



0191-2615(93)E0002-3

THE CELL TRANSMISSION MODEL: A DYNAMIC REPRESENTATION OF HIGHWAY TRAFFIC CONSISTENT WITH THE HYDRODYNAMIC THEORY

CARLOS F. DAGANZO

Department of Civil Engineering and Institute of Transportation Studies,
University of California, Berkeley CA 94720, U.S.A.

(Received 23 October 1992; in revised form 13 July 1993)

Abstract—This paper presents a simple representation of traffic on a highway with a single entrance and exit. The representation can be used to predict traffic's evolution over time and space, including transient phenomena such as the building, propagation, and dissipation of queues. The easy-to-solve difference equations used to predict traffic's evolution are shown to be the discrete analog of the differential equations arising from a special case of the hydrodynamic model of traffic flow. The proposed method automatically generates appropriate changes in density at locations where the hydrodynamic theory would call for a shockwave; *i.e.*, a jump in density such as those typically seen at the end of every queue. The complex side calculations required by classical methods to keep track of shockwaves are thus eliminated. The paper also shows how the equations can mimic the real-life development of stop-and-go traffic within moving queues.

1. INTRODUCTION

Accurate descriptions of highway traffic flow over transportation networks, whether at the planning or operations level, must recognize that the vehicles traveling on any section of the network must be bound for specific destinations.

Static traffic assignment models used for transportation planning (see Sheffi, 1985, for example) achieve this goal by describing the flow on a link of the network by its components by final destination; *e.g.*, by specifying a variable y_{id} that represents the amount of flow on link i that is ultimately bound for destination d . Unfortunately, this is much more difficult to do for dynamic network flow problems (with time-dependent origin-destination (O-D) flows) because the functional dependence of the link flows at time t , $y_{id}(t)$, on the collection of all past flows is quite complex. This problem manifests itself both at the planning level, where networks are quite complex, and at the operations level, where networks are simpler, but more detail is sought about the system's evolution.

Although dynamic traffic assignment models—planning level models involving large networks—typically recognize that traffic travels to many destinations, the models are based on simplistic flow relationships that are not perfectly consistent with the conservation laws of traffic. A planned sequel to this paper will discuss this in more detail.

Traffic operations models can be microscopic or macroscopic. Microscopic simulations (*e.g.*, Schwerdtfeger, 1984; Cremer and Ludwig, 1986; Nagel and Schreckenberg, 1992) assume that the behavior of an individual vehicle is a function of the traffic conditions in its environment. Although microscopic simulations usually keep track of each vehicle's destination, their assumptions are difficult to validate because humans' behavior in real traffic (not in contrived "car-following" experiments) is difficult to observe and measure. This is unfortunate because for a simulation to work the microscopic details have to be just right. (As is well known from nonlinear system theory, microscopic details have a way of affecting the macroscopic world unpredictably.) In this paper, thus, we focus on macroscopic models.

Macroscopic models assume that the aggregate behavior of sets of vehicles, easier to observe and validate, depends on the traffic conditions in their environment. The hydrodynamic theory of traffic flow (Lighthill and Whitham, 1955; Richards, 1956)

underlies most of these models. While not perfect, its shortcomings are well understood. It is justifiably and routinely used in traffic engineering analyses. Unfortunately, most macroscopic models do not differentiate their component flows by destination; each stream is treated as the flow y_i of a single commodity. Thus, when the flow reaches a fork in the road, or some other junction where there is a route choice, the models usually specify either a fixed proportion of turns or a fixed exit flow. In reality, though, neither is fixed; both depend on when vehicles bound for the various destinations reach the junction in question.

This deficiency was addressed by Vaughan, Hurdle, and Hauer (1984), Vaughan and Hurdle (1992), and Hurdle (1992), who formulated the problem for a simple network consisting of a series of links connected by junctions. Unfortunately, the solution to the differential equations resulting from the hydrodynamic model of traffic flow, using the classical method of characteristics, is tedious even for a single link, let alone a series of links. In an attempt to solve the multiple link problem, Newell (1993) proposed a shortcut solution method for the hydrodynamic model of one link. The new method predicts the variation of traffic flow at one end of the link from the behavior of traffic at the other end, *without evaluating the behavior at intermediate points*. (Based on the cumulative flow curves at the entrance and exit to the link, the method had also been proposed by Luke (1972) for the solution of a geological erosion problem.) This ability to “jump” from one end of the link to the other is exploited by Newell (1991) to solve the freeway traffic flow prediction problem formulated by Vaughan, Hurdle, and Hauer.

This paper presents an alternative way of predicting traffic behavior for one link by evaluating flow at a finite number of *carefully selected* intermediate points, including the entrance and exit. As in Newell (1993), the difference equations that form the basis for this procedure are shown to be discrete approximations to the differential equations of the hydrodynamic theory for a special form of the equation of state. Although it requires more computer memory than Newell’s method, the proposed procedure can be readily extended to complex networks such as those involving loops and diverging branches; this will be shown in a planned sequel.

The paper is organized as follows: After introducing the basic model (in Section 2), it is shown to be a discrete equivalent of the classical hydrodynamic model: first when the traffic density is continuous in time and space (Section 3), and then in the presence of discontinuities (Section 4). Sections 5 and 6 present and evaluate two simple generalizations aimed at (a) representing more general equations of state (Section 5), and (2) capturing instability phenomena (Section 6). Section 7 concludes the paper with some remarks of importance for network extensions.

2. THE CELL-TRANSMISSION MODEL

This paper examines the evolution of traffic over a one-way road without any intermediate entrances or exits, so that vehicles enter at one end and leave at the other. We propose to simulate the system with a “time-scan” strategy where current conditions are updated with every tick of a clock.

To this end the road is divided into homogeneous sections (*cells*), numbered consecutively starting with the upstream end of the road, from $i = 1$ to I . The lengths of the sections are not chosen arbitrarily; they are set equal to the distances traveled in light traffic by a typical vehicle in one clock tick. Under light traffic then, all the vehicles in a cell can be assumed to advance to the next with each tick; it is unnecessary to know where within the cell they are located. Thus, the system’s evolution obeys:

$$n_{i+1}(t + 1) = n_i(t),$$

where $n_i(t)$ is the number of vehicles in cell i at time t . We will assume that the above recursion holds for all flows, unless traffic is slowed down by queuing from a downstream bottleneck. This seems reasonable because, for crowded conditions as might arise during the rush hour, most of the delays can be attributed to queuing at bottlenecks where flow

temporarily exceeds capacity, rather than to any dependence between flow and speed. (It has been empirically observed that the space-mean speed of freeway traffic remains relatively constant, independent of flow, unless flow is close to capacity.)

To incorporate queuing we introduce two constants: $N_i(t)$, the maximum number of vehicles that can be present in cell i at time t , and $Q_i(t)$, the maximum number of vehicles that can flow into cell i when the clock advances from t to $t + 1$ (time interval t). The first constant is the product of the cell's length and its "jam density," and the second one is the minimum of the "capacity flows" of cells $i - 1$ and i . It will be called the "capacity" of cell i —short for input capacity—with the understanding that it represents the maximum flow that can be transferred from $i - 1$ to i . We allow these constants to vary with time to be able to model transient traffic incidents. The number of vehicles that can flow from cell $i - 1$ to cell i when the clock advances from t to $t + 1$ (the flow into " i " for the time interval after " t "), $y_i(t)$, is assumed to be the smallest of three quantities:

$$\begin{array}{ll} n_{i-1}(t) & , \quad \text{the number of vehicles in cell } i - 1 \text{ at time } t, \\ Q_i(t) & , \quad \text{the capacity flow into } i \text{ for time interval } t, \text{ and} \\ N_i(t) - n_i(t) & , \quad \text{the amount of empty space in cell } i \text{ at time } t. \end{array}$$

(This last quantity ensures that the vehicular density on every section of the road remains below jam density.) Our proposed simulation, called the "cell-transmission model," will be based on a recursion where the cell occupancy at time $t + 1$ equals its occupancy at time t , plus the inflow and minus the outflow; *i.e.*,

$$n_i(t + 1) = n_i(t) + y_i(t) - y_{i+1}(t) \quad (1a)$$

where the flows are related to the current conditions at time t as indicated below:

$$y_i(t) = \min\{n_{i-1}(t), Q_i(t), N_i(t) - n_i(t)\}. \quad (1b)$$

The simulation would step through time, updating the cell occupancies (for all i) with each tick of the clock.

Boundary conditions can be specified by means of input and output cells. The output cell, a sink for all exiting traffic, should have infinite size ($N_{I+1} = \infty$) and a suitable, possibly time-varying, capacity. Input flows can be modeled by a cell pair. A "source" cell numbered "00" with an infinite number of vehicles ($n_{00}(0) = \infty$) that discharges into an empty "gate" cell "0" of infinite size, $N_0(t) = \infty$. The inflow capacity $Q_0(t)$ of the gate cell is set equal to the desired link input flow for time interval $t + 1$. The gate cell then acts as a metering device that releases traffic at the desired rate while holding (as a parking lot would) any flow that is unable to enter the link.

Note that *the result of the simulation is independent of the order in which the cells are considered at each step*. This important property of the cell-transmission model will permit the analysis of complex networks, *e.g.*, with loops. The property arises because we are specifying that the number of vehicles that enter a cell is unrelated to the number of vehicles that will leave it; thus, only current conditions influence the inflow to a cell. Although the first two restrictions ($y_i(t) \leq n_{i-1}(t)$ and $y_i(t) \leq Q_i(t)$) are reasonable in this regard, one may argue that if a cell is rapidly being emptied then the third (occupancy) restriction $y_i(t) \leq N_i(t) - n_i(t)$ may be overly conservative. The counter-argument is that empty slots for cars can only travel backwards at a finite speed (the wave propagation speed) unlikely to be greater than the free flow speed. Therefore, the effects of the outflow should only be noticed upstream after some time. Because this lag is one tick of the clock for our model, this is akin to assuming that density waves propagate backwards at the free flow speed. Sections 3 and 4 demonstrate that our simulation indeed behaves as a hydrodynamic model of traffic flow with this property. An extension in which backward waves move more slowly is introduced in Section 5.

3. EQUIVALENCE TO A HYDRODYNAMIC MODEL

We now consider a homogeneous highway and show that eqns (1) are a discrete approximation to the Lighthill, Whitham, Richards hydrodynamic model with a density-flow (k - q) relationship in the shape of an isosceles trapezoid, as in Fig. 1. This relationship can be expressed as:

$$q = \min\{vk, q_{\max}, v(k_j - k)\}, \quad \text{for } 0 \leq k \leq k_j, \quad (2)$$

where v is the free-flow speed (and the speed of all backward moving waves), k_j is the jam density, and $q_{\max} \leq k_j v/2$ is the maximum flow.

If eqn (2) is replaced into the flow conservation equation:

$$\partial q(x,t)/\partial x = -\partial k(x,t)/\partial t,$$

we obtain the differential equation that would define the evolution of the system under the hydrodynamic model:

$$\partial \min\{vk(x,t), q_{\max}, v(k_j - k(x,t))\}/\partial x = -\partial k(x,t)/\partial t. \quad (3)$$

We now show that eqn (3) is equivalent to our simulation. Note that for our homogeneous highway, the cell characteristics would be independent of i and t : $N_i(t) = N$ and $Q_i(t) = Q$. To demonstrate the equivalence of the discrete and continuous approaches, we now define the tick of the clock to be equal to dt and choose the unit of distance such that $vd t = 1$. Then the cell length is 1, v is also 1, and the following equivalences hold: $x \equiv i$, $k_j \equiv N$, $q_{\max} \equiv Q$, and $k(x,t) \equiv n_i(t)$. With these conventions, the variable in braces in eqn (3) is equivalent to:

$$\min\{n_i(t), Q, N - n_i(t)\},$$

which coincides with the definition of $y_{i+1}(t)$ of eqn (1b), except for the subindex of n in the last term which should have been $i + 1$ instead of i . This, however, is immaterial unless the density is discontinuous. Because the hydrodynamic differential equations are only meaningful when the density is differentiable in x (discontinuities—shocks—are treated outside of the differential equations with the hydrodynamic model), we can state that the variable in braces is equivalent to $y_{i+1}(t)$, and as a result the left hand side of eqn (3) is:

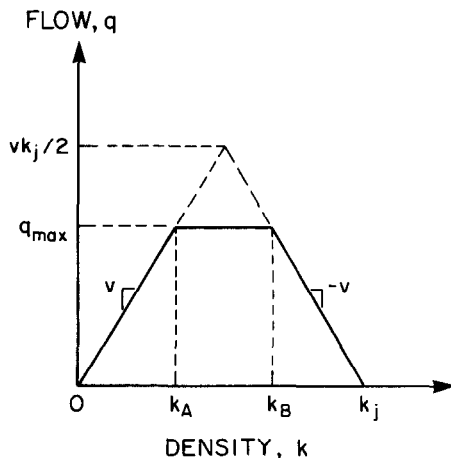


Fig. 1. Flow-density relationship for the basic cell-transmission model.

$$y_{i+1}(t) - y_i(t).$$

The right side of eqn (3), of course, is:

$$- [n_i(t+1) - n_i(t)],$$

and the equality of these two quantities justifies the simulation recursion, eqn (1a).

To be sure, there are other forms of eqn (1b) that would also be consistent with eqn (3). For example, the flow into cell i , $y_i(t)$ could have been specified only as a function of the occupancy of the sending cell, $i-1$, so that eqn (1b) would have been:

$$y_i(t) = \min\{n_{i-1}(t), Q_i(t), N_{i-1}(t) - n_{i-1}(t)\}.$$

Alternatively, $y_i(t)$ could have been specified only as a function of the occupancy in the receiving cell. While these, and possibly other specifications are consistent with eqn (3) for any slow varying density, their behavior in the face of discontinuities is not equivalent and may not be reasonable.

That consistency with eqn (3) is no guarantee of reasonable behavior across discontinuities is clearly illustrated by the expression above. It should be clear from inspection that such an expression would predict zero flow downstream of any discontinuous drop in density from the jam density to any lower value. Thus, such a model would predict that vehicles queued at a traffic light would not advance even after the light turned green! (Unrealistic predictions also arise if flows are based only on the status of the receiving cell, as then a receiving cell would continue to attract vehicles even after upstream occupancies have dropped to zero.)

Because discontinuities are common—arising spontaneously when low density traffic catches up with denser, slower downstream traffic—their effect on model performance must be examined. We claim here that eqn (1b) will replicate the behavior of the continuous model even across discontinuities. Thus, it should be possible to iterate eqn (1) and automatically track varying densities and the paths of any resulting shocks. This property, very useful for automatic computation, is discussed in the next two sections.

4. PROPAGATION OF DISTURBANCES AND CREATION OF SHOCKS

Here we compare the predictions of the hydrodynamic theory and our model, when the density along the road is known at an initial time ($t = 0$) and discontinuities may be created.

Consider a portion of road where the density varies within a narrow range in the direction of travel; *i.e.*, $k(x,0)$ changes with x , but remains either (a) between 0 and k_A (in Fig. 1), (b) between k_A and k_B , or (c) between k_B and k_j . Then the hydrodynamic theory predicts that the same density profile will be preserved over time, except for a position shift; no shocks are created. That is, $k(x,t) = k(x - w_k t, 0)$, where w_k is the wave speed for the portion of the diagram corresponding to our initial densities. For our chosen time-distance units of measurement (with $v = 1$) the wave speed is 1 for case (a), 0 for case (b), and -1 for case (c). (This result arises because in the special case we are considering the characteristics—loci of constant density in time-space—are parallel straight lines with slope w_k .)

It is not difficult to see that the cell-transmission model also satisfies $k(x,t) = k(x - w_k t, 0)$: According to eqn (1b) the number of vehicles transmitted from cell $i-1$ to cell i is either n_{i-1} , Q_i , or $N_i - n_i$, depending on whether k is in $[0, k_A]$, $[k_A, k_B]$ or $[k_B, k_j]$; and as a result, it should be clear from eqn (1a) that $n_i(t+1)$ will be either $n_{i-1}(t)$, $n_i(t)$, or $n_{i+1}(t)$, depending on whether k is in $[0, k_A]$, $[k_A, k_B]$ or $[k_B, k_j]$. Thus, the density profile (vehicle counts) move with the wave speed.

If the initial density profile is not entirely within a narrow range, the evolution is more complicated—discontinuities are created. We consider first situations in which the density increases only in the direction of travel.

4.1. Increasing density

Let us assume that $k(x,0)$ increases from a point slightly below k_A of Fig. 1 to a point above, and examine a range of x where the increase is roughly linear.

Without loss of generality, we assume that the units of measurement for time and space are chosen so that $v = 1$ (as before) and $\partial k(x,0)/\partial x = 1$. Further, we assume that vehicles are counted in units (*e.g.*, pairs, halves, dozens . . .) such that the maximum flow is $q_{\max} = 50$ units/time, and also assume that the origin of (spatial) coordinates is located where the initial density is 50 count units per unit distance. Thus, without any loss of generality we have reduced all of the problems of interest to a unique one for which the diagram of Fig. 1 has only one free parameter k_j (since $q_{\max} = 50$, $k_A = 50$, and $k_B = k_j - k_A = k_j - 50$), and such that the initial density profile is $k(x,0) = 50 + x$. Furthermore, because neither k_B nor k_j affect the evolution of our density profile as long as the densities remain below k_B , the free parameter doesn't need to be specified (we could select for example $k_B = 100$ and $k_j = 150$).

Figure 2a depicts the map of characteristics for this problem and the shockwave

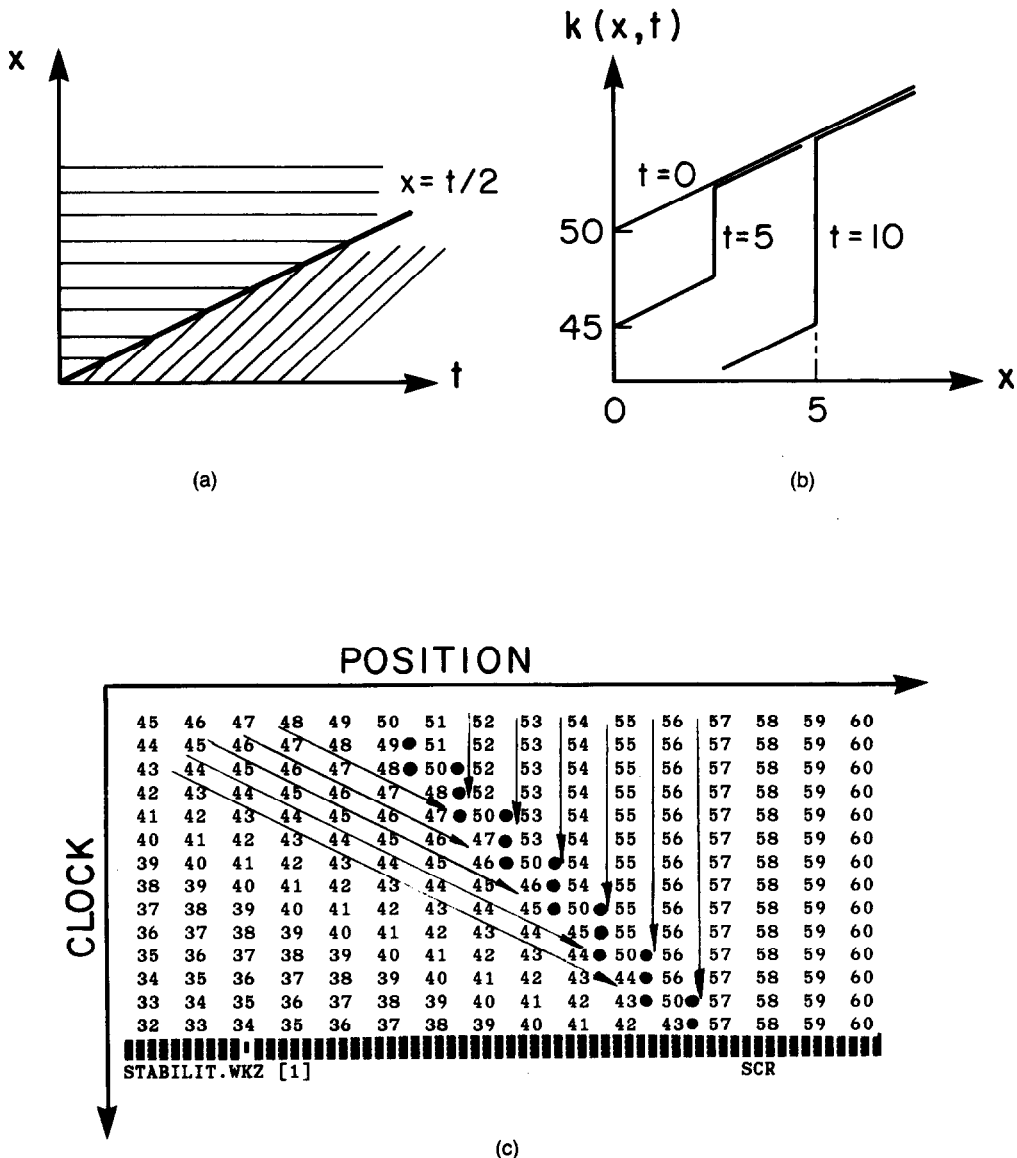


Fig. 2. Traffic evolution with an increasing density. (a) Map of characteristics. (b) Density profile. (c) Numerical result.

resulting from the convergence of the characteristics. The shock path is defined by the line: $x = t/2$. On one side of the shock, the density is defined by the horizontal characteristics and is $k(x,t) = 50 + x$; the flow is $q(x,t) = 50$. On the other side the density, defined by the slanted characteristics, is $50 + x - t$; the flow is $q(x,t) = k(x,t)$. As the reader can easily verify, the curve in the figure is the only $x(t)$ satisfying the vehicle conservation condition:

$$\partial x(t)/\partial t = [q_1(x,t) - q_2(x,t)]/[k_1(x,t) - k_2(x,t)], \quad (4)$$

where the subscripts "1" and "2" refer to the state on either side of the shock. (This condition ensures that vehicles don't disappear; *i.e.*, those entering the shock on one side also leave it on the other side.)

Figure 2b depicts the density profile $k(x,t)$ for different t . Note how the shock moves forward at speed $1/2$, gradually increasing in size.

Figure 2c depicts the result of the cell transmission model, eqn (1), when the initial conditions are as stipulated and the clock tick has been chosen to equal one time unit — this also implies that cells are one distance unit long. Note the similarity of Figs. 2c (viewed sideways) and 2a. The arrows in Fig. 2c follow the characteristics of constant density. The shock path, highlighted by circles between cells, alternates between 1 and 2 cells wide; as in Fig. 2a, it advances 1 cell (distance unit) for every 2 clock ticks. The rows of Fig. 2c are also consistent with the evolution pattern depicted in Fig. 2b.

The close agreement is not a result of the clock speed used. If a clock that was 10 times faster had been chosen, cells would have been 10 times shorter and the entries in the first row of Fig. 2c would have been 10 times smaller. Because Q and N would also have been 10 times smaller, the result of the recursion would reveal the exact same pattern of Fig. 2c, except that the entries would have been 10 times smaller. Note in particular that the shock path would have remained 1 or 2 cells wide so that its actual spatial range would have been 10 times smaller than before. Thus, in agreement with the continuous hydrodynamic model, the spatial dimension of the shock could be made to vanish with infinitely fast clocks and negligible cell sizes.

A density profile $k(x,0)$ that increases past k_B yields similar results. The same set of parameters for Fig. 1 (with $k_B = 100$ and $k_j = 150$) also suffices to describe this case exhaustively as long as the upstream density remains above k_A . As the reader can verify, the map of characteristics is now simply an upside down version of Fig. 2a, with backward-moving characteristics ahead of stationary ones and a shockpath defined by $x = -t/2$. The numerical results of a simulation with first row cell contents increasing past "100" (. . . 98, 99, 100, 101, . . .) are also consistent with this pattern; as a whole they resemble an upside down version of Fig. 2c, including a backward moving shock that is 1 or 2 cells wide.

Suppose now that the density profile $k(x,0)$ increases past k_A and then past k_B . Then, the forward-moving shock generated when the density increases beyond k_A will eventually be met by the backward moving shock generated by the growth past k_B . The shocks would coalesce into a (forward or backward) moving shock that would separate upstream traffic states with density in the interval $[0, k_A]$ from downstream traffic states with density in $[k_B, k_j]$. The speed of the coalesced shock would be governed by eqn (4).

The numerical model replicates the behavior of the coalesced shock with a 1- or 2-cell interface that moves at the correct speed. That the interface must move at the speed given by eqn (4) should be obvious, because eqn (4) is a direct result of vehicle conservation, and the numerical model conserves vehicles.¹ This fact is illustrated by the example in Section 5.

¹To see that the interface spans only 1 or 2 cells, start with an initial set of occupancies such as (. . . , n' , n' , n_0 , n'' , n'' , . . .) where n' would correspond to a density below k_A , n'' to a density above k_B , and n_0 would be in the interval $[n', n'']$. Then, it suffices to check that at the next iteration there would be at most one cell with an occupancy different from n' and n'' , and that cell could only be the cell previously containing n_0 , or one of its neighbors. Care must be exercised in checking this property, because the flows between cells depend on the magnitude of n_0 relative to n' , Q , $N-Q$, n'' , and N .

4.2. Decreasing density

This subsection shows that the cell transmission model is not disrupted by discontinuities in the derivative of the density profile, which are generated by a decreasing density.

With the flexibility to choose the units of measurement for counts, time, and distance as described in Section 4.1, the diagram of Fig. 1 with $q_{\max} = 50$, $k_A = 50$, $k_B = 100$, and $k_j = 150$ still suffices to describe exhaustively the evolution of any density profile decreasing smoothly past either k_A or k_B . The origin of spatial coordinates is chosen so that $k(0,0) = 50$ if the initial density profile decreases past k_A , and $k(0,0) = 100$ if it decreases past k_B .

With these conventions, the maps of characteristics for both cases are as depicted in Fig. 3a. Note the constant-density, wedge-shaped regions that appear in the time-space continuum. Because the characteristics diverge, no shocks are formed.

The figure also displays the density profile predicted by the hydrodynamic theory at $t = 0$ and 1 for both cases. In the first case, densities below k_A are propagated forward at speed 1, and higher densities remain stationary; an expanding road section with density k_A separates the high and low densities. In the second case, the low density section remains stationary, and the high density section is propagated backward at speed -1 .

Figure 3 also displays the results of the cell-transmission model, using as before a one time unit clock tick. The results closely replicate the hydrodynamic predictions. Note in particular the wedge-shaped regions of constant count and the direction of the characteristics.

We have thus established that if the density changes gradually over space, the cell transmission model is equivalent to the continuous hydrodynamic model. Discontinuities in this density (shocks) are also captured adequately by the cell-transmission model; they are represented by transition sections comparable with the lattice width and spanning 1 or 2 cells.

An example involving the build-up and dissipation of a queue, and confirming the phenomena just discussed, is given in Daganzo (1993). An appendix in that reference also provides a spreadsheet that can be used to solve similar problems with the cell transmission model.

5. THE CELL-TRANSMISSION MODEL: GENERAL CASE

Although the equation of state in Fig. 1 allows one to choose three basic engineering parameters (the free flow speed, the maximum flow, and the jam density), the relationship forces the backward wave speed to match the free flow speed. This is somewhat unrealistic because in reality waves move several times more slowly than free flowing traffic, changing the manner in which vehicles approach the bottleneck and the location of queues. In turn, this may change the times when an intersection upstream of a bottleneck might be blocked by the queue because, with slow waves, queues persist for a longer time behind temporary bottlenecks and are dissipated further upstream. Without close intersections upstream, the wave speed discrepancy is less important—it can be shown not to change the time when approaching vehicles would pass the bottleneck and, hence, not to influence the resulting vehicle delay.

Here, thus, we examine an extension of the cell-transmission model that would approximate the hydrodynamic model for an equation of state that allows backward waves with speed $w \leq v$. Said equation of state, depicted in Fig. 4, is:

$$q = \min\{vk, q_{\max}, w(k_j - k)\}, \quad \text{for } 0 \leq k \leq k_j, \quad (5)$$

where $w \leq v$ and $q_{\max} \leq k_j/[1/v + 1/w]$.

The cell-transmission model intended to represent this relationship is identical in all respects to the one in Section 2, except that eqn (1b) is modified slightly and now is:

$$y_i(t) = \min\{n_{i-1}(t), Q_i(t), (w/v)[N_i(t) - n_i(t)]\}. \quad (6)$$

5.1. Equivalence to the hydrodynamic theory

For a homogeneous highway, the differential equation defining the system evolution under the hydrodynamic model, formerly given by eqn (3), is now:

$$\partial \min\{vk(x,t), q_{\max}, w(k_j - k(x,t))\} / \partial x = -\partial k(x,t) / \partial t. \quad (7)$$

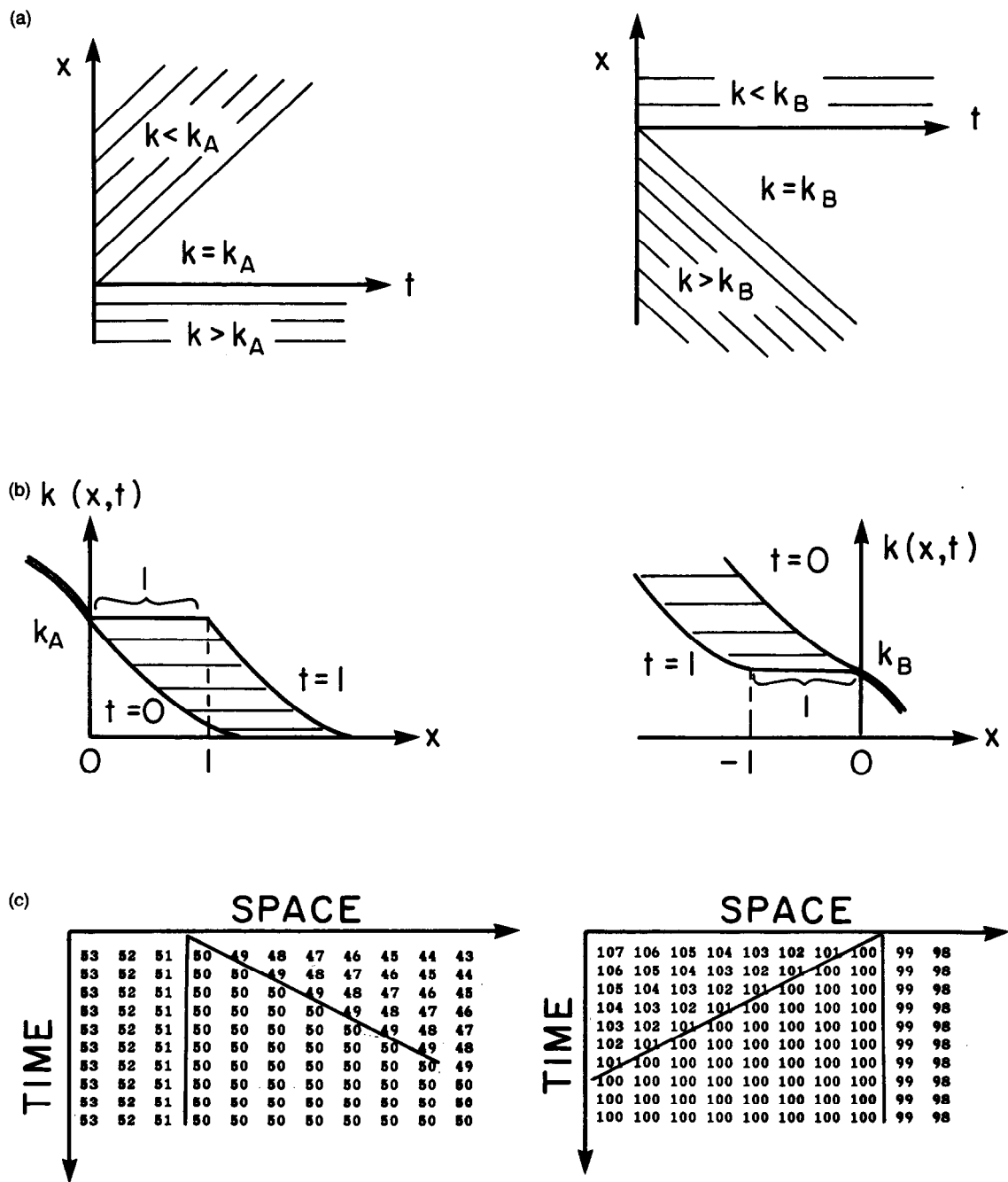


Fig. 3. Traffic evolution with a decreasing density. (a) Map of characteristics. (b) Density profile. (c) Numerical result.

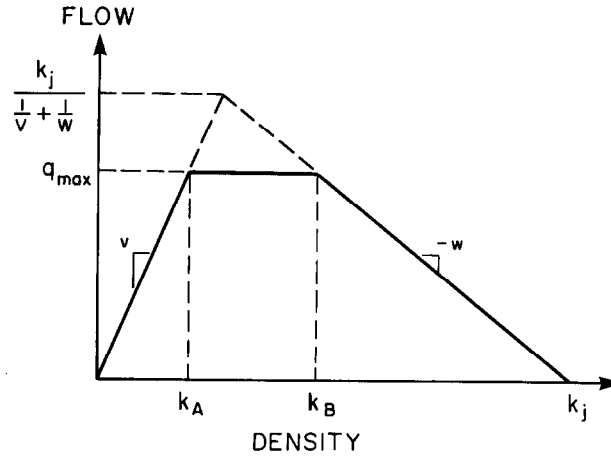


Fig. 4. Flow-density relationship for the generalized cell-transmission model.

It is not difficult to see repeating the arguments of Section 3 (explicitly given in Daganzo, 1993) that eqns (6) and (7) are equivalent wherever $k(x,t)$ is differentiable.

As in Section 3, other subindex choices for eqn (6) could also preserve the equivalence, but the ones adopted here ensure that the recursion behaves properly in the presence of discontinuities. The following subsection explores this fact in more detail.

5.2. Behavior of difference equations with finite clock ticks

As happened for the basic cell-transmission model, finite difference eqns (1a) and (6) solve the continuous hydrodynamic model of Fig. 4 when an infinitesimally small clock tick is used. We now explore the model behavior with finite clock ticks and discontinuous densities. Section 4 demonstrated that for finite clock ticks the difference between the continuous and discrete solutions to the basic model is minimal: the largest discrepancy between the two arises with the representation of shocks, which take no space in the continuous model but span either 1 or 2 cells in the discrete model. The generalized model, although equivalent to the continuous model in the limit, is not as well behaved for finite clock ticks.

To illustrate the issues, the current subsection describes the evolution of both a shock and an acceleration wave as predicted by both the continuous and discrete equations. Then, the following subsection introduces a modification to eqn (6) that improves the finite clock tick performance of the model.

In this subsection, we use the diagram of Fig. 4 with $v = 1$ and $k_A = 50$ (no loss of generality here) and also assume when an illustration is needed that $w = 0.25$ (a reasonable value) and $k_B = k_A$. As a result, $q_{\max} = 50$ and $k_j = 250$.

Consider first a density profile, representing traffic running into the back of a stationary queue at position $x = 15$; i.e.:

$$\begin{aligned} k(x,0) &= k, & \text{if } x < 15 \\ &= k_j, & \text{if } x \geq 15. \end{aligned}$$

The discontinuity at $x = 15$ is a shockwave that will propagate backward. If k is greater than k_B (e.g., $k = 70$), then the shock will propagate at speed -0.25 (see Fig. 4).

In a plot of cumulative vehicle count (recall that we may be counting vehicle pairs, dozens, etc.) vs. distance, the shock will appear as a convex break in the slope of the curve. The position of this break is independent of the count label assigned to the vehicle at $x = 0$, and for this reason we will always assign label 0 to the vehicle immediately upstream of $x = 0$; then the cumulative count $K(x,t)$, called from now on the "cumulative profile," is the integral over x of $k(x,t)$. (Note that this scheme does not identify individual vehicles.) For our example, the initial cumulative profile is:

$$\begin{aligned}
 K(x,0) &= xk & , & \quad \text{if } x < 15 \\
 &= 15k + (x - 15)k_j, & \quad \text{if } x \geq 15.
 \end{aligned}$$

Because the shock travels at speed 0.25, the break in slope is at $x = 15 - 0.25t$ for any $t > 0$. Hence, for any $t < 60$, the cumulative profile predicted by the hydrodynamic theory is:

$$\begin{aligned}
 K(x,t) &= xk & , & \quad \text{if } x < 15 - 0.25t \\
 &= (15 - 0.25t)k + [x - (15 - 0.25t)]k_j, & \quad \text{if } x \geq 15 - 0.25t.
 \end{aligned}$$

The cumulative profile can also be evaluated with the cell-transmission model. If the clock ticks once every time unit and cells are defined so that cell 1 extends from $x = 0$ to $x = 1$, then the cumulative profile at any (integer valued) x is simply the sum of the vehicle counts in cells 1 to $x - 1$. For the case with $k_j = 250$ and $k = 70$, the initial cell contents would then be: $(\dots, 70, 70, 250, 250, \dots)$. Figure 5 displays the approximation to $K(x,t)$ at various times; dashed lines show the exact result. Notice how as time passes the discrete approximation “softens” the shock, so that it spreads to a growing number of cells.

A similar spreading phenomenon occurs with acceleration waves (*i.e.*, concave bends in the cumulative profiles). Daganzo (1993) contains an illustration.

We claim that the prediction errors caused by numerical spreading are minor because as illustrated in Fig. 5: (a) the maximum error in the count increases at a decreasing rate with the passage of time, (b) the disturbances still travel at the appropriate speed, (c) the vehicle counts and densities on either side of the shock (or wave) are not corrupted, and (d) the accuracy of the count near a shock (or wave) can be controlled by modifying the clock tick—as per dimensional arguments already seen.

5.3. A modification that eliminates spreading for certain shocks

Repetition of the arguments in Section 3 shows that the following modification to eqn (6) preserves its consistency with eqn (7) wherever $k(x,t)$ is differentiable:

$$y_i(t) = \min\{n_{i-1}(t), Q_i(t), \alpha[N_i(t) - n_i(t)]\}, \quad (8a)$$

where

$$\alpha = 1 \quad , \quad \text{if } n_{i-1}(t) \leq Q_i(t) \quad (8b)$$

$$= w/v, \quad \text{if } n_{i-1}(t) > Q_i(t). \quad (8c)$$

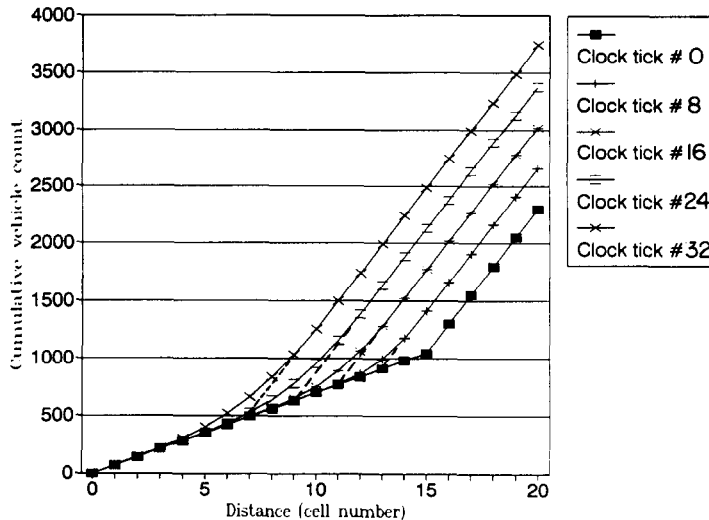


Fig. 5. Generalized cell-transmission model: softening of a shock with distance.

The modification ensures that shocks separating an upstream density lower than k_A from a downstream density greater than k_A cannot spread. This should be clear because for shocks of this nature $\alpha = 1$; eqn (8a) is then identical to eqn (1b), which does not exhibit spreading.

We claim that with this improvement any errors in the cumulative count near a shock (wave) should disappear once upstream light traffic ($v = 1$, $k < k_A$) catches up with the spreading shock (wave); as a result, any numerical errors should be short lived. We base this observation on the fact that the shock preceding the advancing light traffic is of the nonspreading type, and that it eventually must separate accurate upstream and downstream counts.

Perhaps, the above argument can be best understood if it is illustrated with an example. Consider the following initial density profile:

$$k(x,0) = k_j, \quad \text{if } 8 \leq x \leq 11 \\ = 25, \quad \text{otherwise.}$$

A diagram of $k(x,t)$ for this problem, also displaying the acceleration waves and shocks predicted by the hydrodynamic theory, is provided in Fig. 6b. We can see from it that all the changes in density for this problem are discontinuous and that, as a result, the cumulative profile $K(x,t)$ at any t should be a piecewise linear function, as shown in Fig. 6c. The solid lines in Fig. 7 are graphs of the cumulative profile for $t = 0, 8, 16, 24$ and 32 as predicted with eqns (1a) and (8); the dashed lines in the figure are the (piecewise linear) exact results. Although the acceleration waves spread (as shown by the concave bends of the curves), the (convex) shocks remain sharp. The figure clearly illustrates how the spreading wave is gradually "eaten up" by the advancing shock, until at time 24 and thereafter the exact and approximate results essentially coincide.

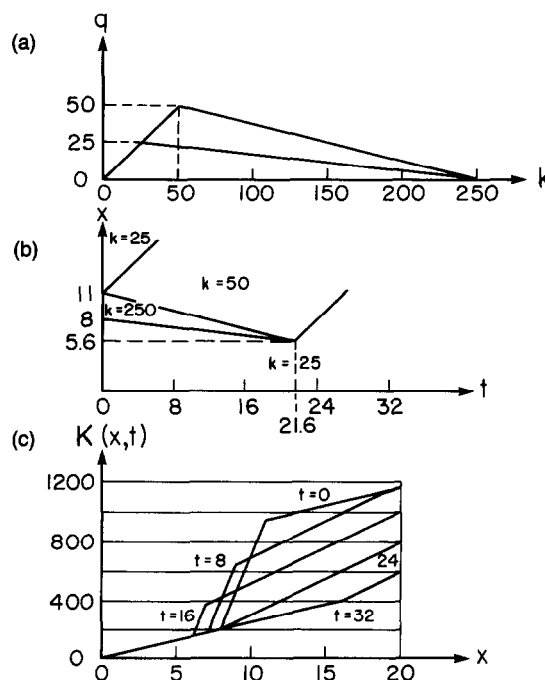


Fig. 6. Hydrodynamic solution to an example with a generalized flow-density relationship. (a) the q - k relationship. (b) The resulting density map, $k(x,t)$. (c) Cumulative density profiles at different times.

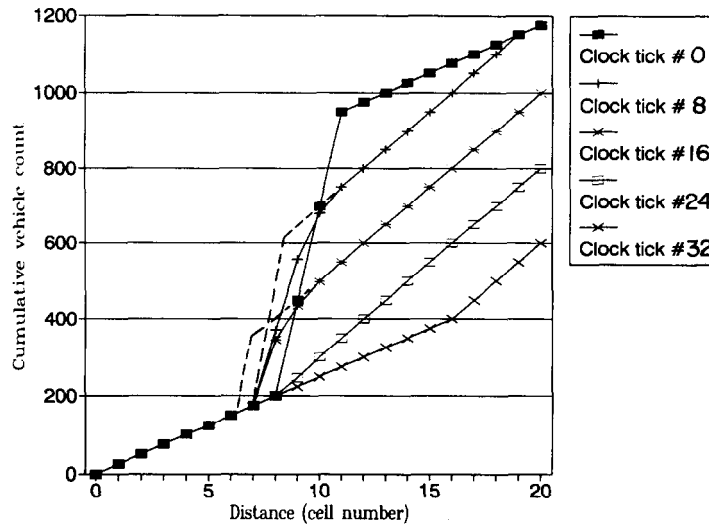


Fig. 7. Cumulative density profiles predicted by the generalized cell-transmission model for the example of Fig. 6.

6. INSTABILITY

So far we have argued that the cell-transmission model in one of its forms can easily produce results consistent with the hydrodynamic theory of traffic flow. This section takes an extra step; it shows that the cell-transmission model has the potential for capturing real-life instability phenomena, not included in the hydrodynamic theory.

Research on the stop-and-go phenomenon of congested freeway traffic dates back at least to Edie and Foote's (1961) and Edie's (1963) observations at the Lincoln and Holland Tunnels, and the George Washington Bridge in New York. Despite substantial efforts on the subject since then (witness for example the extensive car-following literature of the 1960's) a model of traffic instability that would account for the long periods of oscillation (approximately 1 minute long) observed in practice seems to have eluded researchers.

Newell (1963) had stated that an instability would arise if drivers catching up with denser/slower traffic ahead were to delay braking, perhaps in the hope that traffic would clear up before they had to slow down. This behavior would result in average spacings shorter than usual when traffic was decelerating (as observed by Edie) and would cause an instability. Of course, a microscopic car-following model to mimic this behavior would be extremely difficult to build (and validate) because in making their decisions drivers consider the recent evolution of the traffic stream immediately ahead, and not just the current status of one vehicle. A macroscopic model, consistent with Edie's observation that spacings are shorter when platoons are compressing, may be a more sensible goal to pursue.

Perhaps because of these difficulties, current efforts (see, for example, Ferrari, 1991) typically avoid seeking a behavioral explanation for instability and tend to focus instead on its control, using on-line macroscopic traffic measurements. This section also examines instability at the macroscopic level. It shows that the cell-transmission model, with a very simple modification that makes it consistent with Edie's observations, can duplicate real-life instability features.

Here we postulate that drivers operate in one of two modes depicted by the two q - k diagrams of Fig. 8a, depending on the traffic conditions prevailing in a "look-ahead" road section extending a fixed distance ahead of their current location. If the traffic density in this "look-ahead" section is greater than the traffic density in the vehicles' immediate neighborhood (assumed to be smaller than the "look-ahead" section), then we will assume that the vehicles would advance position as if they were regulated by the top curve of the figure. Otherwise, they will advance according to the lower curve².

²Other conditions for the on-off switching mechanism could have been postulated. For example we could have stipulated that drivers will advance more aggressively (according to the upper curve) if the traffic density

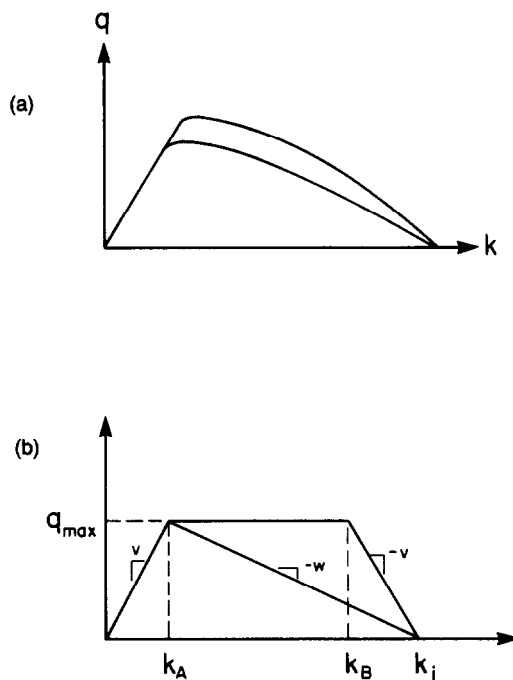


Fig. 8. Flow-density relationships for the generalized cell-transmission model with built-in instability. (a) General case. (b) Specific case.

The cell-transmission model can capture such conditions quite naturally. Without loss of generality, we define the cell length to be the distance of “the immediate neighborhood,” and examine a special case where the “look-ahead” distance is two cells. Then, only the occupancies of two neighboring cells need to be compared to determine any inter-cell flow. For the purposes of illustration the q - k curves are assumed to be of the special form depicted in Fig. 11b, as then no additional parameters need to be introduced. For this special case, the cell-transmission model will capture the desired behavior if the variable α of eqn (8a) is redefined as:

$$\alpha = 1 \quad , \quad \text{if } n_{i-1}(t) \leq n_i(t) \text{ or } Q_i(t) \quad (9a)$$

$$= w/v, \quad \text{otherwise.} \quad (9b)$$

Although much experimental evidence would be needed to make a strong claim for realism, the results described below at least seem to be qualitatively consistent with real-life behavior.

6.1. Results

Here we describe the numerical results of a number of simulations that were done with the cell-transmission model, using eqns (8a) and (9). The tests involve a homogeneous finite section of highway with a large input flow and a restricted steady output. Because the output is constant, the (stable) hydrodynamic steady state solution is a spatially uniform density. (The value of said density is given by the point on the congested part of the q - k curve that has q equal to the output flow.) If the system is assumed to be in steady state at time $t = 0$, eqns (9) don't generate any instability; they produce the same result as eqn (6) and eqns (8), of course, matching the hydrodynamic prediction.

We explore here the evolution of the system as predicted by eqns (9) when a single

in their neighborhood has been increasing *in time*. Alternatively, aggressive behavior could be stipulated when both conditions are met (or else either one of them). Fortunately, the specific trigger is not important because, with backward moving waves (relative to a vehicle) as result from our model, if one of the conditions is met so are the others; all the models should be similar.

random disturbance to the exit flow is introduced at time $t = 0$ (*i.e.*, a slightly larger or smaller output flow for the first clock tick only). Experiments have been conducted for a number of examples, and these will be discussed here qualitatively. Numerical results will be given for one of the examples (for illustration only) corresponding to the following parameters of Fig. 8b: $q_{\max}(Q_i) = 50$, $k_j(N_i) = 150$, $w/v = 0.5$; and the following initial conditions: $Q_{\text{output}} = 25$ and $n_i = 100$.

In all the cases tested, and independently of the magnitude of the initial impulse, a flow disturbance was generated that grew in duration and magnitude as it traveled backward. The disturbance is first noticed by a temporary increase in flow, which is then followed by a shorter but sharp reduction, finally terminating with a gradual return to normalcy. The disturbance does not generate any lasting effects; after its passage, vehicles are observed to go by any location at the time they would have passed without the disturbance. However, affected vehicles are observed to pass . . . earlier! We are not sure whether this would also occur in reality, and unfortunately it seems difficult to design an experiment to verify it. In a way the answer to this question is moot because, with a constant output flow, every vehicle will pass the bottleneck at its appointed time. A small disturbance cannot cause any delay.

Figure 9a contains plots of cumulative count vs. time at 5 locations separated by 3 cells. Notice how the disturbance grows as described, eventually including time intervals with nearly zero flow. From the time at which the shock front (the crest of each cumulative count curve) passes the various locations we can infer the disturbance's speed. In the figure, as in the other cases tested, the disturbance traveled (roughly) with the wave speed w . This should not be surprising for a disturbance that is no longer growing: the disturbance must move with the speed of its decreasing density portion, which must be w since eqns (9) and (8) are equivalent when the density decreases.

We also performed experiments in which a steady random noise was added to the exit flow—the noise was identical to the described impulse at $t = 0$, but operating at all

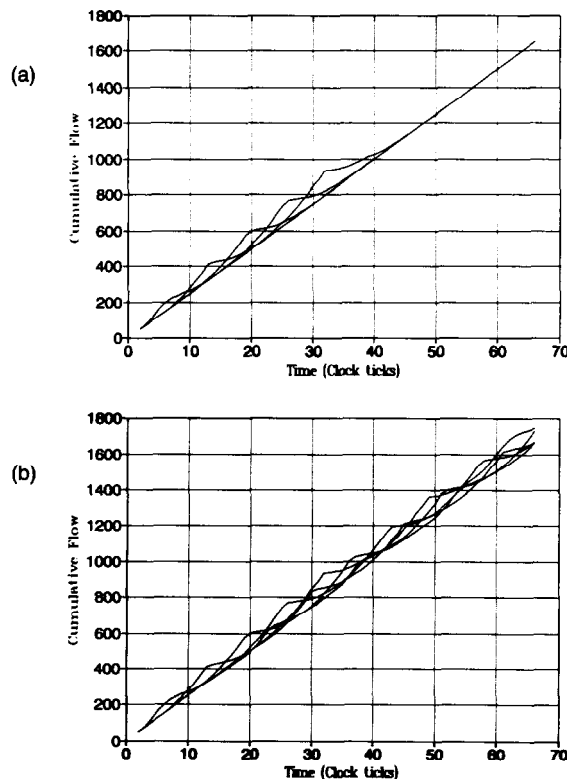


Fig. 9. Curves of cumulative flow versus time observed at different positions. (a) A single impulse at the exit. (b) Steady noise at the exit.

times. With such noise, regularly oscillating waves would invariably develop upstream from the exit, with a period of oscillation that was the same at all locations and a phase shift of speed w . Both the wave growth pattern and the oscillation period were surprisingly reproducible, *i.e.*, largely independent of the set of generating impulses, perhaps providing an explanation for the regularity of stop-and-go traffic waves. Figure 9b illustrates these comments; the figure represents a situation identical in all respects to 9a, except that the single impulse has been replaced by steady noise. This figure can be compared with the flow-wave results of Edie and Foote (1961), given in Fig. 10. Note that a certain regularity was also observed in these experiments and that, as in Fig. 9, the disturbances also grew with the distance from the tunnel's exit.

For the same combination of parameters, Fig. 11 depicts the oscillations recorded in the q - k plane. Each point in the figure corresponds to a combination of $y_i(t)$ and the average of the occupancies in the sending and receiving cells during the two time intervals immediately preceding and following clock tick i :

$$(1/4)[n_{i-1}(t) + n_i(t) + n_{i-1}(t+1) + n_i(t+1)].$$

The pattern of the figure, insensitive to the impulse set, is also observed at other locations, although, obviously, there will be a scaling discrepancy wherever the disturbance is not fully grown. Different patterns, also insensitive to the impulse set and the specific location, are obtained for other wave speeds and initial densities.

The number of ticks for one oscillation was found to decline with w and to depend on the metering rate; the number was close to $8/(w/v)$ for $q \approx q_{\max}/2$. Thus, a period of a few minutes, as reported in Edie and Foote (1961), would arise with $w/v = 1/6$ and a clock tick of about 5 seconds—the “look-ahead” section would then be a few hundred feet long.

Our findings seem to be consistent with the observations of Treiterer and Myers (1974), who documented the growth of one disturbance. Interesting to note is the similarity of Fig. 11 and Fig. 4 of this reference (reproduced here as Fig. 12a); the similarity should be qualified, however, because Treiterer and Myers' figure depicts the q - k oscillations seen by a moving observer, and Fig. 11 does not. Other comparisons can also be made. From the trajectories in Treiterer and Myers' Fig. 2 (reproduced here as Fig. 12b), the following parameters can be estimated: $v = 54$ MPH, $q_{\max} = 1850$ VPH, and $k_j = 240$ VPM. For a diagram such as the lower one of Fig. 8b, this corresponds to a wave speed of about 9 MPH (consistent with the observed speed of the disturbance). We used a cell-transmission model with a 1.0-sec clock tick and 0.015-mi cell length, which ensures that $v = 1$ cell per tick = 54 MPH. With these basic features, the parameters of the cell-transmission model are $N = 3.6$, $Q = .51$, and $w/v = 1/6$. The original state observed in the figure ($k \approx 80$ VPM $> k_{op}$) corresponds to $n = 1.2$. In rough agreement with the trajectory pattern of Fig. 12b, the numerical results revealed a disturbance that grew in spatial dimension by about 60 meters per minute, traveling at about 7 MPH during this transitional phase.

The above comparisons are not meant to establish the validity of the model in Fig. 8b; only that it may have the potential for reproducing real-life phenomena. A very large empirical effort would be needed to pinpoint more precisely a “correct” model. After all, typical instability features such as a disturbance's growth rate and its final equilibrium size may depend on local conditions, perhaps even changing across facilities just as k_j , q_{\max} , w , and v do. Besides changing the clock tick (and cell size), reducing the slope of the upper curve of Fig. 8b, and increasing the number of cells for the “look-ahead” section might also capture these variabilities. The clock tick should be adjusted to match observed oscillation periods, and the upper slope to match the disturbance's growth pattern. (We have found from our experiments that with a shallower upper branch for the diagram in Fig. 8b the system becomes more stable, and when disturbances grow they grow at a slower rate.) We have not explored the effect of longer “look-ahead” sections. Another generalization might involve a model where the transition flows (eqn 8a) themselves include random noise.

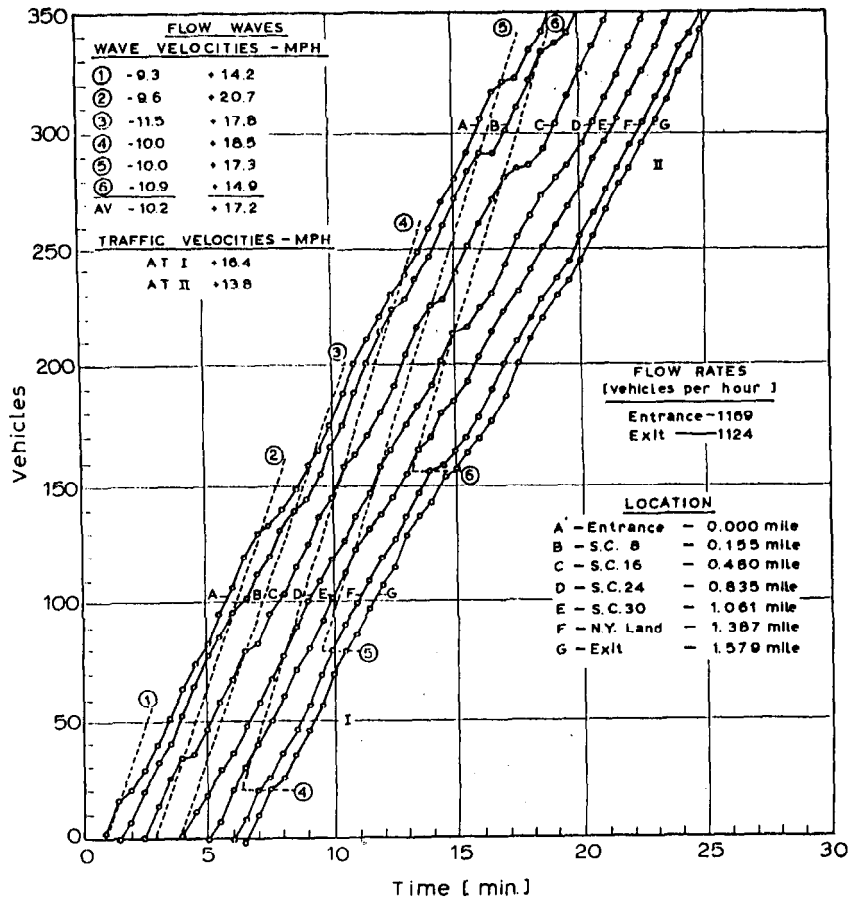


Fig. 10. Empirical cumulative curves of flow versus time, Edie and Foote (1961). Holland tunnel, south tube, fast lane. Flow wave experiments.

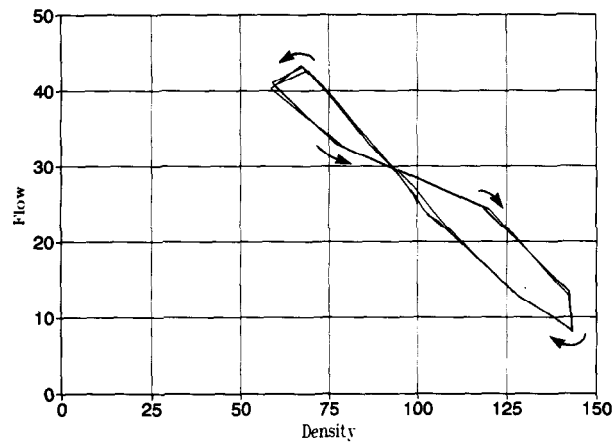


Fig. 11. Flow-density hysteresis curve seen by a stationary observer.

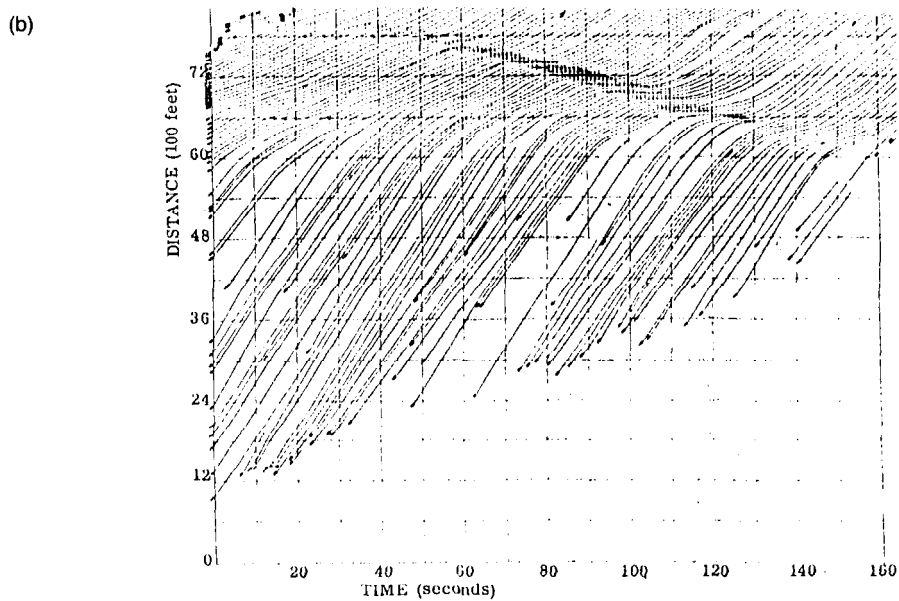
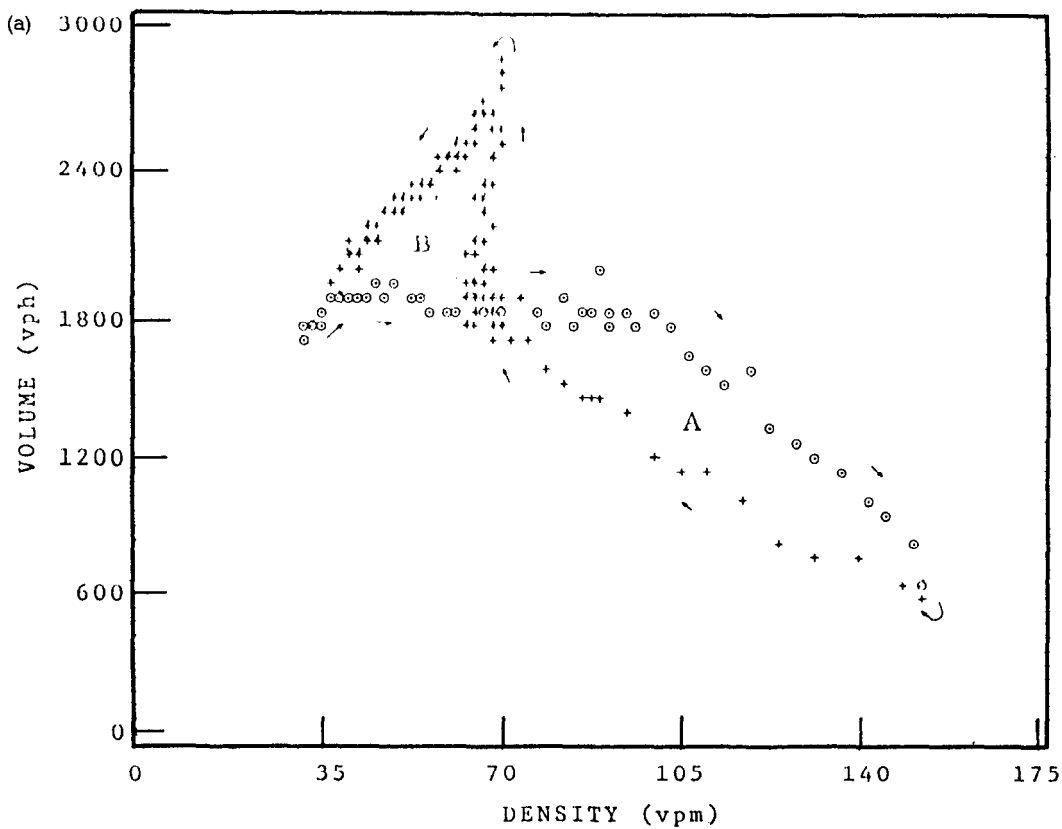


Fig. 12. Empirical observations of Treiterer and Myers (1974). (a) Flow-density hysteresis. (b) Vehicle trajectories.

7. CONCLUSION

Although the results in Sections 5 and 6 have been developed for the specific form of the q - k relationship depicted in Fig. 4, the form is rather general. It offers 4 degrees of freedom: the free flow speed, the maximum flow, the jam density, and the wave speed. As pointed out by Newell (1993), these are the most important determinants of traffic evolution; and even if a more general model were available, it is unlikely that in any practical application an engineer would have reliable data beyond these parameters. Arguably, thus, a model with so much flexibility should be accurate enough for modeling complex networks. Field experiments are currently being planned to assess the accuracy of the model.

The method is very robust and matches the results of the hydrodynamic theory, even for long clock intervals and large cell sizes. Errors in timing are comparable to the clock interval, and errors in location are comparable to the cell size. Thus, in any practical application, one should choose the longest clock step consistent with one's objectives. An illustration and further discussion of this property can be found in Daganzo (1993).

Acknowledgements—This research was supported by PATH MOU 90, Institute of Transportation Studies, Berkeley, CA.

REFERENCES

- Cremer M. and Ludwig J. (1986) A fast simulation model for traffic flow on the basis of Boolean operations. *Math. Comp. in Simul.* **28**, 297–303.
- Daganzo C. F. (1993) The cell transmission model. Part I: A simple dynamic representation of highway traffic. Research Report UCB-ITS-PATH-RR-93-7 University of California, Berkeley.
- Edie L. C. (1963) Traffic stream measurements and definitions. In J. Almond, (Ed.), *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, London, pp. 139–154. Published by OECD, Paris, France, 1965.
- Edie L. C. and Foote R. S. (1961) Experiments on single-lane flow in tunnels. In R. Herman, (Ed.), *Proceedings of the 1st International Symposium on the Theory of Traffic Flow*, Warren, MI, 1959, pp. 175–192. Published as *Theory of Traffic Flow*, Elsevier Publishing Co., New York.
- Ferrari P. (1991) The control of motorway reliability. *Trans. Res.* **25A**, 419–427.
- Hurdle V. F. (1992) A theory of traffic flow for congested conditions on urban arterial streets. II: Four illustrative examples. *Trans. Res.* **26B**, 397–415.
- Lighthill M. J. and Whitham J. B. (1955) On kinematic waves. I: Flow movement in long rivers; II: A theory of traffic flow on long crowded roads. *Proc. Royal Soc. A*, **229**, 281–345.
- Luke J. C. (1972) Mathematical models for landform evolution. *J. Geoph. Res.* **77**, 2460–2464.
- Nagel K. and Schreckenberg M. (1992) A cellular automaton model for freeway traffic. *J. de Physique I*, **2**, 2221–2229.
- Newell G. F. (1963) Instability in dense highway traffic: a review. In J. Almond, (Ed.), *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, London 1963, pp. 73–83. Published by OECD, Paris, France, 1965.
- Newell G. F. (1993) A simplified theory of kinematic waves. I: general theory; II: Queuing at freeway bottlenecks; III: Multi-destination flows. *Transpn. Res.* **27B**, 281–314.
- Richards P. I. (1956) Shockwaves on the highway. *Opns. Res.* **4**, 42–51.
- Schwerdtfeger T. (1984) DYNEMO: A model for the simulation of traffic flow in motorway networks. In J. Vollmuller and R. Hamerslag (Eds.), *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, pp. 65–87. VNU Science Press, Utrecht, The Netherlands.
- Sheffi Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, NJ.
- Treiterer J. and Myers J. A. (1974) The hysteresis phenomenon in traffic flow, In D. J. Buckley (Ed.), *Proceedings of the 6th International Symposium of Transportation and Traffic Theory*, pp. 13–38. A. H. & A. W. Reed, Sydney, Australia.
- Vaughan R. and Hurdle V. F. (1992) A theory of traffic flow for congested conditions on urban arterial streets. I: Theoretical developments. *Trans. Res.* **26B**, 381–396.
- Vaughan R., Hurdle V. F., and Hauer E. (1984) A traffic flow model with time-dependent O-D patterns. In J. Vollmuller and R. Hamerslag (Eds.), *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, pp. 155–178. VNU Science Press, Utrecht.