Remarks on Traffic Flow Modeling and its Applications

Carlos F. Daganzo

Department of Civil and Environmental Engineering
University of California, Berkeley, CA 94720, USA
daganzo@ce.berkeley.edu

1. This document presents some recent results and ideas from the University of California (Berkeley) traffic operations group, and at the same time discusses the role of traffic flow modeling in traffic management and control. It stresses the steps that can be taken to reduce congestion and improve traffic efficiency, and how traffic models and theories fit within this picture.

1 Introduction

2. Few drivers would dispute the fact that congestion is caused by bottlenecks, both recurrent and non-recurrent, and that the resulting queues can cause further problems if they become too long. Long queues can entrap cars that do not wish to pass through the bottleneck that generated them, compounding the problem and causing spillovers. These can have widespread effects, such as "gridlock".

3. It is therefore important to learn more about the behavior of bottlenecks and the spatial extent of queues. The game in congestion management is queue avoidance and containment. This can be achieved through a combination of active control measures such as ramp metering and also by means of passive measures such as route guidance and information delivery.

4. From a practical point of view, it is most important to have models that can predict reliably the things that matter; i.e. bottleneck behavior and queue dynamics. Models should be tested by verifying how well they can predict these performance measures (the generation of queues and their spatial evolution), more so than other things.

5. The following text will discuss two important types of bottlenecks (merges and diverges), queue storage issues and the behavior of systems with interconnected bottlenecks. Theoretical and experimental issues will be addressed.

2 Active bottlenecks

6. We say that traffic is unqueued ("uncongested" or "free flowing") if small speed disturbances introduced at a location are not detected upstream of that location. Experimental methods have been developed for determining if this is the case. [1] Conversely, if changes in speed introduced at a location are felt later at an upstream location we say that traffic is queued (or congested).

7. We say that an active bottleneck exists in between two locations if traffic is detected to be queued upstream of the location and unqueued downstream. The identification of bottleneck activity at unexpected locations (non-recurrent congestion) has shown promise for incident detection. [2]

3 Merges

8. A simple theory for merge bottlenecks states that there is a maximum ***sustainable***

flow downstream of a merge when downstream conditions are uncongested, and this quantity is called the “capacity”. According to this theory, if the sum of the entering flows over an extended period of time exceeds the merge capacity, then the downstream flow will drop to the capacity level and a queue will form. The queue may grow on either one or on both approaches to the merge, as explained under item 11.

9. Experiment shows that in some locations the period of time with over-capacity flows lasts for five or ten minutes and at other locations almost nothing at all. Experiments also show that the queue discharge flow (the capacity) is relatively stable. This is most clearly seen from an examination of cumulative count curves. [3]

10. Whether over-capacity flows at merges (if they arise) can be maintained for longer periods of time by doing something to the traffic stream upstream of the merge has not been established. Proper experiments need to control for conditions downstream, which must be uncongested for the duration of the experiment.

11. Although experiments have not been performed to test the following hypothesis (apparently proposed by K.Moskowitz in the late 1950's) it seems rather plausible. One would expect traffic from the 2 approaches to a merge to flow in a fixed reproducible ratio if the bottleneck is active and both approaches are queued. Moreover, if one of the approaches is not queued one would expect it to have less flow than its share, and the other approach to take up the slack.

12. These rules are sufficient to formulate a model that would predict what happens to two traffic streams that compete for space in a merge of insufficient capacity. The model would predict approximately when queues would form and the delay that individual vehicles would experience in passing through the bottleneck. It would not predict the spatial extent of the queues, however. [4]

4 Diverges

13. Diverges, such as the one shown in Fig. 1, are other common locations for active bottlenecks. Here the simplest theory consists of assuming that each off-branch of the diverge has a “capacity” and that flows higher than the capacity cannot be sustained.

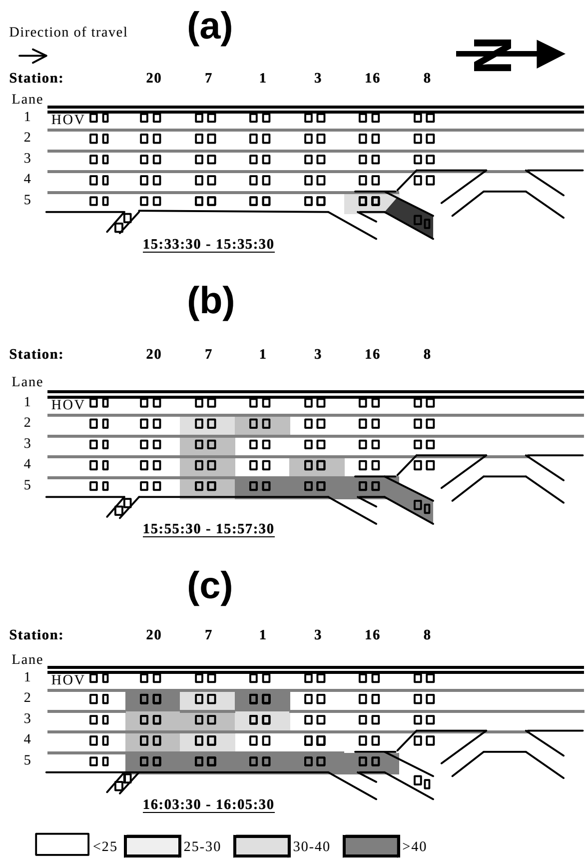

FIG. 1. Queue Evolution at a Diverge Bottleneck:
Shading intensities indicate occupancy rates
Source: Daganzo, C., Cassidy, M., Bertini, R. (1998) *Possible Explanations of Phase Transitions in Highway Traffic* , to be published in Transportation Research.

14. If the flow that passes either branch of the diverge exceeds its capacity for a while then a queue should grow in the common approach. If the approach is narrow this queue should operate with a first-in-first-out (FIFO) discipline, so that it will entrap and delay

vehicles destined for the other branch of the diverge. Something qualitatively similar may happen even if the approach is wide, as in the case of Fig. 1, but not enough experimentation has been carried out at different locations to determine if this happens in general.

15. In any case, once a queue has grown upstream of the diverge the combined flow past the bottleneck should depend on the mixture of exiting vehicles present in the queue, which should vary. Thus, ***flows upstream and downstream of queued diverges should not be expected to be steady***, although they could be.

16. If one knows the desired (virtual) exit flows by destination as a function of time, one can use these bottleneck rules and the FIFO approximation to predict when queues grow upstream of a diverge and the ensuing vehicle delays; see item 17. [5]

17. A simple geometrical construction with cumulative curves, that can and has been programmed, can be used to illustrate the procedure (see Fig. 2). The example in the figure depicts a curve (labeled "TOTAL") that corresponds to a steady arrival stream. A queue is created in this stream because the percentage of exiting cars varies. The evolution of delay until dissipation is shown, assuming that each branch of the diverge has a well defined capacity and that the queue discipline if FIFO; i.e., that the horizontal separation between the three pairs of arrival/departure curves is the same for vehicles that arrive at the same time.

18. Experiments reveal that mainline freeway flows can collapse in this way next to an off-ramp (even if the mainline flow is steady) with no evidence of congestion downstream of the diverge on either branch. [6]

5 Other bottlenecks

19. Merges and diverges are not the only bottleneck locations. Bottleneck activity can also be detected at weaves, sags, curves, tunnels, lane drops and other locations where the freeway characteristics change. Bottleneck activity can also be generated by temporary exogenous causes such as an incident, some unusual activity next to the road, or even a distracting variable message sign.

20. In my opinion, it is important to improve our understanding of bottleneck behavior more than almost anything else in traffic theory. Especially since bottlenecks come in many different flavors, and some can be rather peculiar.

21. As an illustration of this, and perhaps somewhat surprisingly, I should note that there are situations where imposing a speed limit on a freeway bottleneck may actually increase its "capacity". A typical example occurs in a 2-lane uphill section of California State Highway 17 out of Los Gatos, California, where fast cars, which saturate the passing lane at about 110 Km/hr, avoid the shoulder lane because this lane is traveled by just a few but much slower trucks (at 80 Km/hr.) The result is a markedly underutilized shoulder lane, which creates a bottleneck at the beginning of the

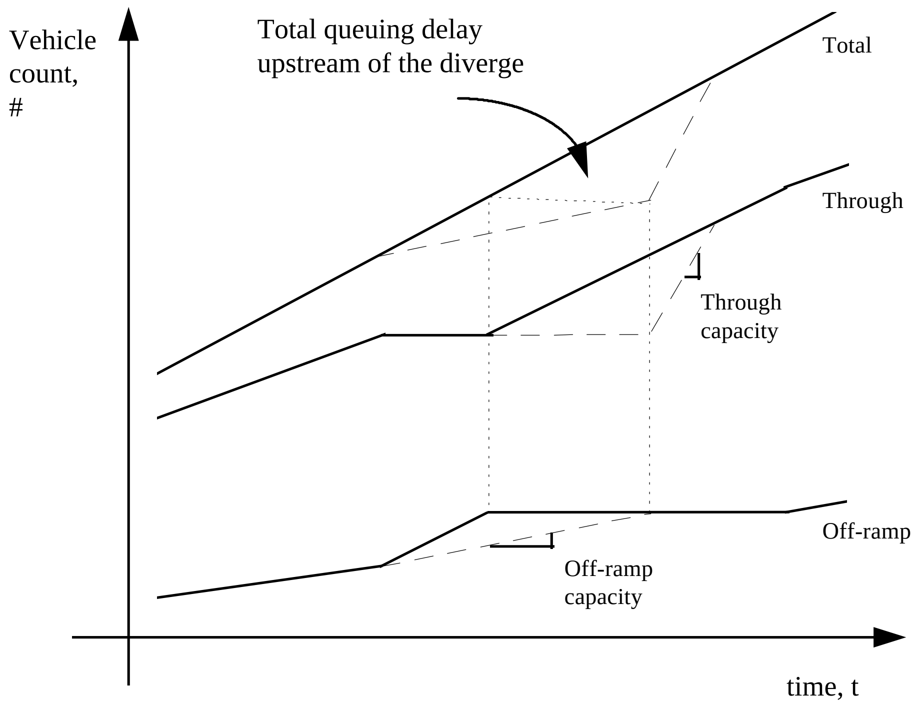

Figure 2. Queue formation upstream of a diverge due to a fluctuation in traffic composition

grade. A speed limit could encourage cars to use the right lane, resulting in higher flows through the bottleneck. It could perhaps eliminate the very long waits observed on the approach to the hill.

22. The above remarks have shown how queues may be triggered. (Information of this sort is very valuable because it can be used by traffic management schemes in an attempt to avoid queues.) Furthermore, in all the theories reviewed vehicle delays only depend on two things: (1) on people's "schedules", i.e. on the times when individual vehicles would have passed the bottleneck in the absence of queuing, and (2) on the behavior of the bottleneck itself. Nothing else matters. In particular, ***the delay does not depend on the structure of the queues***. This suggests that control actions should be directed at the bottlenecks more than at the queues themselves.

23. This is not to say that the spatial extent of the queues does not matter. If a queue is allowed to spill back past an intersection it may interfere with other ***upstream*** traffic and create another bottleneck. This is usually undesirable.

24. **Gridlock effect**: Figure 3a shows an active merge bottleneck and a freeway queue that spills back past a diverge, starving the exit for flow. According to the foregoing theories if we were to restrict the on-ramp flow, the freeway would take up the

slack and the downstream flow would not change. If the ***upstream*** freeway flow could be increased (by metering the on-ramp in this way) beyond the demand level, then the queue would abate. It would no longer block the off-ramp and restrict its flow, as shown in Fig. 3b.   As a result the total system flow would have increased. This illustrates the importance of queue containment and the need for theories that will predict the distance spanned by queues. [7]

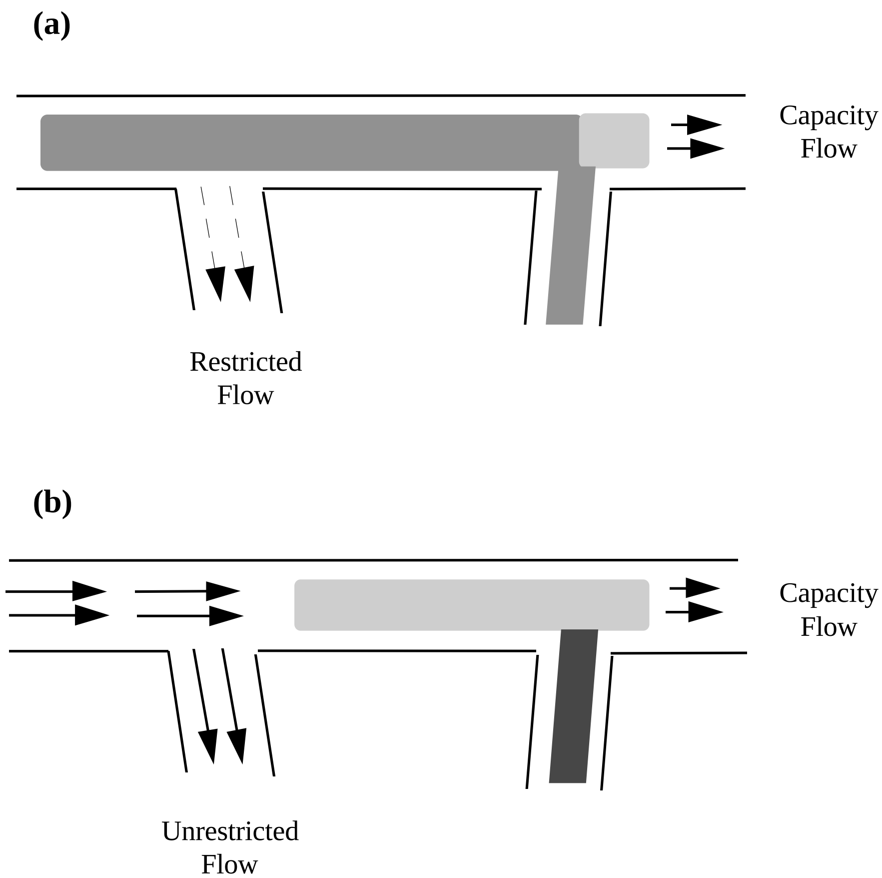

Figure 3. Effect of spillbacks on exit flow and the effect of ramp metering: (a) No Metering; (b) Metering: More System Flow

6 Storage theories

25. The simplest storage theory would say that the amount of distance per unit vehicle consumed in a long queue (the average spacing) is only a function of the queue discharge rate. (Note that this is not very different from saying that there is a "fundamental diagram" between flow and density.)

26. Evidence in this respect is very sparse. Nonetheless, it appears from one experiment [8] (see Fig. 4) that the number of vehicles in between two distinct stationary observers fluctuated within reasonable norms despite the occurrence of stop-and-go episodes, and that this average number increased with declining discharge flows. Thus, hypothesis “25" may have some practical value, even if it is not correct at a microscopic level.

27. Hypothesis “25" implies the existence of “waves”. If carried to its microscopic limit it becomes the kinematic wave (KW) theory. These waves are manifested by linked changes in slope of the curves of vehicle arrival, as suggested in Fig. 5.[9] In the figure, a downstream change in slope is detected upstream later, which is what one would expect in congested traffic.

28. Over short distances, waves are noticeable and appear to behave reproducibly. [10] Thus there is hope that this simple model of traffic storage, or a simple modification of it, can predict vehicle accumulations and queue distances reasonably well. In my view, the use of more complicated models for freeway traffic prediction and management seems a bit premature, since it is not clear yet that there is a need.

29. Macroscopic traffic models that include the above rules for bottlenecks and are consistent with this simple storage theory have been developed. A computer demonstration that illustrates the lasting and widespread effects of an incident downstream of a diverge, and where to some drivers it would appear that the congestion they experienced was caused by nothing, can be seen by visiting the author’s home-page:

“www.ce.berkeley.edu/~daganzo/”.

A computer program for Windows PC’s and its user manual can also be downloaded from this website. [11]

7 Spillovers, information and system capacity

30. Spillovers are rather peculiar and they must be treated with caution. Situations exist where provision of information or even the improvement of a bottleneck will actually reduce the flow that can go from point A to point B.

31. That information can have a negative effect is clear. If we inform drivers on a freeway close to saturation that “there is congestion ahead”, some may decide to exit

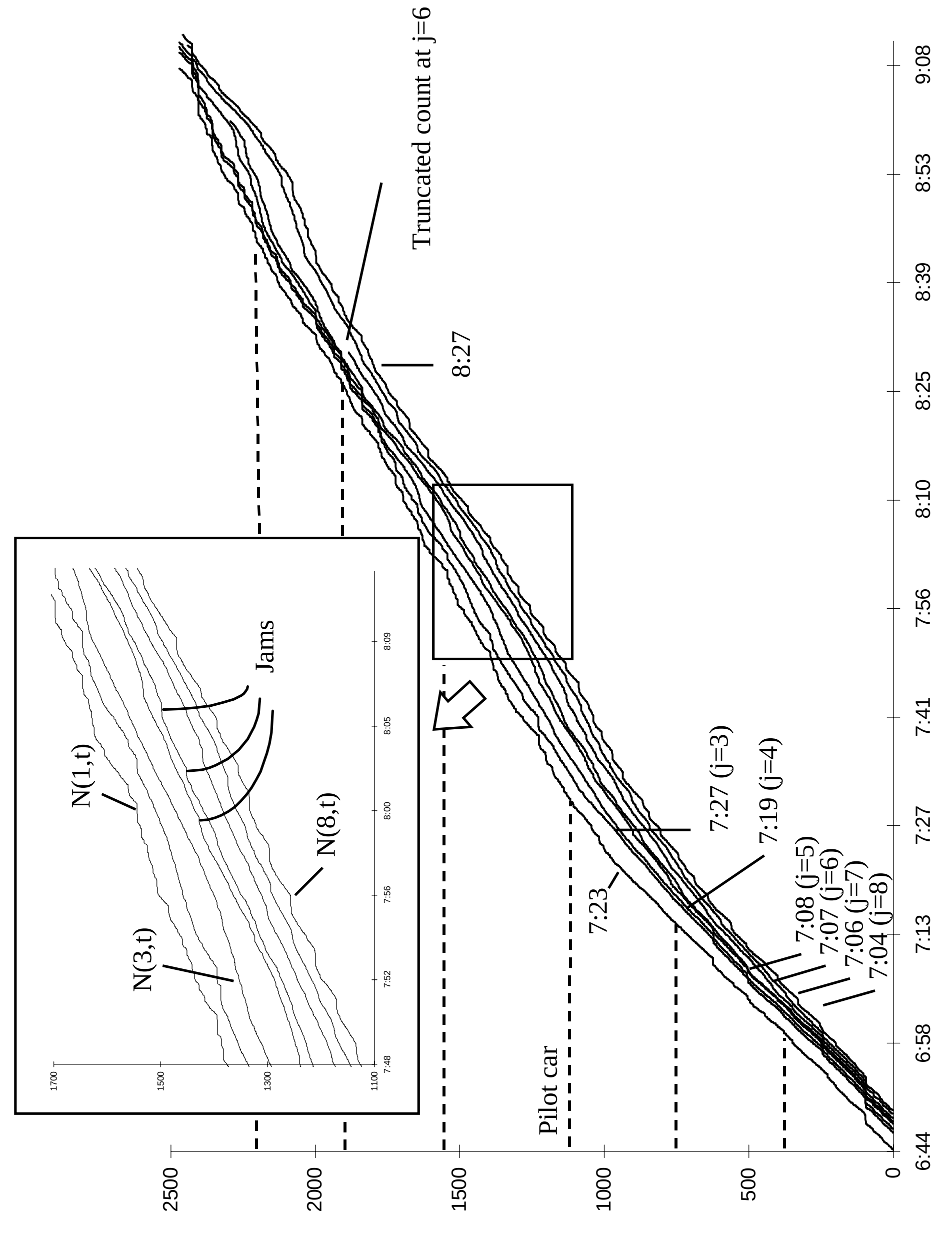

Figure 4: Filtered Curves of N(j,t) for Day Two (Source: Smilowitz, K., Daganzo, C., Cassidy, M., Bertini, R. (1998) *Some Observations of Highway Traffic in Long Queues*, Research Report, Institute of Transportation Studies, University of California at Berkeley, USA.)

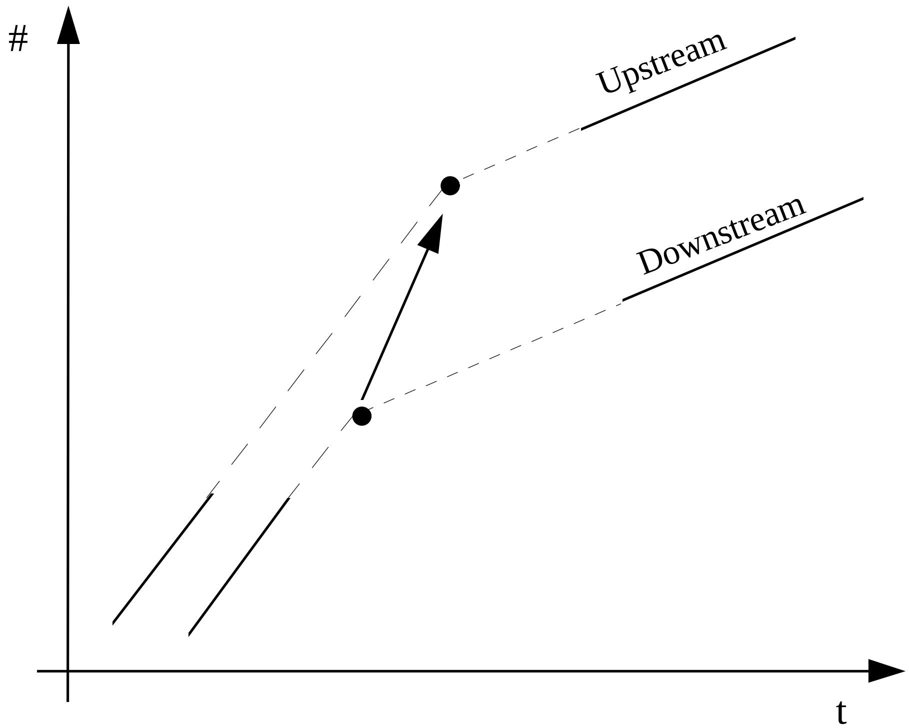

Figure 5. Reproducibility of vehicle accumulations implies a kinematic wave

the freeway. If the exit was already close to saturation, the added flow may push the exit "over the edge," converting it into a bottleneck that would spill onto the freeway. In the process we would have created a worse problem than the original one.

32. Improving an active bottleneck (expanding its capacity) can also be undesirable. This may happen if the bottleneck is located in a preferred road between two points A and B (e.g. a freeway) and the congestion it generates diverts some traffic to less desirable but uncongested alternative routes (e.g., to surface streets). With an improved capacity, the bottleneck should attract enough flow from the alternative routes to equalize approximately the trip times on the preferred and the alternative routes. If the trip time on the uncongested alternative routes is insensitive to decreases in flow, then the delay at the bottleneck should not change either after its capacity is expanded. Thus, its queue will grow... ***longer*** ! The longer queue can spill over upstream intersections. It is not difficult to think of many situations where such a spillover can have dire consequences. [12]

8 Multiple stable states and chaos

33. In fact, it is possible to construct very simple networks with steady origin to destination flows, as in Fig. 6, that can be stable in two ways: (a) with an over-saturated configuration where some of the queues have backed up to the origins, blocking their flow, and (b) with an under-saturated configuration of queues that allow all the origins to discharge their flows.

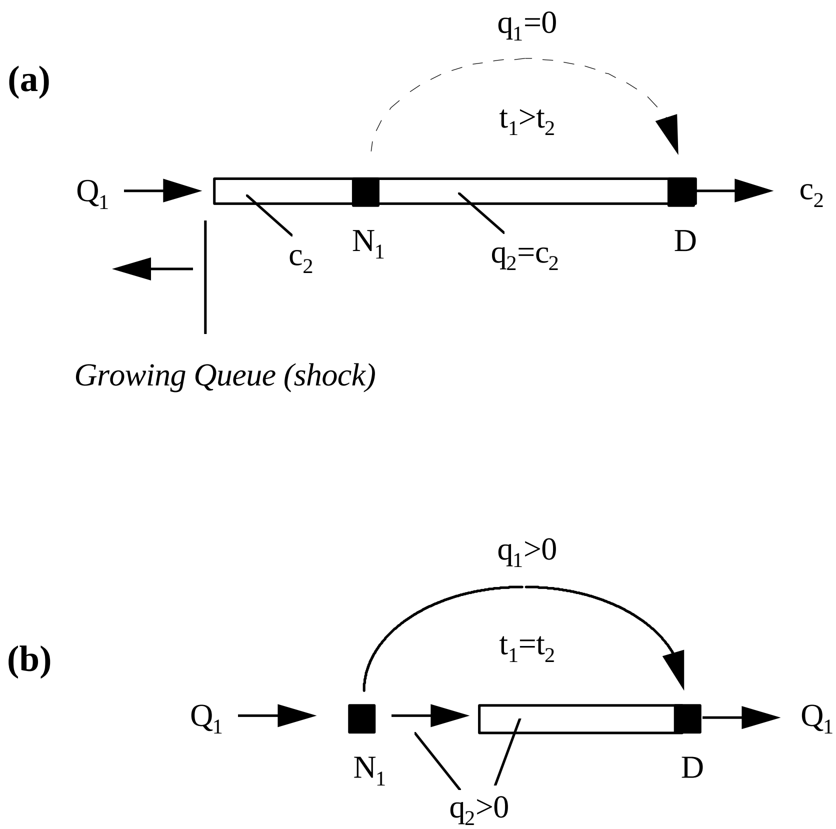

Figure 6. A simple network that can be in two different steady states:
(a) Oversaturated; (b) Undersaturated

34. Furthermore, in these types of networks a ***temporary*** disturbance can change ***permanently*** the saturation state of the network. When conditions are changing with time (as occurs in real life), this means that a small localized and temporary variation in conditions can be magnified and spread through the network. The resulting changes, may in turn, trigger others as in a "chain reaction". This suggests that the behavior of traffic in congested networks with interconnected bottlenecks and spillovers is chaotic in nature. This fact, combined with our present inability to predict accurately some basic inputs, such as how people choose routes on a time-dependent network, should force us to rethink the role of predictive models in traffic engineering and management.

9 Role of modeling

35. In view of the above, it seems that models cannot be expected to provide precise (or even reasonable) predictions on a link by link basis for large congested networks. Therefore, it seems rather fruitless to base system control procedures (for congested systems) on model predictions.

36. Nonetheless, models can perhaps play some role in an evaluation process. If one decides for example, that traffic into the central part of a city should be metered (so as to prevent crowding), it may be reasonable to use a ***simple and reasonable*** traffic model to predict the ***overall*** effects of the strategy in the city center. Of course, those predictions would still have to be taken with a "grain of salt".

37. Simulations are most useful for small networks (or subnetworks such as weaving areas, interchanges and small freeway sections) without route choice [13] where one can hope to obtain the necessary information to make predictions, and where problems of sensitivity of the results to the initial conditions do not arise. In instances like these, methods of dynamic traffic management (e.g., access control and ramp meters) based on models become possible. It is my opinion that in any particular application one should always choose the simplest model that can predict what is at stake. [14]

38. I believe that we stand to gain the most from focusing more attention on two basic questions. Bottleneck behavior is the first one. It is my opinion that too much effort is spent by our research community today developing models of traffic behavior ***for homogeneous roads***, and that not enough experimental work is being done to see how things really are. Homogenous roads are important, but we should really bring the focus of our lenses to the things that matter most. And these are the bottlenecks.

39. The spatial extent of queues would be the second issue. Experimentalists today measure things such as "occupancy", "speed" and "flow" and use these quantities to assess the goodness of models. I would much prefer to measure the things that the models should really be predicting (e.g.., the "queue distances" and "vehicle trip times" between detectors) and then use these to choose among models.

40. Of course, there are many other issues deserving attention, such as an improved understanding of driver route choice behavior but, in my opinion, the above two research directions are so fundamental that they should be explored first.

Acknowledgments

41. Much of this work has been supported by PATH MOU-305. The artwork was done by K. Smilowitz.

Endnotes

[1] One possible approach that seems to be reliable is explained in: "*Methodology for assessing dynamics of freeway traffic flow*, (M. Cassidy and J. Windover), Trans. Res. Rec. **1484**, 73-79, 1995."

[2] See "*Incident detection with data from loop surveillance systems: The role of wave analysis.* (W.H. Lin) PhD thesis, Dept of Civil and Environmental Engineering, University of California, Berkeley, 1995." This reference also discusses the state of the art in incident detection.

[3] Some interesting figures can be found in "*Some traffic features at freeway bottlenecks.* (M. Cassidy and R. Bertini) Institute of Transportation Studies Research Report ITS-RR-97-07, U. of California, Berkeley, CA, 1997; Trans. Res. A. (in press)," and also in not yet published work by M. Mauch.

[4] A detailed explanation of the theory can be found in Sec. 2 of "*The nature of freeway gridlock and how to prevent it*, (C. Daganzo) in *Transportation and Traffic Theory*, pp. 629-646, J.B. Lesort, editor, Pergamon-Elsevier, New York, N.Y., 1996." This reference also describes the implications of the theory for ramp metering.

[5] A detailed explanation of the theory, including its discretization for computer calculation, can be found in Secs. 3.3 and 4 of "*The cell transmission model part II: network traffic*. (C. Daganzo) Trans. Res. **29B**, 79-93, 1995."

[6] This phenomenon has been reported in "*Experimental properties of phase transitions in traffic flow*. (B. Kerner and H. Rehborn) Phys. Rev. Let. **79,** 4030-4033, 1997."

[7] If queue spillovers fill a ring road with a queue, a gridlock process is started where traffic conditions gradually worsen. See "*The nature of freeway gridlock* (op. cit.)"

[8] The data for this experiment have been posted on the World Wide Web at:

"http://www.ce.berkeley.edu/Programs/Transportation/Daganzo/spdr.html"

[9] The connection between these changes in slope and kinematic waves was noted in "*A simplified theory of kinematic waves in highway traffic, I general theory, II queuing at freeway bottlenecks, III multi-destination flows* (G.F. Newell) Trans. Res. **27B**, 281-313, 1993".

[10] See "*Empirical studies of the dynamic features of freeway traffic.* (J.R. Windover) PhD Dissertation, Dept. of Civil and Environmental Engineering, University of California, Berkeley, 1998" for an extensive statistical analysis of these waves.

[11] See "*The NETCELL simulation package: technical description* (R. Cayford, W. Lin, and C. Daganzo) California PATH research report UCB-ITS-PRR-97-23, University of California, Berkeley, 1997."

[12] This idea and its disturbing logical consequences are explored in: "*Effect of queue spillovers on transportation networks with a route choice*, (C. Daganzo) Institute of Transportation Studies, Research Report, UCB-ITS-RR-96-1, Univ. of California, Berkeley, CA, 1996; Trans. Sci. in press."

[13] The dangers of controlling a system in which drivers can change routes by simple feedback mechanisms are well known; see for example: "*Traffic control and route choice; a simple example*. (M.J. Smith) Trans. Res. **13B**, 289-295, 1979."

[14] A recent review of the ramp metering literature can be found in "*Traffic control on metered networks without route choice* (D.J. Lovell) PhD Dissertation, Dept. of Civil and Environmental Engineering, University of California, Berkeley, 1997"