

Institute of Transportation Studies
University of California at Berkeley

**A Behavioral Theory of Multi-Lane Traffic Flow Part II:
Merges and the Onset of Congestion**

Carlos F. Daganzo

RESEARCH REPORT
UCB-ITS-RR-99-6

October 1999
ISSN 0192 4095

A BEHAVIORAL THEORY OF MULTI-LANE TRAFFIC FLOW. PART II: MERGES AND THE ONSET OF CONGESTION

Carlos F. Daganzo
Department of Civil and Environmental Engineering
and
Institute of Transportation Studies
University of California, Berkeley, CA 94720
(October 19, 1999)

Abstract

This paper examines the behavior of multi-lane freeway traffic past on-ramps, building on the continuum model of part I and focusing on the onset of congestion. The main complication with merges is that rabbits (fast vehicles) entering from an on-ramp usually stay on the shoulder lane(s) of the freeway for some distance before merging into the fast lane(s). An idealization is proposed where this distance is taken to be the same for all vehicles. As a result, the system behaves as if there was a fixed buffer zone where entering rabbits cannot change lanes. The model of part I is extended to capture the peculiarities of traffic within such a buffer zone, including its two end-points: the “entrance” and the “merge”. The onset of congestion is described by means of waves. The paper shows that this highly idealized model (and its more realistic cousins) explain qualitatively all the puzzling facts discussed in part I without introducing obviously unreasonable phenomena.

A typical sequence of events during the onset of congestion is predicted to be as follows: (i) increased on-ramp flows and the ensuing merging maneuvers into the passing lane generate a (fast-moving) queue at the merge that grows on the passing lane(s) of the buffer; (ii) when the speed in this queue drops past a critical level (v_f) a 1-pipe queue forms at the merge, which then grows upstream across all lanes; (iii) if the front of this queue moves slowly forward, as per the model of part I, then the lane-flows at the merge point would be at “capacity” from then on (with roughly the same speed across all lanes) but downstream of the front there would be a “discharge state” with *less total flow*. Therefore, an observer downstream of the merge would see this discharge state before the capacity state, and would record a drop in flow followed by a recovery. If the front of the queue would move slowly backward, then the sequence of events following (ii) is somewhat different, as described in the paper.

1. INTRODUCTION

Part I of this report (Daganzo, 1999) introduced a model of traffic dynamics for homogeneous, multi-lane freeway sections. Although in this model traffic could be in a “2-pipe” over-saturated regime for an extended period of time (in agreement with experiment) the model included no mechanism for the creation of the over-saturated states in the first place. Therefore, the model of part I does not explain their existence. The present paper completes the picture; it proposes that merging traffic can get traffic into the over-saturated state through a “pumping mechanism” that follows necessarily from the driver psychology model of part I. It also describes the queue generation process in qualitative detail (by-lane and by longitudinal position) in terms that drivers could understand, and quantifies the process systematically with the different kinds of waves introduced in part I. Such detailed predictions should allow the model to be falsified (or verified) with new experiments. If it is falsified, the new empirical observations should lead to improved theories.

The focus of the paper is on the generation of queues by merging traffic. The paper is organized as follows. Section 2, below, highlights the most relevant results of part I. Section 3 describes the pumping mechanism. Section 4 examines the consequences of over-pumping and how queuing begins; it also shows that the predictions of this theory are consistent with what is observed. Finally, Sec. 5 offers some closing comments pertaining to experimentation and further theoretical work.

2. REVIEW: HOMOGENEOUS SECTIONS

Part I of this report proposed that under moderate to heavy traffic levels where passing is possible but not free, drivers in homogeneous freeway sections do two things: (1) they

segregate themselves by lane according to maximum desired speed (so that the speed of a lane corresponds to the speed of the slowest drivers willing to use that lane), and (2) they accept lower headways while passing. The special case of this theory that pertains to freeways with only two driver classes (rabbits and slugs) and/or two lanes (passing and shoulder) was examined in detail.

It was argued that there would be 3 kinds of stationary states characterized by the speed of the rabbits, V : free-flow/uncongested ($V = V_f$), semi-congested ($v_f < V < V_f$) and congested/queued ($V \leq v_f$). In all cases the speed of the slugs would be $v = \min(v_f, V)$. Accordingly, two regime types were defined: "1-pipe" (congested), where the speed of all lanes and vehicles is the same and "2-pipe" (uncongested or semi-congested), where vehicles are segregated by type and by lane. Because the lane-flows and lane-densities are related, a state can be characterized by only two quantities, e.g., the respective densities of rabbits and slugs. Given these two quantities, one can identify uniquely the flow-density-speed of all the traffic stream components (by vehicle type and by lane) and draw representative sets of vehicle trajectories on the (t, x) plane.

A recipe was given to solve dynamics problems in which the (t, x) data were piecewise constant on piecewise linear boundaries. For well posed problems the (t, x) solution is a unique partition of the solution space into polygonal areas of stationarity separated by "stable" interfaces. These interfaces represent (t, x) regions where traffic changes in character, and can be of three types: slips, simple (or kinematic) waves and mixed (or regime changing) waves. Slips are changes in vehicular density that propagate with the vehicular speed, i.e., where vehicles do not interact. Simple waves induce changes in speed as vehicles are hit by the wave, without lane-changing. Mixed waves involve concurrent speed, lane and psychology changes. Situations involving regime transitions such as incidents and their removal were

analyzed in detail. Of particular interest were the transitions marking the rear and front of a 1-pipe queue.

If the queue is generated when traffic flow is high then 1-pipe precursor states with a vehicular speed that can be anywhere between 0 and v_f will form upstream of it. These states grow spatially with time. The details depend on the particular conditions of the problem as per the geometrical construction in Fig. 6 of part I.

The evolution of the front of a 1-pipe queue is described by Fig. 4 of part I. The front of the queue is characterized by a “fan” of states with progressively higher speeds. In the prototypical case all vehicles leaving the queue first accelerate to speed v_f , entering the “capacity” 1-pipe state, and some time later rabbits change lanes and accelerate into a 2-pipe “discharge” state. The passing-lane flow of the discharge state, Q_d , is reproducible and smaller than the maximum possible flow on that lane, Q_c . The regime transition between the capacity and discharge states will travel either upstream or downstream, depending on the proportion of slugs in the queue, which dictates the shoulder lane flow. For multi-lane freeways the process is assumed to be similar; i.e., involving a sequence of discharge states where lanes reach their maximum speed sequentially from the shoulder to the median (see Fig. 8 of part I.)

The present paper extends the above to explain evidence “E” of part I. The following aspects of “E”, noted in Cassidy and Bertini (1996), are worth recalling: (E1) prior to the onset of queuing there is a period of time where the flows on both the passing lane and the freeway as a whole are very high; (E3): (i) delays for vehicles passing the merge increase slightly towards the end of this period; (ii) ... ; (iii) afterwards, the total flow drops sharply and queuing delay increases significantly; (iv) the total flow finally rises to a level, slightly below the original, which is sustained with minor fluctuations while the queue is present.

3. THE PUMPING EFFECT

Although extensions are possible, it will be assumed for the remainder of this paper that there are only two driver types (rabbits and slugs) and that slugs behave near the merge as they do in a homogeneous section, i.e. always staying in the right lane. We recognize that it would be more realistic to use more driver classes and to assume that some of the slugs could use the left lane in the neighborhood of the merge (i.e., temporarily behaving as rabbits), but generalizations like these would also introduce more parameters into the model. We prefer to keep the model simple until experimental evidence indicates that complications are necessary.

It is well known that the transition between the congested and uncongested traffic regimes at an active merge bottleneck occurs somewhere downstream of the merge rather than at the merge itself.¹ It is conjectured here that as main-line and on-ramp flows increase, main-line rabbits stay on the passing lane and become increasingly motivated to accept short headways, especially while proceeding past the merge. (Still triggered by passing, this psychological mechanism is the same as in part I.) At the same time, on-ramp vehicles of both types enter at the junction and stay on the shoulder lane. Further downstream, rabbits vacate the shoulder lane by accelerating and forcefully merging into the passing lane. These merges decrease the downstream headways on the passing lane, which is possible because drivers are motivated, and increase the downstream passing-lane flow beyond the discharge level. We call this the “pumping mechanism” and the resulting state an “over-saturated” or “pumped state”.

If merging and/or through traffic increase so much that the ensuing passing-lane flow would exceed the maximum possible motivated flow, Q_c , then only this flow would advance, and a queue would grow on the passing lane generating a semi-congested state. We say that the merge has been “critically-pumped”. Further increases in on-ramp flow and the ensuing decline

¹ See for example the theoretical proposal made in Buckley and Yagar(1974) in recognition of this phenomenon.

in passing-lane speed may cause the passing-lane queue to spillover to the shoulder lane and create a 1-pipe queued state. The result would then be an “over-pumped” state of affairs in the neighborhood of the merge. The remainder of this section only describes the pumping effect and the transition into criticality. The collapse into the over-pumped state is described in Sec. 4.

As in part I, it is assumed that rabbits (and slugs) are identical in all respects. In particular, all entering rabbits are assumed to accelerate and merge into the fast lane a fixed distance, D , downstream of the entrance and all slugs are assumed to remain in the shoulder lane. This is modeled, as shown in Fig. 1a, by introducing an imaginary set of lane markings (a baffle) that forbids shoulder lane vehicles from crossing into the passing lane between the entrance and the merge. Vehicle trajectories could then be idealized as shown in the figure. Note that the headway adjustments due to merging are confined to a narrow “awareness zone”. To keep the figure simple, the entering traffic stream includes no slugs. This, however, is not a requirement of the theory. The (t, x) diagram in part b of the figure is a continuum representation of this process in which the length of the awareness zone has been neglected and where capital letters (A, B, S) have been used to indicate the stationary states. The corresponding flow-density points are shown in the diagram on the left; q'_e is the entering rabbit flow.²

Traffic outside the buffer zone spanned by the baffle is assumed to behave as in part I. Buffer traffic is slightly different, however, because it is a superposition of ordinary upstream traffic and entering traffic, which must stay on the shoulder lane. As a result, the 2-pipe buffer regime can contain shoulder-lane rabbits. In recognition of this difference, “S” and later letters

² The notation in this paper is identical to that of part I.

of the alphabet will be reserved to denote buffer states. As shown on the flow-density diagram of Fig. 1b the buffer point denoting the total flow of rabbits may be off the curve. Note that flow conservation for rabbits requires $Q'_B = Q'_S = Q'_A + q'_e$,³ as shown on the flow-density diagram of Fig.1b.

Parts a and b of Fig. 2 show the transition into criticality. The (t, x) diagram of part a is a continuation of Fig. 1a, where the inflow of rabbits is assumed to have increased by 25% at $t = 1.5$. (Slug trajectories are not shown in this figure to avoid clutter.) The increased flow saturates the passing lane downstream of the merge and a simple shockwave denoting the back end of a passing-lane queue is kicked back upstream. The shock velocity and other macroscopic properties of this process are easily displayed with a continuum representation, as in part b of the figure. The (t, x) portion of this figure includes all the interfaces that would arise as a result of two successive increases in the flow of entering rabbits (at $t = -0.25$ and $t = 1.5$). The three interfaces on the left are slips, as are the interfaces separating S and S', and A' and B. The flow-density diagram shows how the shock velocity is determined. All the points in this figure, except the solid dot labeled "z" are data. This point is located on the down-sloping part of the Q(K) curve, with flow $Q_C - q'_e$, as shown. The shock velocity is obtained as usual. This shock is the only simple wave of the figure. State "Z" deserves to be called a critical state because the total flow of rabbits is Q_C . Of note: (i) U must be negative for otherwise we would still be in the case of Fig. 1; and (ii) the geometrical construction is not valid if solid dot "z" is below the ray with slope v_f . This would imply a traffic speed below v_f , which forbids a 2-pipe state. In this case the system would transition directly into a 1-pipe state. These types of transitions are described in Sec.4

³ Similar conservation relations hold for the flows of slugs and the total flows, but not for the lane flows.

Small increases (or decreases) in on-ramp flow while the system is in the critically-pumped state would continue to propagate as slips from the entrance to the merge. When each one of these slips reaches the merge a flow reducing (or increasing) wave with velocity W would be issued from the merge in order to maintain flow conservation at it without changing the downstream flow on the passing lane. The downstream state would remain “critical”. This wave may overtake the shock and alter its velocity, as in conventional kinematic wave theory. With successive (small) increases in on-ramp flow the solid dots representing the newly introduced semi-congested states upstream of the merge (e.g, points such as “z” in the figure) would move gradually down the $Q(K)$ line until eventually a new point would have to be below the ray with slope v_f . This is the moment of the collapse, which is depicted in Fig. 2c and examined below.

4. OVER-PUMPING AND THE ONSET OF QUEUING

When an increase in on-ramp flow such as the one issued from the entrance at time $t = 4$ in Fig. 2c reduces the passing-lane speed in the neighborhood of the merge to v_f or below, rabbits lose the incentive to follow closely and may want to move into the shoulder lane. When the reduction in speed spans several car lengths, some through rabbits will act, creating an ordinary 1-pipe state of unmotivated drivers. Because there is nothing holding the front of this queue, drivers in this collapsed 1-pipe state accelerate into the 1-pipe capacity state and later reach a discharge state. This process was described in connection with Fig. 4 of part I.

On the upstream side, given a velocity W' for cross-regime deceleration transitions (assumed to be independent of Z' as in the examples of part I), one can identify the collapsed 1-pipe state with the construction of Fig. 6b of part I. One may think of this state as a “queue precursor” state that introduces the capacity state.

The final result may be as in Fig. 2c. Note that the solution includes two regime transitions: (i) a “collapse wave” which involves a general speed reduction, lane-changes away from the passing lane and a loss of driver motivation, and (ii) a “frontal wave” or “acceleration transition” which marks the return of all the rabbits to the passing lane. The solution also includes two slips and a simple wave. Because the proportion of slugs in the traffic stream is low in our example the velocity of the frontal wave was positive but this does not have to be the case. The case with negative velocity is discussed latter.

Downstream behavior: Note that an observer well downstream of the merge would see a gradual increase in the freeway flows (states A, A' in our example), particularly in the passing lanes. Also note that maximum flow levels on the passing lane(s) (state B) could be reached and sustained for an indefinite period. This state of affairs, however, could also end (as occurred in the example) with a sudden drop in the overall flow and the introduction of a discharge state “D”. The drop in flow would be particularly severe on the passing lanes, although traffic on these lanes would continue to move faster than on the shoulder lanes.⁴ This state of affairs cannot persist, however. Eventually, the slow frontal wave would reach the observer and one would see a reduction in the speed of the passing lanes, a more even redistribution of flows across lanes and an increase in the overall flow. This state of affairs could then be sustained indefinitely. The qualitative description in this paragraph, which would also hold for the multilane freeway model embodied in Fig. 8 of part I, is consistent with puzzling empirical observations E1 and E3(vi).

Still downstream but closer to the entrance (upstream of our imaginary merge) the model predicts a slightly different pattern. Total flows also increase but the lane distribution is less

⁴ Flow on the right lanes could conceivably increase.

biased toward the passing lanes since entering rabbits have not yet had the time to merge. This is typified by states A, S and S' in our example. Eventually, once the passing lanes have been saturated with merging rabbits and maximal passing-lane flows are seen downstream, a speed reduction should also be noted on the passing lanes as a queue of slightly delayed passing vehicles begins to grow (state Z in the figure). In agreement with item E3(i), this state manifests itself by way of increased average trip times between detectors, particularly on the passing lanes. (Further increases in entering flow would result in even higher shoulder lane flows, lower passing lane flows/speeds and still unchanged passing lane flows well downstream.) Semi-congested states such as Z can propagate upstream of the merge and can last forever. However, an incident, or further increases in entering flow (to a level such as that of state Z' in part c the figure) would cause a collapse into the 1-pipe regime as shown.⁵ Our imaginary observer would then see the passage of the collapse wave which, in agreement with item E3(iii), introduces a sharp reduction in the speed of all lanes and an increase in vehicular delay. The precursor 1-pipe queue state so introduced would then be followed in short order by the capacity state and continued delays. (Because W' and w are likely to be similar and the length of the "buffer" is likely to be short, the precursor state should not last long at this location.)

Upstream behavior: Traffic behavior upstream of the entrance is influenced by the arrival of the collapse wave and the onset of the 1-pipe regime. From then on the merge should behave approximately as described in Daganzo (1994 and 1996), i.e., with the maximum flow allowed to enter being a given proportion of the flow downstream of the entrance. This proportion may be location- and flow-specific. Although the conventional wisdom is an entering proportion from the on-ramp equal to $1/2L$ (if L is the number of freeway lanes) a number closer

⁵ The model predicts that increases in upstream flow cannot cause the collapse.

to $1/L$ may be more reasonable if one finds that main-line vehicles make room for the on-ramp flow by moving to the passing lanes upstream of the entrance. In any case, a queue will grow on the on-ramp if its cumulative arrivals exceed the maximum allowable cumulative flow. Of course, a reduction in downstream flow will also reduce the upstream main-line flow. Congested main-line flows can be modeled with the KW wave model, albeit not precisely.⁶

Negative front velocity: It was shown in part I that the case of Fig. 2c with $w^* > 0$ arises if $Q_c > Q_d$, i.e., if there are few slugs. Otherwise, the velocity of the front would be negative and a pattern such as that in Fig. 3 should emerge; i.e., as in Fig. 4c of part I. The precise form of the pattern can be determined with the geometrical construction of said figure. The only difference is that the front velocity within the buffer, w_b^* , should now be determined as if the entering rabbits, who cannot change lanes within the buffer, were slugs. This leads to a discharge state with heavier flows in the shoulder lane, and flow Q_d on the passing lane. Insofar as $w_b^* < 0$, the total flow of the discharge state should now be (slightly) greater than capacity flow. On exiting the buffer at the dotted line of Fig. 3, the crypto-rabbits should force their way into the passing lane, increasing its flow to an over-saturated level (higher than Q_d). The figure assumes that this level does not exceed the critical level, and therefore that a queue is not created by this action. This should be the case in most instances. (Otherwise, and quite unlikely, a semi-congested state may appear upstream of the merge and downstream of the front.)

Therefore, in the case of Fig. 3, an observer well downstream of the merge would see

⁶ We recognize that oscillations exist in the 1-pipe regime and that these oscillations may grow in amplitude as one moves upstream from an active bottleneck. We suspect that the initial oscillations may be caused by random variations in the number of vehicles that make room for entering traffic by changing into the passing lanes forcefully. We also recognize that speed changes so generated will travel upstream as waves and that the changes may be magnified by further on-ramps and more congested traffic. Unfortunately, not enough detailed observations have yet been made to speculate further about this mechanism.

a less drastic reduction in flow than in Fig. 2 and would never see the capacity state. After a while, however, the observer may see a slight increase in flow, still in the 2-pipe state. The explanation for this is as follows: The arrival of the collapse wave at the entrance may create an on-ramp queue; if it does, the subsequent arrival of the frontal wave, with its reduced flows in the shoulder lane, would allow the queue to be released and more traffic to enter the freeway; this would be manifested upstream by a higher flow on the passing lanes $Q_d + q_e'$. If this were to happen, we would expect the final equilibrium to be an active bottleneck at the entry point with a 1-pipe freeway queue directly upstream of it, and a 2-pipe state downstream.⁷

5. CLOSING COMMENTS

The theory in this paper and its companion is promising because it is qualitatively consistent with all the observed phenomena listed at the beginning of part I. This is not enough, however. Since the model was developed so as to be consistent with *previously* known facts, it is only providing a “post-prediction”. The theory would pass a more meaningful test if it were to make some *new* predictions that turned out to be true. Some of these predictions have been mentioned in Sec. 4. Other investigators can make additional ones, perhaps aided with the detailed calculation procedure introduced in part I. New experimental evidence should then either put the theory on a firmer foundation or falsify it.

When comparing the theory with reality, however, it should be remembered that the theory does not apply to very light traffic, where drivers obey the rules of the road, and that the proposed mathematical models are highly idealized. In reality, for example, one would expect

⁷ We speculate that to maintain this state of affairs, through rabbits would merge into the passing lines immediately upstream of the merge in anticipation of the higher speeds, creating the bottleneck. The bottleneck will be stable at this location if it does not have a tendency to move upstream; i.e., if the speed of the wave front, w^* , would be positive upstream of the entrance. (This should be the case, given the scarcity of slugs—vehicles willing to be in the right lanes-- upstream of the on-ramp.)

waves (and in particular mixed waves involving lane-changes) to have a characteristic length comparable with tens of vehicle spacings, perhaps extending for the better part of a kilometer. If such a wave were to travel at a speed of 5 km/hr, it could take many minutes to pass over a fixed location. Thus, one should not expect to see the “clean” changes in flow predicted by the model. Furthermore, in reality, drivers will choose to merge at different locations, rather than at an idealized merge point. They will also choose to change lanes less predictably. Driver differences like these should result in less regular wave propagation than in the models; and one should expect the relative timing of the various events depicted in the (t, x) diagrams of the two papers to deviate slightly from the predictions in random ways. Interestingly, if these effects are factored in, one can readily see that the lane-specific flow-density scatter plots recorded by imaginary detectors in the examples of Figs. 2 and 3 would be qualitatively consistent with the patterns described under item “A” of part I.

Another prediction of the theory, which could have important practical ramifications if it is correct, is that prior to the collapse of a merge bottleneck two things must happen in sequence in the neighborhood of the on-ramp: (i) flows in the passing lanes must grow to be very high ($Q \approx Q_c$), and (ii) the speed in these lanes should decline slightly due to lane-changing mergers. Moreover, according to the theory these speed drops can be reversed by reducing the on-ramp flow. It therefore follows that one could prevent the collapse (while maintaining over-capacity flows) by metering the on-ramp so as to ensure that the speed on the passing lanes remains within an acceptable range.

Although the information given in these two papers should be sufficient for experimental falsification/verification, if the theory is found to be valid one should probably try to develop simplified recipes and computer models for systems involving more than one on-ramp. This should be relatively easy since all the waves in the proposed theory obey rather simple rules.

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 Prof. Michael Cassidy (U. California, Berkeley) for some background information.

REFERENCES

- Buckley, D.J. and Yagar, S. (1974) "Capacity funnels near on-ramps" in *Transportation and Traffic Theory*, Proc. 6th Int. Symp. Trans. Traffic Theory (D. J. Buckley, ed) pp. 105-123, A.H. and A.W. Reed, London, U.K.
- 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.
- Daganzo, C.F. (1994), "The cell-transmission model. Part II: Network traffic", PATH working paper UCB-ITS-PWP-94-12, Univ. of California, Berkeley, CA. Also in *Trans. Res.*, 29B(2), 79-94.
- Daganzo, C.F. (1996) "The nature of freeway gridlock and how to prevent it" in *Transportation and Traffic Theory*, Proc. 13th Int. Symp. Trans. Traffic Theory (J.B. Lesort, ed) pp. 629-646, Pergamon-Elsevier, Tarrytown, N.Y.
- Daganzo, C.F. (1999) "A behavioral theory of multi-lane traffic flow. Part I: Long homogeneous freeway sections" Institute of Transportation Studies Research Report, Univ. of California, Berkeley, CA.

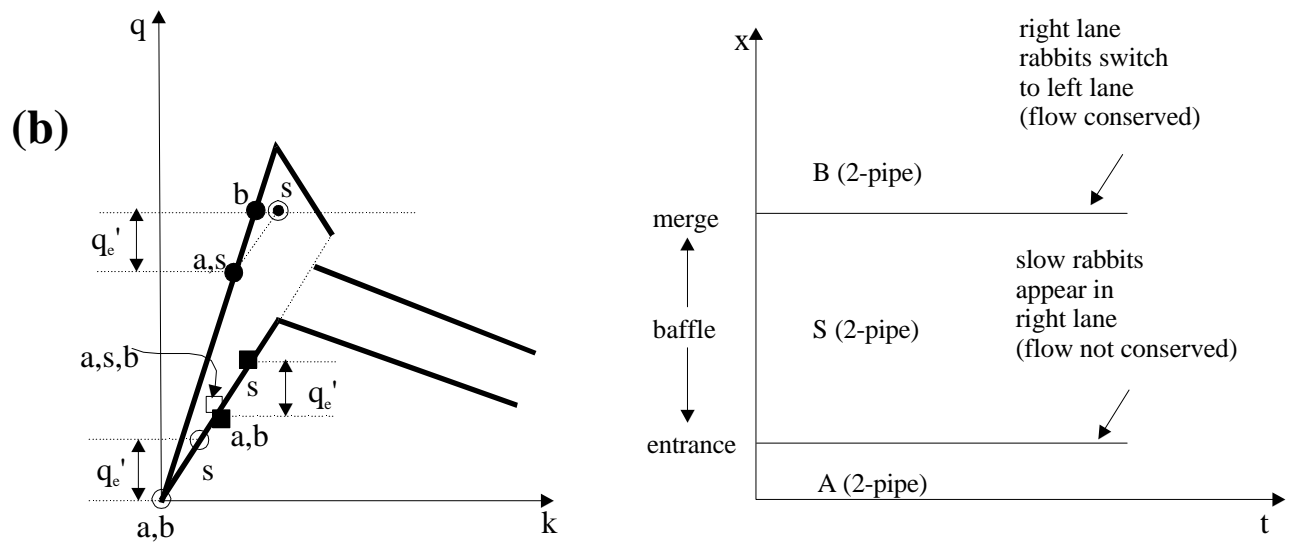
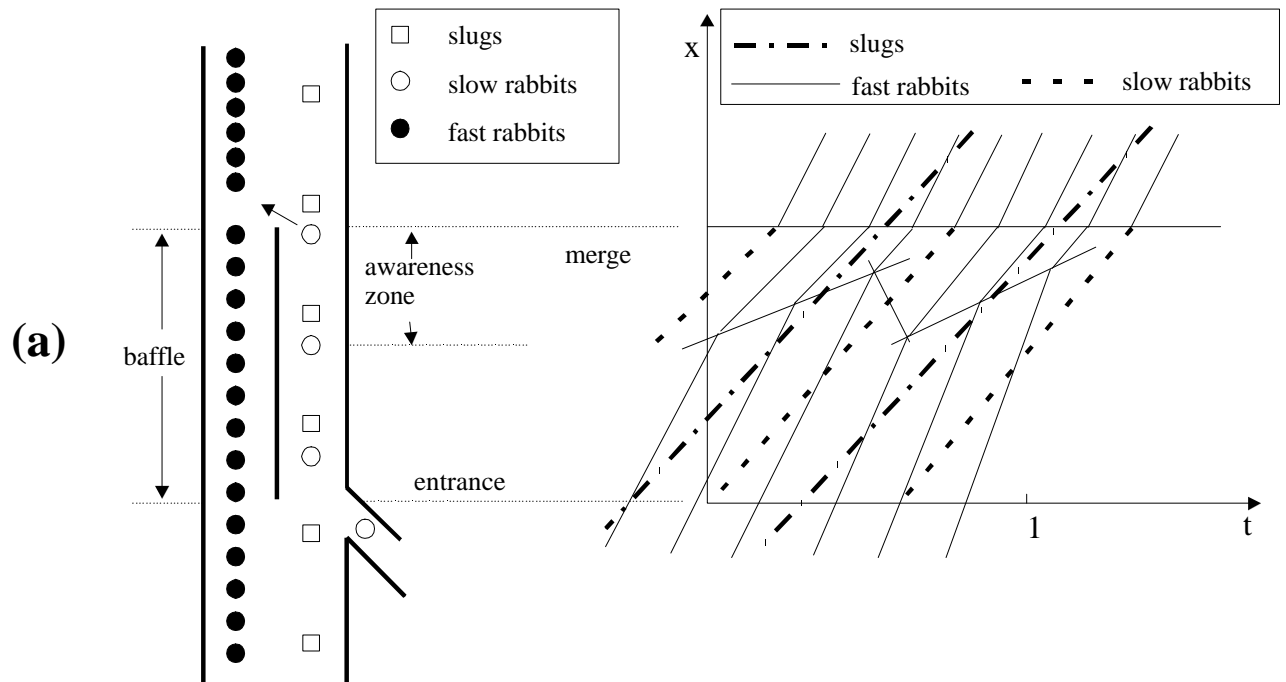


Figure 1. The pumping effect: (a) example vehicle trajectories in two-pipe regime in the vicinity of a merge; (b) schematic representation

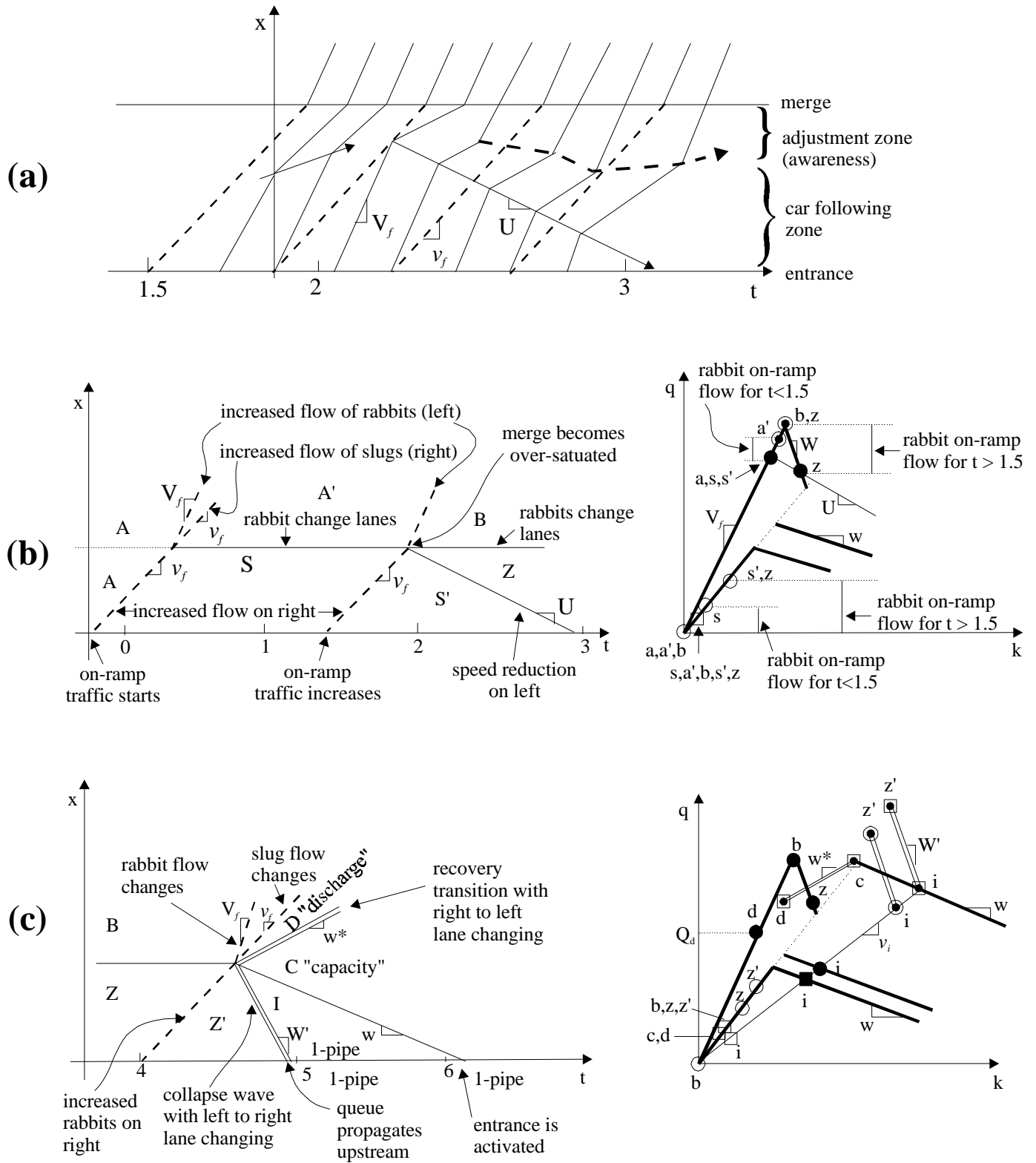


Figure 2. Effects of overpumping: (a) speed reduction on the left lanes with $V > v_f$; (b) schematic representation; (c) schematic representation for case with $V < v_f$, including "collapse" into 1-pipe regime when the velocity of the downstream recovery front is positive.

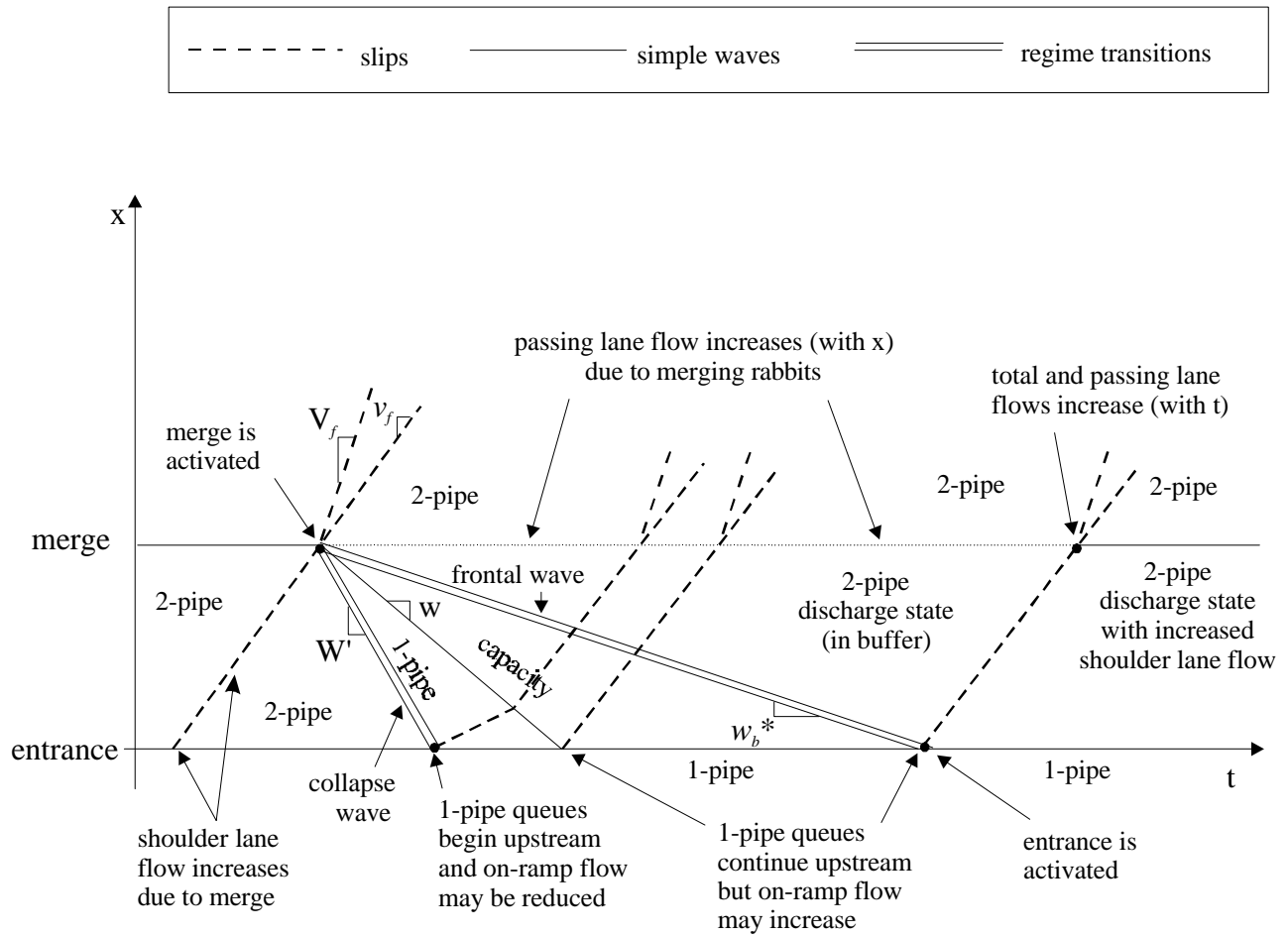


Figure 3. Example of system evolution when the velocity of the downstream recovery front is negative.