections**

Carlos F. Daganzo

RESEARCH REPORT
UCB-ITS-RR-99-5

September 1999
ISSN 0192 4095

A BEHAVIORAL THEORY OF MULTI-LANE TRAFFIC FLOW. PART I: LONG HOMOGENEOUS FREEWAY SECTIONS

Carlos F. Daganzo

Department of Civil and Environmental Engineering

and

Institute of Transportation Studies

University of California, Berkeley, CA 94720

(September 21, 1999)

Abstract

This paper proposes a macroscopic behavioral theory of traffic dynamics for homogeneous, multi-lane freeways. The theory makes predictions for separate groups of lanes while recognizing that the traffic stream is usually composed of aggressive and timid drivers. Its principles are so simple that non-scientist drivers can understand them. The simplest version of the theory, which is described in its full complexity without calculus, is shown to be qualitatively consistent with experimental observations, including the most puzzling. Its predictions agree with the following phenomena: (i) the 'reversed lambda' pattern frequently observed in scatter-plots of flow versus occupancy and the lane-specific evolution of the data points with time, including the 'hysteresis' phenomenon, (ii) the lane-specific patterns in the time series of speed (and flow) in both queued and unqueued traffic, and (iii) the peculiar ways in which disturbances of various types propagate across detector stations. The latter effects include the evolution of both, stoppages and transitions between the queued and unqueued traffic regimes. The simple model is specified by means of eight observable parameters. The paper gives a recipe for solving any well-posed problem with this model and does so in sufficient detail to allow the development of computer models. A few approaches and possible generalizations are suggested. A sequel to this paper, devoted to freeway sections near on-ramps, will attempt to explain in more detail than previously attempted how queuing begins at merges.

1. BACKGROUND

Empirical freeway traffic flow data have accumulated for almost 50 years. While some of this evidence supports certain theories under particular conditions, it is fair to say that to date no theory explains all that is observed, and that some puzzles remain unsolved. Fortunately, new observations obtained with recently developed ways of processing data contain fresh clues that should make it easier to come up with a theory that puts all the pieces of the puzzle together. This paper is an effort in this direction; it presents a theory of traffic dynamics for homogeneous freeways that is qualitatively consistent with all the empirical observations (old and new) known to this author.¹

The theoretical development will begin in Sec. 2. The remainder of this introductory section lists empirical evidence. Facts that have been generally known since the 1950's by traffic engineers and/or widely reported in the literature are simply noted without discussion. To keep a historical perspective, older works are cited when possible. Facts have been grouped into classes (A, B, C, ...) and labeled individually (A3, B1, C2, ...). In this way the reader may proceed directly to Sec. 2, and later refer to this section as needed. The list follows.

(A) *Flow-occupancy data from single detectors on a single lane:* (A1) Plots of this type always show considerable scatter (see, e.g., Koshi et al., 1983, among many others) and the scatter is reduced when the counting intervals are increased. (A2) Plots for the lanes closest to the median (passing lanes) often exhibit a “reversed lambda” shape with very high flows and a discontinuity at the tip of the lambda; this discontinuity seems to have been first noted in Edie (1961). (A3) The very high passing-lane flows (exceeding 2500 vehicles per hour) can last for many minutes as noted for example by Cassidy and Bertini (1999). Therefore,

¹ A sequel to this paper will extend the theory to inhomogeneous sections that include on-ramps.

these high flows (and the reversed lambda) cannot be explained away by statistical variations due to driver differences. (A4) Such consistently high flows have not been reported for the shoulder lanes. (A5) There appear to be two clearly distinct traffic regimes with different properties: an 'uncongested' or 'unqueued' regime corresponding to the left part of the lambda and a 'congested' or 'queued' regime corresponding to the right part; see for example Edie and Foote (1958). (A6) When the data for the passing lanes are recorded for a period encompassing a rush hour with high traffic flows the sequence of points on the flow-density plane often includes a sudden drop from the tip of the lambda (or close to it) to a point on the right side. (A7) The evolution of the state on the right (queued) side is somewhat chaotic (e.g., Mika et al., 1969) with transitions that occur in no particular pattern, albeit grouped along an imaginary right leg of the "lambda"; flows rarely exceed the tip of the lambda. (A8) The quality of this grouping improves with the length of the sampling interval. (A9) Toward the end of a rush hour the data transitions back to the left side. (A10) It usually does so toward the middle of the leg (see e.g., Figs. 3-14 and 3-15 in M.E.A., 1971, and Figs. 2-2-31 and 2-2-32 in J.S.T.E., 1973), and flows near the tip of the lambda are not observed again until the following day. This hysteresis behavior, called the "traffic collapse" by some authors, has been known for a long time and has led many researchers to speculate about the existence of a "two-capacity phenomenon"; see, e.g., Banks (1991).

(B) Comparison of data from different lanes at the same location: It has also been known for a long time (e.g., Edie and Foote, 1958, Forbes et al., 1967) (B1) that for unqueued but heavy traffic conditions, the speeds and flows on the passing lane(s) of a freeway are consistently higher than those on the shoulder lanes, and (B2) that in queued traffic, when

passing is difficult, the differences across lanes are smaller.² Mika et al. (1969) also show through cross-correlation methods (B3) that temporal changes in the speed (or flow) time series of detectors for different lanes are synchronized across lanes.

(C) Comparison of data from different locations and the same (set of) lane(s): (C1)

Small disturbances in cumulative count within a single regime have been found to move quite regularly through the traffic stream; forward with the speed of traffic in the unqueued regime (Hillegas et al., 1974) and backward in the queued regime (Edie and Foote, 1958, and Mika et al., 1969). Recent experiments with an improved methodology (Cassidy and Windover, 1995, and Windover, 1998) show that small disturbances propagate without spreading. (C2) Regular propagation occurs in queued traffic even as one observes complex patterns on the flow-density plane (Muñoz and Daganzo, 1999).³ Invariably, (C3) the forward speed of the disturbances is comparable to the speed of traffic and (C4) the backward speed is on the order of 20 Km/hr. Larger disturbances such as transitions between regimes also show no significant evidence of spreading. This can be seen for example in the flow and speed data reported in Foster (1962), Kerner and Rehborn (1996a) and White et al. (1998), and also in the cumulative counts of Cassidy and Bertini (1997). The data in these studies show (C5) that the transition between the queued and unqueued regimes denoting the onset of congestion propagates spatially in the upstream direction rather regularly and sharply, i.e., (C6) as a non-spreading wave. The figures in Cassidy and Bertini (1997) also show (C7) that the transition back into the unqueued regime at the end of the rush hour also moves in the downstream direction in a non-spreading way. It is also evident from Figs. 2 and 3 in Kerner and Rehborn (1996a) (C8) that large stoppages

² Item (B2) is not necessarily true in queues caused by diverges; see e.g. Muñoz and Daganzo (1999).

³ Lane changing and driver differences can contribute significantly to the complex behavior observed on the flow-density plane. Thus, it is logical to find more regular patterns on data that averages these effects.

can propagate with sharply defined boundaries over long distances, and (C9) that both boundaries can travel for long distances with the same constant speed (which is again comparable with 20 Km/hr). It is interesting to note (C10) the absence of systematically spreading waves in these works, and in the works reviewed in Newell (1999). More recently, Kerner and Rehborn (1999) and Mauch and Cassidy (1999) have both found (using different methodologies) (C11) that large oscillations in flow, speed and cumulative count increase in amplitude across the detectors spanning a long freeway queue and its intervening on-ramps. These data, however, do not reveal the cause and/or mechanism of the growth; e.g., whether it is due to the on-ramps or it is an inherent property of queued traffic. In any case, it has been found from experiments on sections without on-ramps, both on freeways and on single lane roads (Cassidy and Mauch, 1999 and Smilowitz and Daganzo, 1999), (C12) that cumulative counts within a queue obey approximately the kinematic wave (KW) model of Lighthill and Whitham (1955) and Richards (1956) with a constant wave speed of approximately 20 Km/hr; i.e., they can be predicted as proposed in Newell (1993). In other words, (C13) accumulations on homogeneous freeway sections appear to behave on a macroscopic scale as if there was a linear relationship between flow and density in the queued regime.

(D) Behavior of car platoons in a moving frame of reference: Edie and Foote (1958) and Treiterer and Myers (1974) tracked vehicle platoons (moving queues) spatially and showed how their flow-density-speed status evolved with time as the platoons underwent deceleration and acceleration cycles. Both references agreed in many details, but the Treiterer and Myers study showed a remarkable effect (cycle B of Fig. 4 of that reference) that was absent in the Edie and Foote data. Treiterer and Myers showed that as the platoon accelerated from 25 to 40 mph, flow within the platoon increased from about 1800 veh/hr to almost 3000 veh/hr *without an appreciable change in vehicular spacing (!)* This behavior (labeled “cycle B” in figures 3 and

4 of Treiterer and Myers) has been a longstanding source of speculation in the transportation literature, often used to justify questionable models.⁴ However, there is a simple explanation for it. First note that the Edie and Foote data (with no “cycle B” effect) were gathered in a tunnel, where lane-changing was prohibited, and that the Treiterer and Myers data (as stated on Fig. 1 of that reference) pertained to the median (fast) lane of an urban expressway where lane-changing occurred frequently.⁵ (Furthermore, as stated on p. 17 of that reference, vehicles were allowed to join and leave the platoon as it was being tracked.) Therefore, it is logical to assume that the “cycle B” effect is a result of lane-changing. Quite likely, as the platoon accelerated past 25 mph, the aggressive followers in the platoon sensed that the (timid) leader was about to move out of the way and they followed more tightly so as to prevent drivers in the neighboring lane from “cutting in” the queue. Some neighbors did cut in, as can be seen from Fig. 2 of that reference, and the combination of these two effects explains the tightness of the platoon during its acceleration from 25 to 40 mph. From then on, the “cycle B” pattern changed sharply and suddenly. The speed continued to increase but flows decreased and the platoon expanded. It is hard to believe that the sharp reversal could have been caused by anything other than the clearance of the lane by the slow platoon leader. (This, unfortunately, cannot be verified precisely from the Treiterer and Myers figures.) However, the expansion that ensued is explained by additional defections from the platoon (which can be seen in Fig. 2 of that reference) combined with drivers’ loss of incentive to close the gap and follow closely at the higher speeds.

⁴ Valuable as they are, it should be remembered that the Treiterer and Myers results say nothing about the gaps between platoons, and that as a result they offer an incomplete description of the traffic stream.

⁵ Lane changes are clearly identified in Fig. 1 of said reference. In fact, the source of the main disturbance (which has been surrounded by mystery) can be traced back to a lane change in front of a highly compressed set of cars. Its dissipation can also be traced to vehicles that avoid joining it by changing lanes.

(E) Traffic behavior near merges and the onset of queuing:⁶ Although much work has been reported to date on merges, it was not until recently that experiments were undertaken specifically aimed at understanding the activation process of a merge bottleneck.⁷ The key reference in this respect is Cassidy and Bertini (1997). This work is particularly noteworthy because unlike prior efforts, it tracked accumulations and delays upstream and downstream of the bottleneck throughout the study, and in this way conclusively demonstrated that all the observed phenomena were happening while the bottleneck was unimpeded by a downstream queue. The results confirm the so-called “two-capacity” phenomenon which engineers have tried to exploit with ramp metering (see, e.g., May, 1964, and Banks, 1991). More specifically, Cassidy and Bertini (1997) shows that (E1) prior to activation there is a period of time where flows on the passing lane and on the freeway as a whole are very high (tip of the reversed lambda), and that (E2) once the queue has formed the flow past the bottleneck is reduced. The reduction in flow was found to be only on the order of 10% when measured across all lanes. They further report that (E3): (i) delays for vehicles passing the merge increase very slightly during the period of high flow, (ii) this period of high flow can last for 20 minutes,⁸ (iii) it is followed by a sharp drop in flow with a rapid increase in queuing delay, and (iv) this flow finally rises again to a level that is sustained with minor fluctuations while the queue is present. (In some cases there is a brief bounce-back period with high flows between steps (iii) and (iv).) O

The empirical evidence for freeway sections near diverges is much more scant and not always in agreement with all that has been said so far; see Lawson et al (1999) and Muñoz and

⁶ This evidence will be not be used until the sequel to this paper, which pertains to merges.

⁷ A merge is said to be active if traffic is queued upstream and unqueued downstream; see Daganzo (1997b). The activation process is the process that creates a queue.

⁸ Periods of up to 40 minutes have been observed more recently; Cassidy and Bertini (1999).

Daganzo (1999). Therefore, the present theory only pertains to (homogeneous) freeway sections where diverge disruptions can be ignored.

In developing the theory, clues obtained from single facility data and/or the author's informal observation of traffic will also be used as guidance, even though most of these facts have not been independently verified. The following is a list:

(F) The “Los Gatos effect”: It was pointed out in Daganzo (1997a) that when the passing lanes are heavily traveled in unqueued traffic (F1) passing vehicles seem to be less willing to follow the “rules of the road” and move out of the way to allow faster drivers to overtake. This may occur because drivers in a fast-moving queue (or platoon) may want to avoid the difficulty of rejoining the fast lane from a slower speed, especially if these platoons are long and frequent, and also because drivers within the platoon may not know for sure that those following them are indeed faster. When/if this happens, the shoulder lanes may be underutilized. The author has experienced this phenomenon at three very different sites that were not affected by the disturbances from entrances and exits: (i) The main uphill grade of California State Highway 17 southbound from San Jose, outside the city of Los Gatos, (ii) U.S. Interstate 80 (eastbound, downhill) from Truckee to Auburn on Sunday evenings during ski season, also in California, and (iii) a few times on the freeway system from Munich to Aachen, in Germany (on level terrain). It should be noted that these events were observed in the U.S. and in Europe, despite considerably different speed limits and driver customs.

(G) Existence of fast waves: (G1) A transition between uncongested flow (about 10% occupancy) and lightly congested flow (about 30% occupancy) has been observed to travel upstream at speeds approaching 30 Km/hr.⁹ (G2) At the same site, denser queued states

⁹ This speed was estimated from low-resolution occupancy data and our estimate is therefore subject to error. The event in question was the result of an incident that was recorded on 3/3/93 on Highway I-880 (lane-3, southbound, leaving Oakland, California) as part of the “Freeway Service Patrol (FSP)” project. Part of the FSP

appeared shortly thereafter within the lightly congested state, but they seemed to grow more slowly and not to propagate as far upstream. (G3) A different kind of fast-traveling transition, which introduced a much more heavily congested state, has also been reported; see Fig. 3 in Kerner and Rehborn (1996b).

(H) Lane specific merging behavior:¹⁰ (H1) Cassidy and Bertini (1999) have found that a precursor of the queue formation at a merge is a slight drop in the speed of the passing lanes.

(I) Very light traffic: (I1) Hurdle et al. (1997) have reported that for very light traffic the (space) mean speed of traffic increases slightly with flow.

(J) Saturation flows: Traffic that has accelerated from a stop on a single lane (e.g., from a traffic signal) can reach flows approaching 2000 veh/hr, but not significantly more, even on freeways. It is conjectured that (J1) the maximum flow per lane in a queue spanning all lanes of a freeway is also in this neighborhood. O

The remainder of this paper proposes a theory of driver behavior that attempts to reconcile all these seemingly disparate observations and then quantifies the theory with an idealized model. The model is only a caricature of reality but does not introduce any obviously unrealistic results.¹¹

data set can now be examined online by visiting <http://www.stat.berkeley.edu/users/fspe/> (Zhang and Rice, 1999).

¹⁰ This information will not be used until the sequel.

¹¹ Although it would be easier to develop and test more complex models with microscopic computer simulations, this was not done in this paper because computer testing can never rule out with certainty the presence of some undesirable phenomena. Continuum analytical modeling, by contrast, allowed us to eliminate very quickly many candidates that were shown to be bad in subtle ways, and to retain a final candidate that was found to be reasonable in all its dimensions. The reader should be able to verify this for his or herself simply by reading this paper. Of course, reasonableness does not prove that the proposed theory is correct, but the existence of a simple model that explains so much suggests that a correct theory must contain elements of human psychology, and lane-specific passing behavior.

2. DRIVER BEHAVIOR

This section describes the rules of driver behavior. A continuum model based on these rules is then presented in Secs. 3 and 4. Section 3 focuses on the steady state solutions, including a way of displaying them graphically, and Sec. 4 on the traffic dynamics. The paper ends (Sec. 5) with: (i) an illustration that shows how the most puzzling facts of Sec. 1 are explained, (ii) a brief list of new predictions that should be checked by experiment, and (iii) a proposal for some extensions to the model within the context of the proposed theory.

Regarding driver behavior, it is argued below that a correct model of freeway traffic flow should not only account for driver differences, as was attempted in some of the early theoretical studies of light freeway traffic (Newell, 1955, Carlsson, 1957, and Andrews, 1970), but also account for the changes in drivers' moods.

Driver differences: Item B1 and clue F1 suggest that different drivers have different tastes for driving fast. Therefore, the proposed theory should capture this effect. Drivers' preferences can be incorporated into a continuum theory much like destinations are captured in network models using the kinematic wave theory, i.e., where destinations (tastes) propagate forward with traffic. Unfortunately, the effects of fast drivers are more complicated than those of destinations in the simple KW model because the propagation of tastes for driving fast affects the character of the traffic stream. A similar complication was overcome in Daganzo (1997c) by means of some graphical diagrams and these diagrams, explained fully in the following section, will be the main solution tool here too.

Driver psychology and behavior: To complicate matters further, it appears that variations in tastes across the population of drivers are not sufficient to explain in a simple way the difference in flows between the high flows observed on the passing lanes (A2, A4, E1) and the lower flows observed within queues (A7) and discharging from queues (E2, J1). It is

therefore proposed that most drivers change psychology when driving fast on the passing lanes.¹² For example, they will allow someone to 'cut in' ahead of them but will drive closer to the car in front from then on. (This effect has been confirmed in informal interviews with drivers and can be occasionally observed when driving.) Drivers in this state will be said to be 'motivated'. It is postulated in this theory that the act of passing (or the anticipation of passing) triggers a psychological change into a 'motivated' frame of mind (for all drivers doing the passing) that allows them to accept headways as short as 1 sec. It is also assumed that this motivation disappears when passing on the fast lanes is no longer possible (e.g., when drivers join a slow queue). This behavioral pattern is consistent with the Treiterer and Myers (1974) effect; see item (E).

In view of (F1) it is also assumed that if multi-vehicle interactions are frequent (moderate to heavy traffic) then drivers will tend to ignore the rules of the road for passing, and will instead try to be in the fastest possible lane without exceeding their maximum desired speed. It is assumed, however, that the maximum desired speed is not changed by psychological changes.

Proposed idealization: To keep things simple, it will be assumed that there are only two types of drivers, 'slugs' with a maximum speed v_f and 'rabbits' with a maximum speed $V_f > v_f$.¹³ It will be shown that this simple model already captures the desired phenomena. Models with more driver types would be more realistic but also more complicated and data-intensive. This is further discussed in Sec. 5.

It will also be assumed that there are two lanes (or two sets of lanes), that slugs do not

¹² Such a change in psychology also seems to be the most logical explanation of effect (I1).

¹³ This terminology is preferred to a more serious alternative such as 'trucks' and 'cars', because we want to stress that rabbits and slugs are not perfectly correlated with vehicle size, and that the driver-type cannot be detected by the usual means.

go into the passing lane(s),¹⁴ and that rabbits always place themselves on the lane(s) with the highest speed. That is, rabbits place themselves on the passing lanes if the speed in these lanes is greater than v_f , and they change lanes so as to equalize the speed on all lanes otherwise. (That speeds below v_f cannot vary significantly across lanes is consistent with facts B2 and B3.)

The above set of assumptions are sufficient to enumerate all the possible stationary traffic states that can exist in this theory, as described below. A few additional assumptions, which are needed to describe the dynamics will be introduced in Sec. 4.

3. STATIONARY STATES

A stationary state is defined by a set of vehicle trajectories on the time-space plane that look invariant to translations in space and time. In view of our behavioral assumptions only two possibilities exist in this model: (i) a “2-pipe regime” where the speed on the passing lane, V , is higher than v_f and the speed of the shoulder lane is $v = v_f$, and (ii) a “1-pipe regime” where all vehicles travel with a speed below or equal to v_f . In the 1-pipe regime vehicles are segregated by lane, with all rabbits on the passing lane(s) and all slugs on the shoulder lane(s). Idealizations of these regimes using perfectly straight vehicle trajectories are displayed on Fig. 1b.

All the information necessary to describe a particular stationary state can be embodied by a set of six points on the flow-density plane, in the sense that if one was given an internally consistent set of points (differentiated by code as in Fig. 1a), then one could draw a set of

¹⁴ We recognize that this is unrealistic because slow vehicles may want to pass each other, and in the process they may create disruptions to the traffic stream on the fast lanes. However, our idealization is not unreasonable if the interruptions are infrequent because then they can be modeled as if they were exogenous and non-interacting.

vehicle trajectories with well-defined speeds and spacings for all the vehicle types, by lane. If diagrams such as those of parts (a) and (b) of Fig. 1 are drawn with scales such that parallel lines correspond to the same speed (as is approximately the case in the figures of this paper) then the time-space trajectories of each vehicle type should be parallel to the ray passing through the corresponding point on the flow-density plane; see Fig. 1. The average separation between the trajectories of a vehicle type, is likewise dictated by the position of the corresponding flow-density point along its ray.

Notation and representation of stationary states: The codes used in Figure 1 for the various types of points on the flow-density plane, and for the displayed vehicular trajectories will be used throughout this paper. Note that only the filled circle, the filled square and the dotted square are observable. In the text, the coordinates of these points (density and flow), and the points themselves will be identified by the letters (k, q) and p, respectively. Different fonts will be used to identify the classes by lane as follows:

capitals, K, Q, P \leftrightarrow passing lane \leftrightarrow solid dot;

lower case, k, q, p \leftrightarrow shoulder lane \leftrightarrow solid square;

boldface capitals, **K, Q, P** \leftrightarrow both lanes total \leftrightarrow dotted square.

These letters will be superscripted by primes to denote rabbits and by double primes to denote slugs; e.g., k," q", p" = shoulder lane slugs (white square). Note that six points describe a stationary state.¹⁵ Subscripts will be used to refer to a particular traffic state; e.g., p"_A and p"_B identify the two white squares of Fig. 1a that are labeled by the corresponding letters.

Note that every state on the flow-density plane can be represented on the (t, x)-plane by three well-defined sets of vehicle trajectories, as in Fig. 1b. White squares p"_A and p"_B of part

¹⁵ A symbol P" for slugs on the passing lane will not be used since it is assumed that P" = 0.

(a) for example correspond to the two sets of slug (slash-dot) trajectories in part (b). Note that only three sets of vehicle trajectories, and therefore only three of the six points defining a state, can be chosen freely because the following is always true:

$$p + P = \mathbf{P} \quad (1a)$$

$$p' + p'' = p, \text{ and} \quad (1b)$$

$$P + p' = \mathbf{P}' \quad (1c)$$

These expressions translate graphically onto the flow-density plane quite neatly because they represent the addition of vectors;¹⁶ e.g., (1b) says that “white circle + white square = filled square”. The reader should mentally verify that (1a), (1b) and (1c) hold for the data in Fig. 1a. Equations (1) allow us to represent a state with only three points; e.g., (p', p'', P) or (p', p, **P**). This will sometimes be done to avoid clutter. In those cases the reader can fill in mentally the remaining points.

Loci of stationary states: Figure 1a also contains a set of lines that further restrict the possible location of points in this theory. The bottom solid triangle defines the loci of the possible states for the shoulder lane (solid squares), and the discontinuous upper solid line the loci for the states of the passing lane (solid circles). A dotted ray with slope v_f marks the separation between the two regimes.

The three slanted lines ($Q(K)$, $q(k)$ and $\mathbf{Q}(K)$) with slope w are the loci of possible stationary points (for the shoulder lane, the passing lane and the total) in the 1-pipe regime, in agreement with fact C13. This assumes the same wave speed for all lanes, as one would expect in view of facts B2 and B3. The location of the three lines is defined by means of two constants,

¹⁶ It is true in general that the superposition of two stationary traffic streams on the (t, x)-plane corresponds to the addition of their vectors on the flow-density plane.

Q_m and q_m , which denote the maximum flows in the 1-pipe regime on each (set of) lane(s). These flows will be called “capacities”. A stationary 1-pipe state is defined by a set of six points satisfying (1) along a ray with slope $V = v \leq v_f$. If a backslash is used to denote the slope of a vector; e.g. $\backslash p''_A = v_f$, these conditions can be expressed mathematically as follows:

1-pipe conditions:

(equal average speeds on all lanes) $\backslash P = \backslash p' = \backslash p'' = v = V \leq v_f$ (2a)

(flow and density averages on the curves) $Q = Q(K), q = q(k), \mathbf{Q} = \mathbf{Q}(K)$ (2b)

Note that a 1-pipe state has two degrees of freedom, e.g., the speed of traffic and the fraction of rabbits (or slugs) on the shoulder lane, or equivalently just the position of point p' (or p'').¹⁷

A ray with slope V_f defines the possible states of rabbits traveling at their maximum speed (on the passing lane) in the 2-pipe regime. Rabbits' states can also be found to the right of this line with a speed V , such that $v_f < V < V_f$. In this case rabbits will be motivated and restricted to a sub-maximal speed by other rabbits ahead; their state will be found on a curve such as $Q_c(K)$. We will call this set of states a “semi-congested” 2-pipe regime because it represents a fast-moving queue of rabbits on the passing lane and a free-flowing stream of slugs on the shoulder lane. Note that in both 2-pipe regimes, $p' = (0,0)$ and $p'' = p$. Thus, there is perfect vehicular segregation by lane and 2-pipe states also have two degrees of freedom, e.g., the position of points P and $p = p''$ on their respective curves. The 2-pipe conditions can be expressed as follows:

¹⁷ The ray on which all the points lie is identified by (2a). It intersects the three loci, (2b), at unique points P , p and P . The fraction of slugs (or rabbits) identifies the position of p' (or p'') relative to that of p . Eq.(1b) then identifies p' (or p'') and (1c) does the same for P' .

2-pipe conditions (uncongested):

(speed conditions) $\lambda P = V_f ; \lambda p'' = v_f ,$ (3a)

(segregation conditions) $p' = (0,0)$ (3b)

(maximum flow conditions) $Q \leq Q_c$ and $q \leq q_m$ (3c)

2-pipe conditions (semi-congested):

(speed conditions) $V_f > \lambda P > \lambda p'' = v_f ,$ (4a)

(segregation conditions) $p' = 0$ (4b)

(maximum flow conditions) $Q \leq Q_c(K)$ and $q \leq q_m$ (4c)

The flows Q_c and Q_s at the intersection points of $Q_c(K)$ and $Q(K)$ with the V_f -ray will be called respectively the “critical flow” and the “saturation flow”. It is assumed that Q_c is considerably larger than Q_s and that $Q_c(K)$ defines a straight line with slope $W < w$ (see figure). The figure also includes another flow parameter, Q_d , which is not needed yet. Including this point, the complete figure can be drawn by specifying 8 measurable parameters. As a final aside, note that the speed of both lanes is uniquely defined by the speed of the rabbits, V , and that these relations are continuous and monotonic. This is true because in all regimes the speed of the left lane is V and the speed of the right-lane (i.e., the slugs) is: $v = \min (v_f, V)$.

4. DYNAMICS

The proposed theory applies to well-posed problems for which data have been specified on a proper boundary. A recipe is only given for problems in which the boundary is piecewise linear and where the data (specified in terms of flows and/or densities, as appropriate) is piecewise constant. Insofar as smooth boundaries and data can be approximated arbitrarily well

by piecewise linear (constant) functions the theory can be applied to any type of data. This simplification allows us to dispense with calculus formalisms, which would only lengthen the discussion in unproductive ways.

The dynamical theory is expressed in terms of 5 reasonable postulates:

- (P1)** (Stationary behavior): Only the stationary states of Sec. 3 can appear in a solution.
- (P2)** (Vehicular conservation): Vehicle trajectories of both slugs and rabbits are continuous.
- (P3)** (Stability): Interfaces between neighboring stationary states on the time-space plane must be stable to perturbations that smooth the speed of the rabbits; i.e. to all the possible continuous changes in the speeds on both lanes.
- (P4)** (Queue discharge model). There is a well-defined rate at which rabbits flow past the first slug of a 1-pipe queue in the capacity state when the restriction is removed and rabbits are allowed to travel at their maximum speed. (The rate could conceivably depend on the separation between slugs in the queue, but it will be assumed that it does not.)
- (P5)** (Maximum wave speed for deceleration): Cross-regime transitions for deceleration can propagate with a maximum velocity W' that is somewhere between w and W . (W' could depend on the upstream traffic state but it will be assumed that it does not.)

Postulate 1 is a reasonable first choice if one wishes to examine the simplest possible theory. Postulates 2 and 3 are self-evident. Postulates 4 and 5 could be relaxed with no substantive changes to the theory, but complications are premature in view of the scant empirical evidence that is available. The physical meaning of these two postulates will be discussed in more detail later.

Postulates 1 and 2 imply that neighboring stationary states must be separated by a straight line on the (t, x)-plane (an interface) where the trajectories of slugs and rabbits may bend, and rabbits may change lanes. This, of course, is an idealization.¹⁸ The slope of such a line must be such that the relative flows of rabbits (on both lanes) and slugs seen by two observers moving on both sides of the transition are equal. Letting U_{AB} denote the velocity of the interface separating two states A and B, we can express this condition as:

$$U_{AB} = \frac{P_A - P_B}{p''_A - p''_B} = \frac{P'_A - P'_B}{p'_A - p'_B} = \frac{P_A - P_B}{p_A - p_B}. \quad (5a)$$

In the 1-pipe regime, it follows from (2b) and Fig. 1a that in addition:

$$U_{AB} = \frac{P_A - P_B}{p_A - p_B} = w; \quad (5b)$$

i.e., that the interface propagates on both lanes synchronously with velocity w . Equations (5) have the same geometrical interpretation as the condition for the velocity of an interface in KW theory. In the present theory, however, interfaces are of two types: “simple”, which correspond to speed changes without any lane changing (as in KW theory), and “mixed” which introduce both effects concurrently. Equations (5) are not sufficient, however, to specify a unique solution for all reasonable problems. Postulates 3-5 are also needed.

It turns out that the unique solution to any reasonable problem can always be found by putting together the unique solutions of elementary (Riemann-type) problems for which the initial data at an arbitrary time consists of two neighboring stationary states separated by a discontinuity. This will be illustrated by means of an example in Sec. 5. Therefore, the solution

¹⁸ In reality these interfaces should have a characteristic length comparable with the sight distance, perhaps spanning tens of vehicle spacings. The characteristic length should be quite large (comparable with ½ Km, or perhaps even more) if the freeway is wide and lane-changing is involved. This means that if the interface is moving slowly, it can take many minutes to pass over a detector.

to all problems can be found if one knows the solution to all possible elementary problems. This set of solutions is developed from principles (P1) to (P5) in Secs. 4.1 and 4.2, below. The first time reader is advised just to skim the case-by-case results, and to focus on the overall logic which is given at the beginning of each sub-section. The list of elementary solutions is given because it establishes unambiguously that obviously undesirable effects (such as vehicles that respond to stimuli reaching them from behind) never arise in this theory, and because such a list can also be used for computerization.

4.1. Riemann problems with no regime changes.

This is the least interesting case because there is no lane-changing. Therefore all the interfaces are “simple”, and the solutions are always as in the KW model. This is illustrated in the three parts of Fig. 2.

Part (a) of this figure displays the typical solution for a problem consisting of two 1-pipe states. As required by Eq. (5) and shown on the left part of the figure, all the adjustments in flow, density and speed (for all vehicle types) must travel through the traffic stream with the same velocity, w , and occur synchronously. This is recognized on the middle part of the* figure, which shows the wave emitted after the acceleration of a lead vehicle. The solution is similar for a deceleration. In both cases it consists of a simple kinematic adjustment where the proportion of vehicle types in the traffic stream does not change. This should not be surprising since there is no passing in the 1-pipe state. The right part of the figure shows the solution to the Riemann problem without the vehicle trajectories. Note that the solution is completely specified by the wave velocity. In this picture (and others like it) legends next to each wave indicate the qualitative characteristics of the traffic stream that are changed by the wave.

Parts (b) and (c) of the figure describe problems consisting of two 2-pipe states. In these

cases slugs do not change their status, but rabbits may decelerate or accelerate. The latter and their effects on the traffic stream remain confined to the passing lane. These effects are as in the KW model with the flow-density curve followed by the (motivated) rabbits. As an example, part (b) of the figure (middle portion) shows a sustained reduction in speed of a vehicle that blocks the passing lane of a freeway. The freeway is initially in state A (left side) and the reduction is from speed V_A to V_B . The decelerated vehicles will form a fast-moving queue (platoon) behind the lead vehicle with minimum (motivated) headways. This queue will be in a state B' , where P is at the intersection of the ray with slope V_B and curve $Q(K)$, as shown on the left side. The back end of this queue will move with a velocity $U_{AB'} = \lambda(P_A - P_{B'})$, as shown. Note that the interface marking the end of the queue does not affect the slugs; under the right conditions it could even pass them. The right side of the figure is the solution to the Riemann problem. Because the density of slugs may be different in states A and B, the solution includes one more wave for the slugs. This wave is not crossed by the slugs; it simply separates the sets of slugs that travel (freely) with different densities. Waves of this type will be called “slips”. Clearly, state B' is characterized by p_A and P_B . As in the KW model several cases can arise depending on whether the problem involves an acceleration or a deceleration and the relative position of P_A and P_B . All cases will have a slug-slip and either one or two rabbit-waves. Part (c) shows an acceleration within the semi-congested state. An acceleration from the semi-congested state into an uncongested state would include an extra wave and a wedge with the critical state for the rabbits.

4.2. Riemann problems with regime changes

Regime changes are more complicated than the simple problems considered in Sec. 4.1 because they involve lane-changing and therefore require the introduction of “mixed” interfaces.

It will be convenient to consider separately three different cases: (1) acceleration transitions from 1-pipe to 2-pipe uncongested flow, (2) accelerations from 1-pipe to 2-pipe semi-congested flow, and (3) decelerations from 2-pipe to 1-pipe flow.

4.2.1. 1-pipe to 2-pipe (uncongested) transitions

As a preliminary step, let us consider what would happen to a long 1-pipe queue of vehicles in the capacity state, which will be abbreviated by the letter “C” from now on, if the obstruction causing the queue was suddenly removed. Imagine that the system is viewed from a frame of reference moving with the traffic at speed v_f and that we record the flow of rabbits passing the first slug. The flow in this frame of reference (the passing rate) is denoted Q_p . Figure 3(a) shows the state in which the system may be found initially and at two future times. Unless the density of slugs is extremely high, there is no a priori reason (from a qualitative understanding of what drivers might do) to expect a significant dependence between this flow and the composition of the queue. The figure shows how the front of the 1-pipe regime would recede (in the moving frame of reference). Although this front may have a characteristic width that should encompass several (or perhaps even many) vehicles, it is assumed that its width stabilizes (see facts C5-C9) and therefore that it should move with an average speed in agreement with (5a) in either frame of reference. Before this can be determined, we need to characterize the state downstream of the front, which we call the “discharge state” and abbreviate by the letter “D” from now on. (Note that “C” \neq “D”.)

Since the density is the same in both frames of reference the rabbit density in the frame of reference fixed to the road can be obtained with the formula:

$$(\text{relative flow, } Q_p) = (\text{density}) \times (\text{relative speed, } V_f - v_f)$$

and P_D can be identified graphically as in Fig. 3(b); the result is:

$$P_D = (Q_p / (V_f - v_f) ; V_f Q_p / (V_f - v_f)). \quad (6)$$

The second (flow) component of P_D is a constant of the problem which will be called the (rabbit) discharge flow on the passing lane, Q_d . We expect $Q_d > Q_m$. Note, however, that the flow on the shoulder lane equals the flow of slugs in the 1-pipe capacity state. Therefore *the total discharge flow is not fixed* and the velocity of the transition zone between regimes, which will be denoted w^* from now on, may vary. The expression for w^* is:

$$w^* = U_{CD} = \frac{1}{2} (P_C - P_D) , \quad (7)$$

as indicated geometrically in the figure. Note that w^* can be positive or negative, depending on the flow of slugs in state C; and that it will be negative if the flow of slugs is high. The determining condition is $Q_C - Q_D = 0$, and since $Q_C = Q_m + (q_m - q''_C)$, we find that w^* will be negative if:

$$q''_C > Q_m + q_m - Q_d. \quad (8)$$

Since we are assuming that $q''_C < q_m$, we see that for negative waves to be possible it is necessary that, $Q_d - Q_m > 0$. In view of this condition, we see that negative wave speeds are more likely in three-lane freeways (if the two lanes closest to the median function as passing lanes) than in two-lane freeways.

It is now possible to analyze all the Riemann problems involving a transition from a 1-pipe

state to the (uncongested) 2-pipe state. The three cases that can arise are summarized in Fig. 4. If we find that $w^* > w$, then traffic in the stable (physically possible) solution includes first a 1-pipe acceleration into the capacity state (as shown in parts (b) and (c) of the figure) and then a transition into the discharge state. This is followed by slug- and rabbit-slips that connect the discharge and the downstream state.

It is also conceivably possible, although very unlikely (and perhaps impossible if Q_d is not much larger than Q_m) that $w^* < w$. In this case, state C cannot appear in the solution, and traffic would have to transition into the discharge state from a different 1-pipe state. Consideration shows that if the flow Q_d does not depend on the 1-pipe state that is discharging then none of the intermediate 1-pipe states between B and C can appear in the solution, and the transition occurs directly from state B.¹⁹ Therefore the stable solution must look as in part (d) of the figure, where the velocity of the transition, U_{BD} , is given by (5) in the usual way.

4.2.2. 1-pipe to 2-pipe (semi-congested) transitions

Now we look at acceleration Riemann problems separating a queued 1-pipe state B and a semi-congested state A. The only difference between this case and the one just concluded is that now one must keep track of the back of the (fast) queue on the passing lane with an additional interface. We focus on the case with $w^* > w$. If the back of the queue that would be formed on the passing lane by the discharge state moves faster than the regime transition, i.e., if $U_{DA} > w^*$, then the solution would look as in part (a) of Fig. 5. The left side of this figure only includes the data points relevant for the question, and the right side only includes the interfaces

¹⁹ If Q_d were to depend on the speed of the upstream state from which the discharge is obtained, then an intermediate 1-pipe state between “B” and “C” could appear in the solution. The solution could then look as that of part (c), with “I” substituted for “C” and U_{ID} substituted for w^* . Because situations with $w^* < w$ are quite unlikely to arise, if they can occur at all, resolving this issue is not urgent.

that affect the speed of the rabbits. Note how they accelerate from v_B to v_f , then to V_f when they discharge, and how they finally decelerate to V_A on reaching the back of the fast queue; slugs remain undisturbed throughout. The right side of the figure should also include two slips (as in Fig. 4), which do not affect the vehicular speeds.

If $U_{DA} \leq w^*$, then the (fast) queue could not stay ahead of the discharge front and the discharge state could not appear into the solution. There would be a direct transition from state C to the semi-congested state, as shown in part (b). The cases with $w^* < w$ are not included in the interest of brevity.²⁰

4.2.3. 2-pipe to 1-pipe transitions

The idea here is to recognize that the rabbits would change speed gradually from V to v_f^+ (in the 2-pipe regime) and then to v_f and below (in the 1-pipe regime), and that stability considerations will tell us which of these intermediate states can grow into the solution and which must be absorbed in interfaces. The intermediate state with speed v_f^+ will be denoted “B” and the final state, “E”.

Figure 6 shows how state B is determined from an initial state, A, of high flow with the rules of 2-pipe decelerations covered in Sec. 4.1. It turns out that the complete form of the transition depends on whether or not the flow after the deceleration to v_f^+ , Q_B , exceeds the capacity flow for both lanes, Q_C , as illustrated by the examples in parts (a, c) and (b, d) of the figure. To understand these examples (and the additional cases that arise when the upstream flow Q_A is low) it is useful to imagine first what drivers would do as they experience the regime transition from semi-congested state B to state E; e.g., if they had to decelerate gradually due to a moving obstruction that gently slows to a speed v_E .

²⁰ We note as an aside that all the solutions in Secs. 4.2.1 and 4.2.2 are stable in the sense of postulate P3. That is, if the initial discontinuity in any of the problems is smoothed by introducing an intermediate step (any state with an intermediate rabbit speed) the solution does not change. (The methodology of Sec. 5 for solving general problems can be used to check that the intermediate state does not grow into the solution.)

The group of vehicles immediately behind the obstruction will decelerate gradually, initially from v_f to v_f^- , and they will change lanes as they do so. This should be clear because if rabbits and slugs decelerated on their respective lanes, the density on the passing lane would remain considerably higher and this would induce rabbits to change lanes. Clearly, the result of the initial deceleration to speed v_f^- and the concurrent lane-changing should be a more even (1-pipe) density distribution across lanes. The corresponding (1-pipe) state will be denoted "B⁻". Further decelerations would then be achieved by our group of vehicles in a purely kinematic way and be transmitted with velocity w , as discussed in Sec. 4.1 and shown in Fig. 2a.

A case that was not discussed in Sec. 4.1 arises, however, if $Q_C \geq Q_B$ because then B⁻ is not an equilibrium state. This requires an additional assumption. Namely, that decelerations from the non-equilibrium 1-pipe state B⁻ propagate faster than ordinary kinematic waves, with a velocity $W' < w$ (see postulate 5). This velocity could in principle depend on the state and be a function of Q_B , but there is no available empirical evidence (other than fact "G") to see if this is the case. Thus, it will be assumed for the remainder of the paper that W' is a constant. A value $W' \approx W$ will be used for illustration purposes. The existence of the fast wave means that vehicles upstream of our group would have to transition abruptly from state B⁻ into a well-defined equilibrium state, "I", when the fast wave hits them, and then settle into state E when the slow wave arrives. This is the only possible stable pattern.²¹ (Figure 6b shows how state I is identified by means of the slanted line with slope W' that passes through P_B .) It is now possible to solve the (Riemann) problems including a discontinuity between state A and state E, and have an intuitive understanding for the results; e.g., the solutions given on the bottom part of Fig. 6.

Consider first the case corresponding to the diagram of Fig. 6a, where $Q_C \geq Q_B$. The

²¹ The logic for this statement, given in Daganzo (1999), can be verified with microscopic simulations.

geometry of interfaces shows that the only states that can grow into the solution as a result of a smooth deceleration through states A, B, B⁻, E are corners in the convex lower envelope of the piecewise linear curve with corner-points \mathbf{P}_A , $\mathbf{P}_B = \mathbf{P}_{B^-}$, and \mathbf{P}_E . All other states must be absorbed in interfaces. Therefore, the solution may or may not include state B depending on the relative positions of points \mathbf{P}_A , \mathbf{P}_B and \mathbf{P}_E . The intermediate state B will appear if and only if $U_{AB} < U_{BE}$. This case is shown in part (c) of the figure. Otherwise, there is a direct transition from “A” to “E”.

The case corresponding to part b of the figure ($Q_C < Q_B$) is similar. One now needs to consider the relative position of points \mathbf{P}_A , $\mathbf{P}_B = \mathbf{P}_{B^-}$, \mathbf{P}_I , and \mathbf{P}_E . Two sub-cases can arise, depending on the relative position of points \mathbf{P}_A and \mathbf{P}_I . If as occurs in part (b) of the figure $U_{AI} < w$, then state B = B⁻ cannot appear in the solution and there must be a direct transition from state A to state I. This is then followed by a 1-pipe kinematic adjustment into state E. The prototype stable solution is as that shown in part (d) of the figure. Note that the 1-pipe change in speed can be positive or negative depending on the relative positions of points \mathbf{P}_E and \mathbf{P}_I on line Q(K). For the other sub-case (with $U_{AI} > w$) intermediate states (B, I) cannot appear in the solution and there is a direct transition from state A to state E.

Note that in every case direct transitions occur if Q_A is low. Otherwise, some intermediate states that expand into the solution occur upstream of the 1-pipe queue. Drivers would experience these precursor states before joining the queue. The propagation velocity of the precursor states can be positive or negative, but never smaller than W' . As illustrated by parts c and d of Fig. 6, the character of the precursor states depends on state A. In general, higher values of Q_A lead to more congested precursor states. Precursor states with $V = v = v_f$ and $V = v = 0$ are theoretically possible within the same model. This is consistent with the seemingly contradictory observations (G1-G2), and (G3).

5. DISCUSSION

The solutions of the Riemann problems (Figs. 2, 4-6) allow us to solve exactly any well-posed problem with a piecewise linear boundary and piecewise constant data, simply by drawing Riemann interfaces from every point of discontinuity on the boundary, and then stepping through time, solving additional Riemann problems where the interfaces intersect.²²

Example

As an illustration let us examine what would happen in this theory if an incident that blocked all flow occurred in a free-flowing but heavily traveled freeway. This is done in Fig. 7. The initial state of the freeway, denoted "A", is sustained by entering flow at $x = 0$ that stops at time $t = 3$. The incident is represented by segment S_0, S_4 in part (b) of the figure. The remaining lines in this sketch are results of the theory, as per the flow-density diagram of part (a).

Because the incident represents an inhomogeneity, a subsidiary theory is needed. It consists of assuming (reasonably) that the incident will separate two regions with no flow, a stoppage with $v = V = 0$ upstream, and an empty freeway downstream. The solution for the problem at S_0 can then be constructed by considering two Riemann subproblems, one at x_0^- (between states A and F) and the other at x_0^+ (between states O and A), and then piecing together the two answers. The result consists of two forward-moving slips and two backward-moving waves, all emanating from S_0 . Because Q_A is very high, the regime transition moves as a fast wave, which is followed by a precursor 1-pipe state and then by a deceleration into the stopped queue, as in Fig. 6d.

The next step takes place when the rabbit-slip denoting the absence of rabbits emitted by the boundary at $t = 3$ meets the fast wave at point S_1 . The Riemann problem at that moment

²² The general procedure is explained in more detail in Daganzo (1997a and b).

is between an upstream uncongested state with p''_A slugs and no rabbits, and downstream state I. With low upstream flow, the solution involves a direct transition, as discussed in sec. 4.2.3 and shown in Fig. 7. This procedure is then repeated at every intersection point by solving additional Riemann problems (always for a homogeneous freeway now). The sequence of points $\{S_i: i = 2, \dots, 6\}$ and the displayed boundaries are the result.²³ From these, one can easily construct flow-density scatter plots and cumulative N-curves.

The solution of this example is interesting because it illustrates that the proposed theory is consistent with several of the most puzzling phenomena in Sec. 1. Note in particular that:

- (i) The stoppage propagates without changing shape, which explains observations (C8) and (C9). This happens even though the flow into the stoppage is smaller than the (capacity) flow leaving it.
- (ii) The onset of congestion (the wave separating state A and I) propagates quickly upstream and a slower wave introduces the more severely congested state, F, which does not propagate as far upstream. This explains circumstantial observation (G).²⁴
- (iii) The flow and density scatter-plots obtained from detectors located upstream and downstream of the incident, e.g., at x_1 and x_2 , are as in Fig. 7c. The pattern matches qualitatively observations (A6)-(A10).
- (iv) If one included periods of time earlier in the day with lighter flows, leading to state A, and one were also to include the effects of a time-dependent bottleneck that lets through a variable amount of flow, instead of a complete stoppage, then it is easy to see that a reversed lambda pattern would be obtained on the flow-density plane

²³ Note that if a high flow stream were to be released from the boundary at time $t > 3$, then other precursor states could be introduced in the solution.

²⁴ It explains both instances because the level of congestion in the precursor state depends on Q_A .

for the passing lane, but not for the shoulder lane. This is consistent with items (A2)-(A5).

(v) In agreement with (C10), there are no spreading waves.

Although the proposed theory is not detailed enough to explain the “stop-and-go” oscillations of the 1-pipe regime, it should predict accumulations and cumulative flows reasonably well--see item “C”.²⁵

Predictions

The theory also makes some predictions that can be tested by experiment. For example:

(i) that a traffic “collapse” from an over-saturated state on the passing lane is possible even if the shoulder lane flow is low; (ii) that a fast wave and a precursor state occur behind the back of a 1-pipe queue only if the approaching flows are high; (iii) that a “reverse collapse” is possible, particularly in 2-lane freeways: i.e., that if the proportion of slugs is so low that (capacity) flow in a 1-pipe queue is higher than the highest possible uncongested flows, then by restricting the speed of an uncongested traffic stream with maximal flows to a value slightly below v_f one could force it into a 1-pipe regime and increase total flow; (iv) that observers downstream of an incident that has just been removed will first see fast cars, then a 1-pipe “discharge state” with a consistent flow on the passing lane(s) (probably on the order of 2000+ veh/hr), and that if the flow on the shoulder lane is high ($w^* > 0$) this will be followed by a capacity state with slightly higher flows on the shoulder lane and slightly lower speeds on the passing lane. (v) Conversely, a 1-pipe discharge state may be seen upstream of the incident after the capacity state if the stream is poor in “slugs”.

The proposed theory can be implemented by computer in a variety of ways: (i) by coding

²⁵ In the author’s opinion, more detailed models of the 1-pipe regime are premature until a better experimental understanding of the causes for instability emerges. If instabilities are triggered by lane-changing, as seems to be suggested by the Treiterer and Myers (1974) data, then the best approach would have to be probabilistic.

the exact procedure that was used to develop Fig. 7; (ii) by developing formulae for the flows observed at the location of the discontinuity in every Riemann problem and then using this information to predict the densities on a lattice with a finite difference approach (Godunov's method); (iii) by using the IT finite difference approach proposed in Daganzo et al. (1997); and (iv) by developing a micro-simulation using the principles of driver behavior that led to the theory and then checking that the macroscopic behavior of the simulation is consistent with what is expected.

It is also desirable to see if the rules of the theory can be simplified when the data are expressed in terms of N-curves (e.g., in terms of minimum principles for important special cases) because this could facilitate the visual interpretation of detector data.

Real driver behavior and modeling extensions

It was assumed in the simplified model that there were only two types of drivers and two sets of lanes. It was also assumed that traffic behaved as if there was a speed limit v_f on the lane (or lanes) closest to the shoulder and that drivers of both types chose to be in the set of lanes that allowed them to travel fastest. However, drivers never exceeded their desired speed and always segregated themselves so as not to encroach on faster lanes.

A more refined model, still within the proposed theory, would recognize that there is a continuum of desired speeds and would also recognize as many lane types as there are lanes, $\ell = 1, 2, \dots, L$. Such a model could still be based on the above-mentioned driving principles, using separate speed limits, v_ℓ , for individual lanes. If the lanes are numbered in increasing order from the shoulder to the median, the speed limits would be assumed to satisfy: $v_f = v_1 \leq v_2 \leq \dots \leq v_{L-1} \leq v_L = V_f$. The results of this model are not very different from those of the simple model; especially with regards to the regime transitions. For example, precursor 1-pipe states upstream of a queue still arise when heavy (over-saturated) traffic is interrupted, as occurred in Figs. 6(c,

d), and the removal of an obstruction that holds back a 1-pipe queue is still followed by a capacity state and/or a discharge state, as illustrated in Fig. 4(b-d). The main difference is that in the new model there is more than one discharge state, depending on the number of low numbered lanes that are flowing at their speed limit. In the present case, if ℓ lanes are flowing at the limit, one would expect the combined passing rate $Q_{p\ell}$ on the remaining lanes ($\ell+1, \ell+2, \dots, L$) to be fixed (and known). Therefore, the discharge flows $Q_{d\ell}$ on these lanes should also be fixed and known. The individual flows on the lowered number lanes ($1, 2, \dots, \ell$) should also be known since they are directly related to the distribution of desired speeds within the queue. Thus, the character of each discharge state, D_ℓ , is known. Clearly then, a group of fast vehicles released from the queue would experience known states $\{C, D_1, D_2, \dots, D_{L-1}\}$, in sequence. Depending on the initial composition of the queue some of these states may or may not propagate into the solution. If they all do, the solution would look as in Fig. 8. If total flow decreases across the sequence, as it is quite likely for low ℓ , then states with low ℓ would propagate in the upstream direction, also as shown. An upstream observer would then see a succession of semi-congested states with increasing speeds, first across all lanes until the shoulder lane reaches its limit, then across the remaining lanes until the second lane reaches its limit, etc. Conversely, a downstream observer should see the reverse process, with speed reductions starting at the median. Figure 8 reveals that the processes last longer if the observers are moved away from the original obstruction. The lane-specific form of these adjustments is the main difference between the generalized and simplified models.

We stress that the proposed theory is not accurate for very light traffic. Under these conditions, drivers should normally follow the “rules of the road”, staying on the shoulder lanes except when passing, and they may not behave as described. As traffic increases, however, queues of fast vehicles invariably form behind slower passing vehicles, and platoons are

observed on the passing lanes. As these platoons grow in length (with increasing flows), it becomes less appealing for platoon leaders to pull off to the shoulder lane and allow the queue to pass (see F1). At the same time, fast vehicles may stop pressuring those in front to pull off because they know they would find queue after queue, and that their travel time savings would be low. Thus, one would expect traffic to self-segregate by lane for medium to heavy uncongested flows, and in such cases the proposed theory seems reasonable.

Another generalization, still within the scope of the proposed theory, would allow vehicles to change their desired speed slightly in response to traffic conditions. It is not clear, however, whether this modification is sufficiently different from the previous proposal to deserve a separate pursuit.

Finally, note that in the proposed theory, a traffic stream can “collapse” from an over-saturated and uncongested state with $Q > Q_m + q_m$ and $V = V_f$ into a 1-pipe queue and an under-saturated capacity state if affected by a sufficiently large disturbance, e.g., a sustained reduction in speed to a level below v_f . The resulting traffic stream, however, cannot be put back into the original over-saturated state without external assistance. Thus, if there is a “collapse” there can be no spontaneous full recovery. In other words, over-saturated states are unstable when subjected to certain large perturbations. This means that one must still look for an exogenous mechanism that induces traffic to become uncongested and over-saturated in the first place. This author believes that merges provide the assistance through a “pumping” phenomenon. These ideas are described in a companion paper that attempts to explain evidence “E”.

ACKNOWLEDGMENT

This research was supported in part by PATH MOU's 305 and 3004. The author wishes

to thank Mr. Yuwei Li (U. California, Berkeley) for the artwork, and Profs. Michael Cassidy (U. California, Berkeley) and Masao Kuwahara (U. Tokyo) for some background information.

REFERENCES

- Andrews, F.C. (1970) "A statistical theory of traffic flow on highways, I- Steady state flow in low density limit. II-Three car interactions and the onset of queuing", Trans. Res. **8**, 359-365.
- Banks, J. H. (1991) "Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering?", Trans. Res. Rec. **1320**, 83-90.
- Carlesson, L. (1957) "Eu matematisk modell for landsvagsstrafik" Nordisk Matematisk Tidskrift **5**, 175-180.
- Cassidy, M. and Bertini, R. (1997) "Some traffic features at freeway bottlenecks" Institute of Transportation Studies Research Report UCB-ITS-RR-97-7, U. California, Berkeley. Also in Trans. Res. B, **33**, 25-42, 1999.
- Cassidy, M. and Bertini, R. (1999) (private communication).
- Cassidy, M.J. and Mauch, M. (1999) "An observed traffic pattern in long traffic queues", Trans. Res. (in press).
- Cassidy, M. and Windover, J. (1995), "Methodology for assessing dynamics of freeway traffic flow", Trans. Res. Rec. **1484**, 73-79.
- Daganzo, C.F. (1997a) "A simple traffic analysis procedure" Institute of Transportation Studies Working Paper UCB-ITS-WP-97-4, Univ. of California, Berkeley, CA
- Daganzo, C. F. (1997b) *Fundamentals of Transportation and Traffic Operations*. Pergamon, Oxford. U.K.
- Daganzo, C.F. (1997c) "A continuum theory of traffic dynamics for freeways with special lanes". Trans. Res. B, **31**, 83-102.

- Daganzo, C.F., Lin, W.H. and Del Castillo, J. M. (1997) "A simple physical principle for the simulation of freeways with special lanes and priority vehicles". Trans. Res. B, **31**, 105-125.
- Daganzo, C.F. (1999) "The lagged cell-transmission model" Proc. 14th Int. Symp. on Transportation and Traffic Theory. (Ceder, A., editor) pp.. 81-106, Pergamon, New York, N.Y.
- Edie, L.C. (1961), "Car-following and steady state theory for non-congested traffic" Opns. Res. **9**, 66-77.
- Edie, L.C., and Foote, R.S. (1958), "Traffic flow in tunnels", High. Res. Board Proc., **37**, pp. 334-344.
- Forbes, T. W., Mullin, J. J., and Simpson, M. E. (1967) "Interchange spacings and driver behavior effects on freeway operations", Proc. 3rd Int. Symp. on the Theory of Traffic Flow, (L.C. Edie, editor), pp. 97-108, Elsevier, New York, N.Y.
- Foster, J. (1962) "An investigation of the hydrodynamic model for traffic flow with particular reference to the effect of various speed-density relationships" Proc. 1st Conf. Aus. Road Res. Board, pp. 229-257.
- Hillegas, B. D., Houghton, D. G., and Athol, P.J. (1974) "Investigation of flow-density discontinuity and dual mode behavior", Trans. Res. Rec. **495**, 53-63.
- Hurdle, V. F., Merlo, M. I. and Robertson, D. (1997). "A study of speed vs. flow relationships on individual freeway lanes." Trans. Res. Rec. **1591**, 7-13.
- J.S.T.E. (1973) "A report on freeway capacity" (in Japanese), Japan Society of Traffic Engineers, February, 1973.
- Kerner, B.S. and Rehborn, H. (1996a) " Experimental features and characteristics of traffic jams". Phys. Rev. E **53**, R1297-R1300

- Kerner, B.S. and Rehborn, H. (1996b) "Experimental properties of complexity in traffic flow" Phys. Rev. E **53**, R4275-R4278.
- Kerner, B.S. and Rehborn, H. (1999) "Theory of congested traffic flow: self organization without bottlenecks" Proc. 14th Int. Symp. on Transportation and Traffic Theory. (Ceder, A., editor) pp. 147-171, Pergamon, New York, N.Y.
- Koshi, M., Iwasaki, M. and Ohkura, I. (1983) "Some findings and an overview on vehicular flow characteristics". Proc. 8th Int. Symp. on Transportation and Traffic Theory. (Hurdle, V., Hauer, E., and Stuart, G., editors) pp. 403-451, U. Toronto press, Toronto, Canada.
- Lawson, T., Lin, W. H. and Cassidy. M. (1999) "Validation of the incremental transfer model" PATH report (in press), Inst. Transp. Studies, Univ. of California, Berkeley, CA
- Lighthill, M.J. and G.B. Whitham (1955), "On kinematic waves. I flow movement in long rivers. II A theory of traffic flow on long crowded roads," Proc. Roy. Soc., A. **229**, 281-345.
- M.E.A. (1971) "A study of traffic congestion phenomena on metropolitan expressway" (in Japanese) Metropolitan Expressway Association, Tokyo, Japan, March, 1971.
- Mauch, M. and Cassidy, M. (1999) (private communication).
- May, A.D. (1964), "Experimentation with manual and automatic ramp control", High. Res. Rec., **59**, 9-38.
- Mika, H.S., Kreer, J.B., and Yuan, L.S. (1969) "Dual mode behavior of freeway traffic", High. Res. Rec. **279**, 1-13.
- Muñoz, J. C. and Daganzo, C. F. (1999) "Experimental observations upstream of an active bottleneck caused by a freeway off-ramp", PATH working paper, in press.
- Newell, G.F. (1955) "Mathematical models of freely flowing highway traffic", J. Opns. Res. Soc. Am. **3**, 176-188.

- Newell, G.F. (1993), "A simplified theory of kinematic waves in highway traffic, I general theory, II queuing at freeway bottlenecks, III multi-destination flows", Trans. Res. B, **27**, 281-313.
- Newell, G. (1999) "A lower order model." Institute of Transportation Studies Research Report UCB-ITS-RR-99-1, Univ. of California, Berkeley, CA
- Richards, P.I. (1956), "Shockwaves on the highway," Opns. Res., **4**, 42-51.
- Smilowitz, K. and Daganzo, C.F. (1999) "Experimental verification of time-dependent accumulation predictions in congested traffic," PATH working paper UCB-ITS-PWP-99-5, Institute of Transportation Studies, Univ. of California, Berkeley, CA
- Treiterer, J. and Myers, J. A. (1974) "The hysteresis phenomenon in traffic flow", Proc. 6th Int. Symp. on Transportation and Traffic Theory, (D.J. Buckley, editor), pp. 13-38, A.H. & A.W. Reed, London.
- White, J., Garside, S. and Whitaker, J. (1998) "A verification of a cell-type transmission model". World Conference on Transportation Research, Antwerp, Belgium.
- Windover, J. R. (1998) "Empirical studies of the dynamic features of freeway traffic", PhD thesis, Dept. of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, Berkeley, CA.
- Zhang, X. and Rice, J. (1999) "Visualizing loop detector data," Working paper, Dept. of Statistics, Univ. of California, Berkeley, CA.

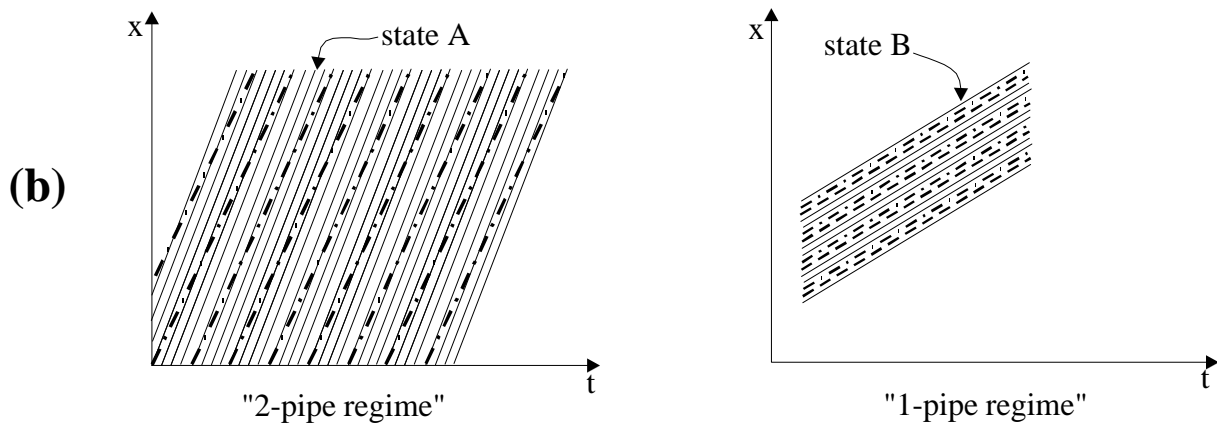
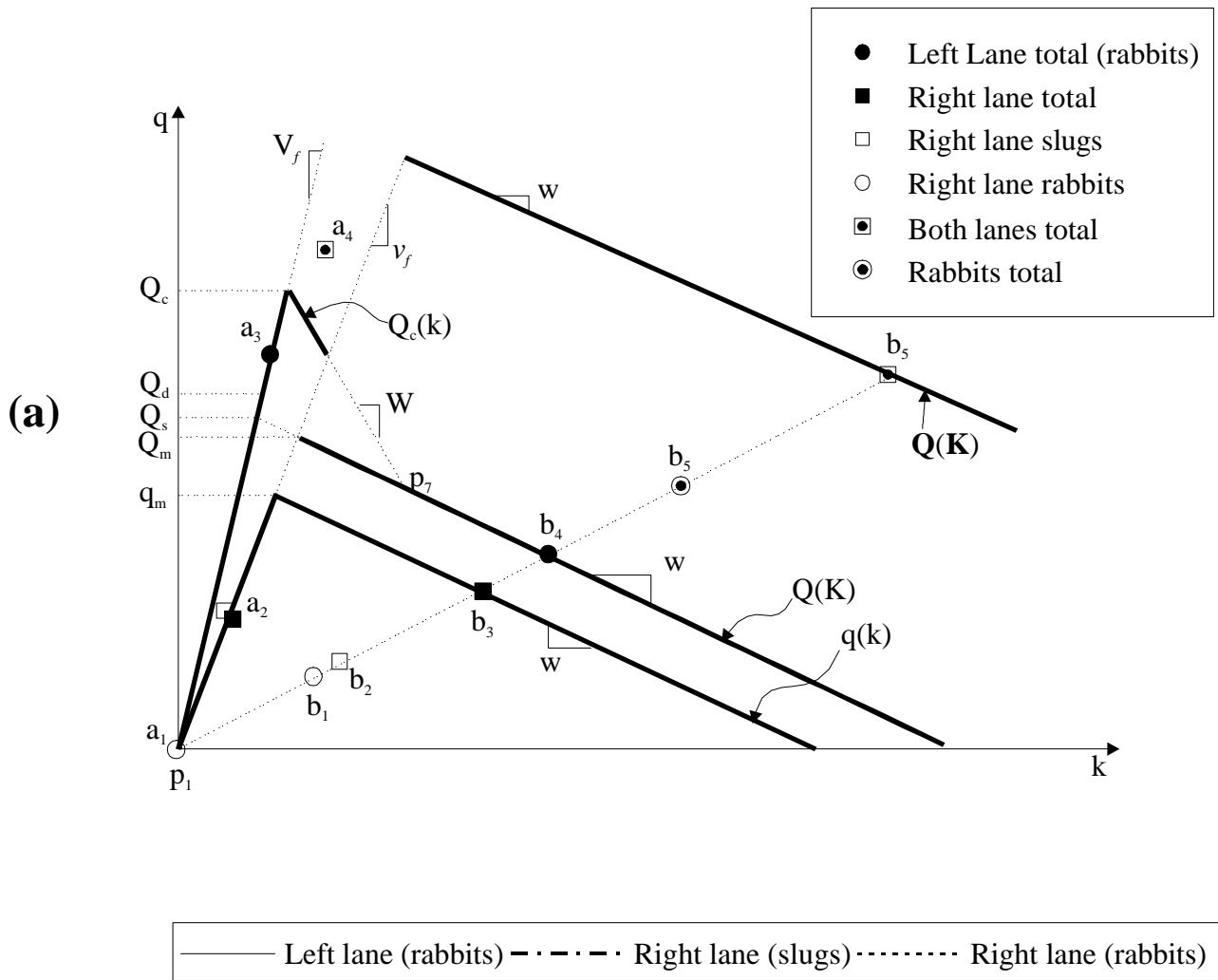


Figure 1. Possible stationary states: (a) Representation on the flow-density plane; (b) Realizations in space-time.

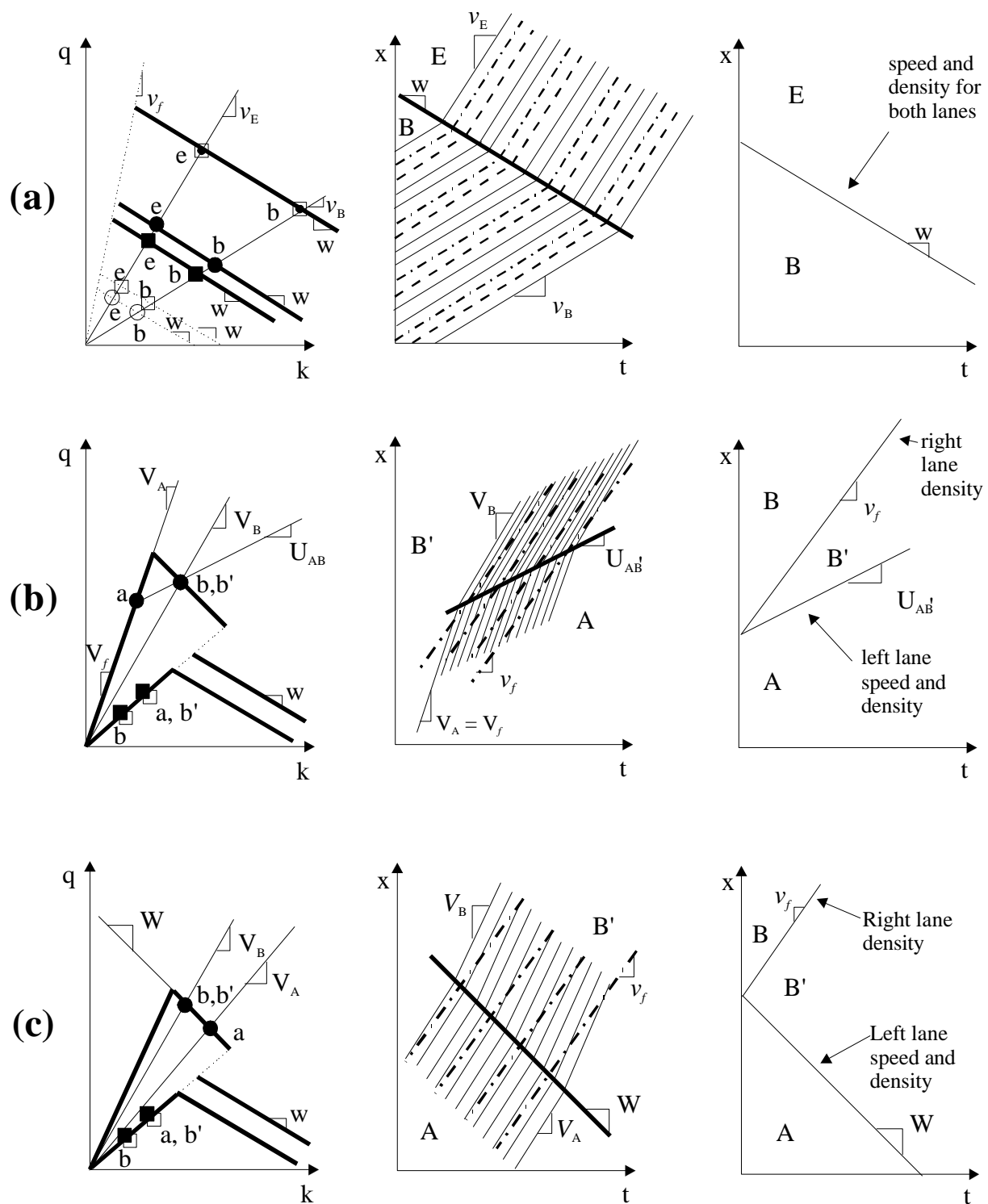


Figure 2: State transitions within the same regime for the lead-vehicle problem and the initial value problem: (a) 1-pipe regime; (b) 2-pipe regime (deceleration); (c) 2-pipe regime (acceleration)

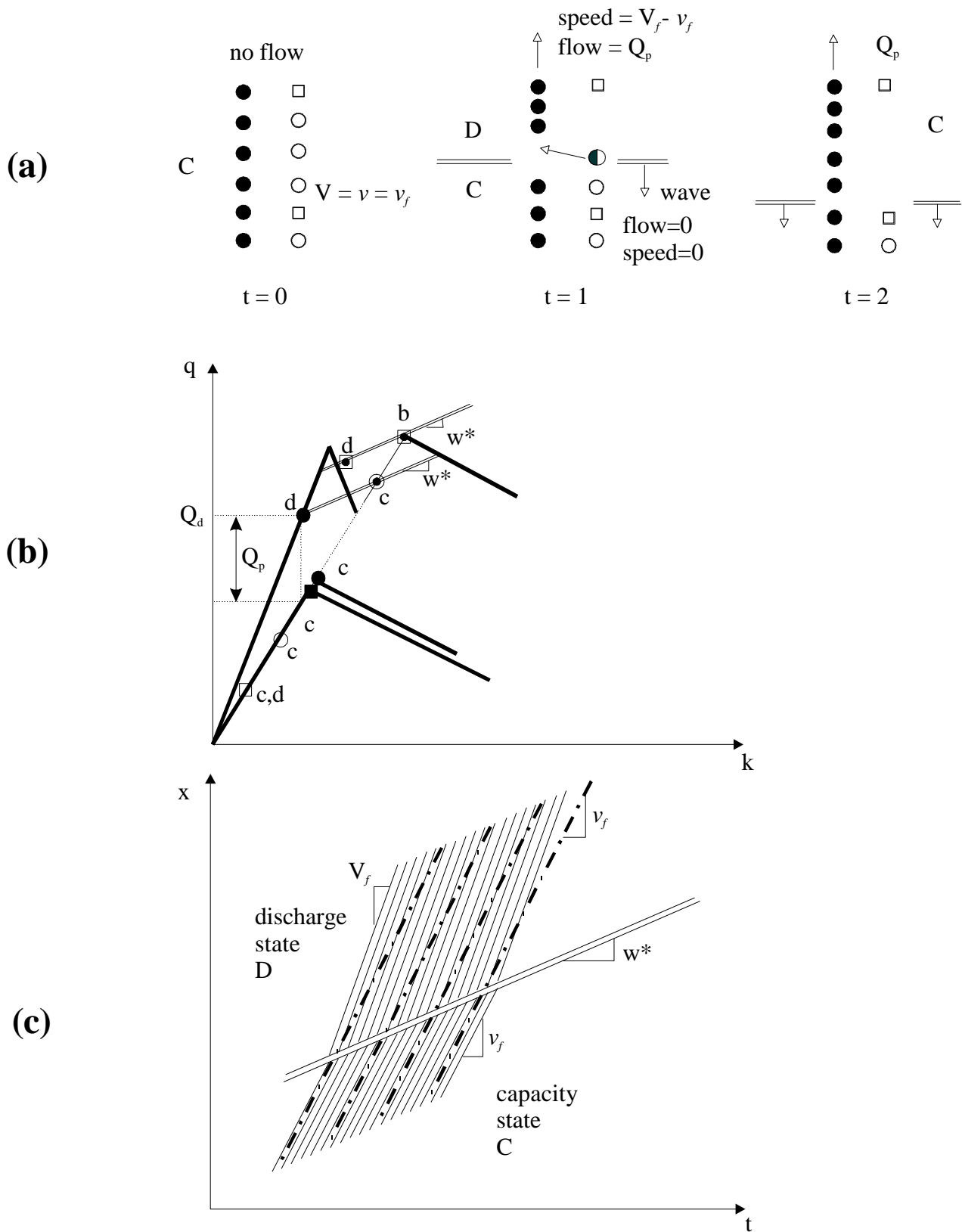


Figure 3. Model of the 1-2 (acceleration) regime transition:(a) Time-lapse pictures of system in a frame of reference moving with velocity v_f ; (b) and (c) q - k and t - x representation of the wave in the frame of reference attached to the road.

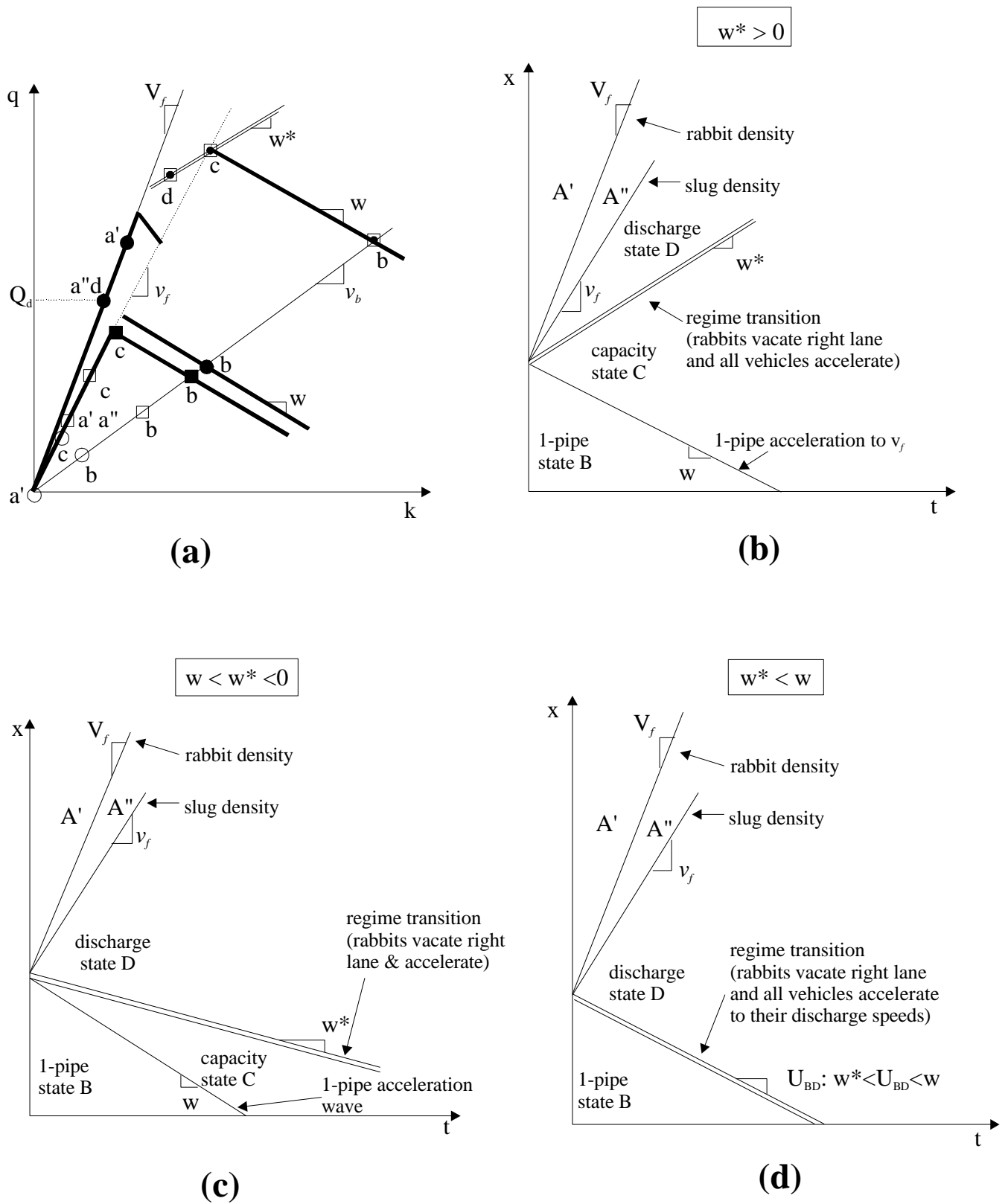


Figure 4: Initial value problems involving cross-regime transitions from 1-pipe to 2-pipe uncoupled flow: (a) identification of states on q - k plane; (b) (c) and (d) the three possible t - x solutions.

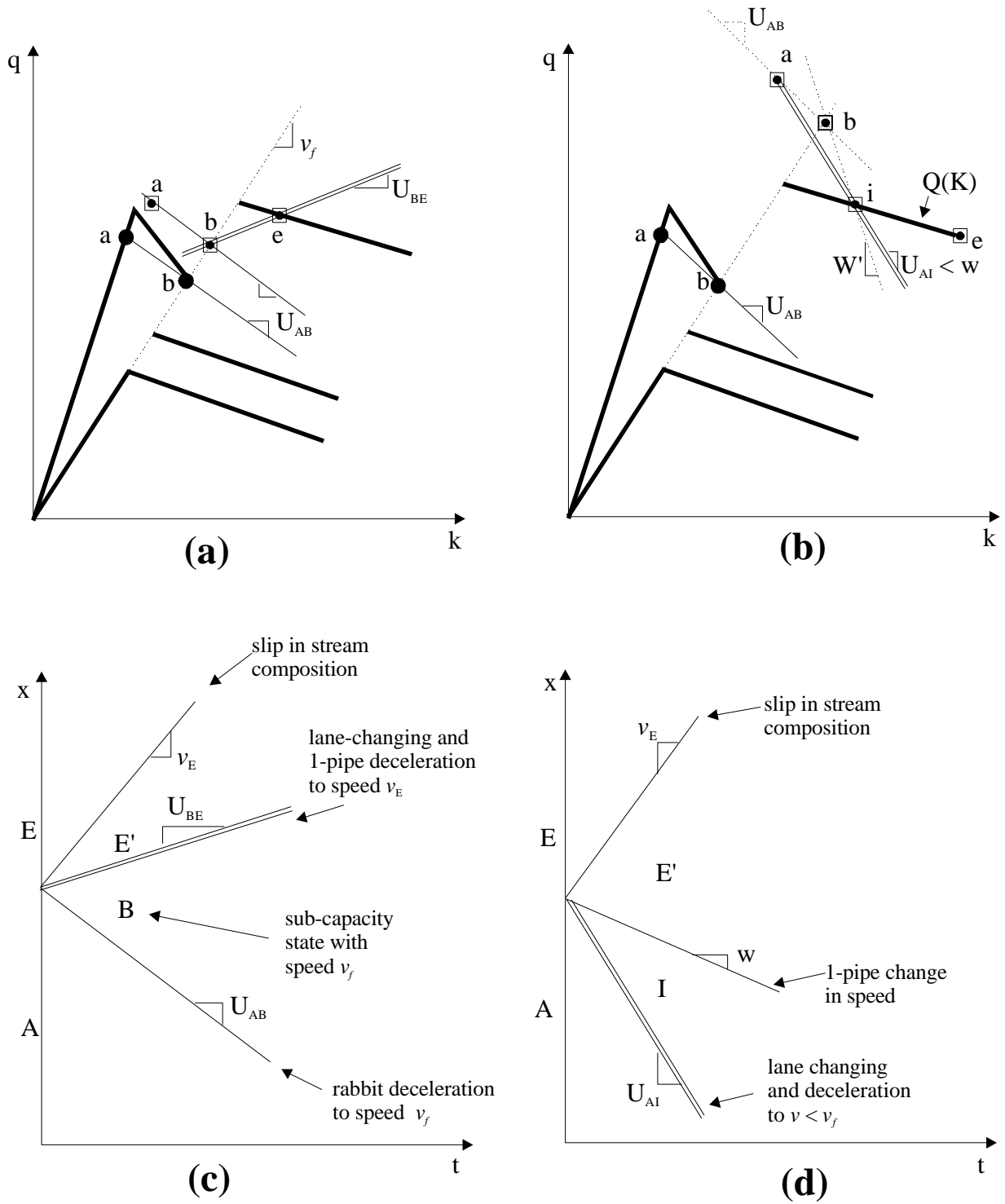


Figure 6: Cross-regime transition from 1-pipe to 2-pipe flow : (a) and (b), identification of states on the flow-density plane; (c) and (d), solution with "queue precursor" states.

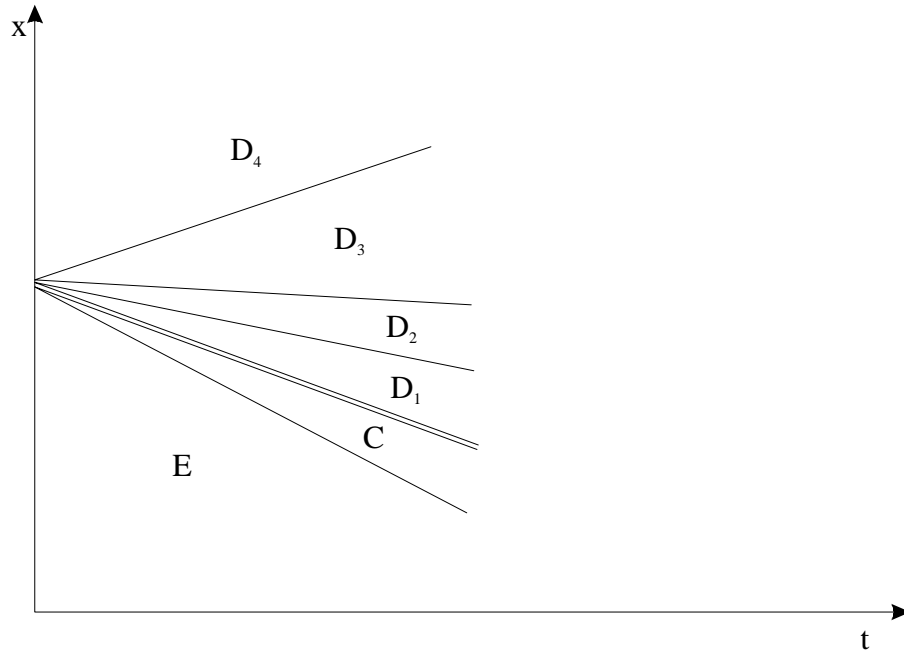


Figure 8: Disentangling of a 1-pipe state in the generalized model. (Slips are not included in the picture.)