CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

The Cell Transmission Model: Network Traffic

Carlos F. Daganzo

University of California, Berkeley

California PATH Working Paper

UCB-ITS-PWP-94-12

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department Transportation, Federal Highway Administration.

August 1994

ISSN 1055-1417

THE CELL TRANSMISSION MODEL: NETWORK TRAFFIC

by

CARLOS F. DAGANZO

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

THE CELL TRANSMISSION MODEL: NETWORK TRAFFIC

by

CARLOS F. DAGANZO

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

California PATH Working Paper UCB-ITS-PWP-94-12

Abstract

This paper shows how the evolution of multicommodity traffic flows over complex networks can be predicted over time, based on a simple macroscopic computer representation of traffic flow that is consistent with the kinematic wave theory under all traffic conditions. The method does not use ad-hoc procedures to treat special situations.

After a brief review of the basic model for one link, the paper describes how three-legged junctions can be modeled. It then introduces a numerical procedure for networks, assuming that a time-varying origin-destination table is given and that the proportion of turns at every junction is known. These assumptions are reasonable for numerical analysis of disaster evacuation plans.

The results are then extended to the case where, instead of the turning proportions, the best routes to each destination from every junction are known at all times. For technical reasons explained in the text, the procedure is more complicated in this case, requiring more computer memory and more time for execution. The effort is estimated to be about an order of magnitude greater than for the static traffic assignment problem on a network of the same size. The procedure is ideally suited for parallel computing.

Hopefully, the results in this paper will lead to more realistic models of freeway flow, disaster evacuations and dynamic traffic assignment for the evening commute.

Keywords: Traffic Flow, Traffic Assignment, Traffic Delay

Executive summary

This paper shows how the evolution of multicommodity traffic flows over complex networks can be predicted over time, based on a simple macroscopic computer representation of traffic flow. An experimental computer program which carries out the procedure has been written; a polished version for public use should be available next year.

A congested traffic network is modeled by a set of cells (representing freeway segments) and their interconnections. It keeps track of the overall traffic state over time with an algorithm that is consistent with the kinematic wave theory of traffic flow. As such, the model is able to keep track of the location of moving queues in the network, predicting queue spillbacks and dissipation in a reasonable way. (Its results match satisfactorily the predictions of the kinematic wave theory for special cases that can be solved by hand.) Validation efforts are currently under way; they seem to indicate that the predictions of the theory are weakest near congested exit ramps, but the results are still preliminary.

At each point in time, the state consists of the number of vehicles located in each cell classified by destination and time of entry. It appears that the memory requirements implied by this structure are about an order of magnitude larger than for an equivalent static traffic assignment problem. Recent theoretical results, not included in this report, indicate that it might be possible to use larger cells (e.g. 1Km long) on long homogeneous freeway segments and shorter cells (e.g. 100 m long) on ramps and interchanges. The memory requirements of the models could then be reduced at the cost of some accuracy.

The execution time of a single time step is also comparable to the time for an iteration of the static equilibrium traffic assignment model, which is proportional to the number of the destinations but grows supralinearly with the number of nodes in network. Because we may need to perform more iterations than are normally done for static equilibrium models in order to simulate a rush hour period, we estimate that the execution time may also be an order of magnitude greater than for static equilibrium models. Fortunately, the nature of our formulation allows a large regional network to be decomposed into small subnetworks corresponding to different subareas of the study region; and to operate on them simultaneously and independently.

The paper does not address the dynamic route choice problem, but the results can be used to model drivers who choose their routes prior to starting a trip. For example, one could assume that drivers respond to some weighted average of their past driving experiences and that such experiences determine their choices. The computer program described in this paper could then be used to simulate a sequence of days to see if an equilibrium is reached. The model could also be used to determine the type of "real-time" infortation that is likely to be of most value to drivers.

Preliminary analyses with the model indicate, to not great surprise, that driver advisories should anticipate the projected evolution of the system; and that otherwise one can do more harm than good!

Dynamic network models can also be applied to evaluate the performance of emergency evacuation plans developed in response to the possibility of a disaster such as a nuclear meltdown. Because in such an instance time is of the essence, a realistic model of network performance under a dynamic load is necessary. Evacuations are easier to study than the general dynamic traffic assignment model because traffic can be modeled as a single commodity flowing to a unique hypothetical destination, "safety".

TABLE OF CONTENTS

Abstract ii
Executive Summary iii
Table of Contents iv
1 Introduction **1**
2 The Cell-transmission Representation of a Network **3**
- **2.1** Background **3**
- **2.2** The Network Representation **5**
- **2.3** Network Topologies Allowed *5*

3 The Difference Equations: Percentage of Turns Known **6**
- **3.1** Ordinary Links **7**
- **3.2** Merges **8**
- **3.3** Model of a Diverge **11**

4 The Model with Known Routes: Turning Percentages not Specified **12**
- **4.1** The Procedure with Disaggregated Occupancies and Flows **12**
- **4.2** Minimum Link Waits **13**

5 Computational Issues **15**
6 Applications and Comments **16**
7 Further Work **16**
References **17**

FIGURES

Figure **1:** The equation of state of the cell-transmission model **4**
Figure **2:** Representation of a merge and diverge **6**
Figure **3:** Valid and invalid representations **6**
Figure **4:** Diagram of feasible flows for a merge junction **9**
Figure *5:* Graphical representation of $y_{kd\tau}(t,a_{Bk})$ and $yk(t,a_{Bk})$ **14**

1. INTRODUCTION

Despite all the attention that "dynamic assignment" is receiving today in the literature (Transportation Research, 1991), most of the research efforts in that field seem to be directed at improving the route choice mechanisms of drivers with various levels of information, and/or at refining the algorithms. Little work is aimed at enhancing the realism of the basic building blocks of the predictions — the underlying traffic performance models. A recent overview of this literature is given in Ran (1993) and Janson and Robles (1993).

Some of the works reviewed in these references attempt to predict some form of system "equilibrium" assuming that the travel time on an arc of the network can be expressed as an increasing function of the flow on the arc at the time. This, however, is a futile exercise for a rather obvious reason: if a bottleneck causes a queue spanning a whole arc to form, the result would be high travel times and low flow — an outcome that cannot be predicted with a simple flow-time relationship.

In an attempt to correct this deficiency, some of the more advanced models define the arc travel time for an entering vehicle as a function of the current arc occupancy as well as entering and exiting flows. Unfortunately, such a generalization doesn't have the desired effect: absurd results are still obtained whenever a vehicle's travel time on an arc is allowed to depend in any way on the arc entry and/or exit flow at the time of entry (Daganzo, 1993c).[1]

The above comments are not meant to be critical of specific works[2]. Rather, they are meant to illustrate that the state of the art in dynamic traffic assignment is not yet very advanced, and to establish the need for improvements. A very basic theory of dynamic assignment would have to include at the very least realistic models of:

(i) traffic behavior when the vehicle paths are known (so that arc travel times can be predicted),

(ii) path choice when the time-varying arc times are known, and

(iii) equilibrium to reconcile the predictions of (i) and (ii)

Arguably, item (i) is the least understood of these three items, and will thus be the focus of the following discussion.

The conventional macroscopic approach to simulate freeway traffic behavior over networks keeps track of the number of vehicles in discrete sections of the network as time passes. The vehicle occupancy in each section is increased by the number of vehicles allowed to enter the section in each time interval, and is decreased by the number of vehicles allowed to leave it.

When simulating the kinematic wave model of Lighthill and Whitham (1955) and Richards (1956) — the LWR model — the outflow is typically specified to be a function of the occupancy of

[1] A formulation based on arc occupancy alone can make some physical sense in the special case where it can be iterpreted as a network of dimensionless (or Point) queues but such a representation could only apply if congestion is so mild that queues do not spill over arcs. If this happens, the predictions with a point ueue model can be diametrically opposed to those of a model with spatial queues (Daganzo and Lin, 19933.

[2] The coarseness of models, in a way, should be expected because transportation networks are more complicated than those arising in other fields (than e.g. electrical networks or hydraulic networks), as they involve particles that are destined for specific points in the network with little understood interactions.

the section emitting the flow and not to be an explicit function of the downstream occupancy. Such approaches don't converge to the desired solution and cannot produce reasonable results (Newell, **1988)**[3]**;** obviously, traffic could be sent into a section even after it reaches its maximum occupancy. Perhaps in recognition of this problem some researchers have proposed to introduce constraints on the flows to ensure that section occupancies remain between zero and the maximum possible (Sheffi et al, 1982, Chang et al, 1985, Algadhi and Mahmassani, 1990). Unfortunately, such improvements do not guarantee convergence; a symptom of non-convergence is that stopped traffic is still predicted not to flow into an empty freeway.

For higher order models, alternative formulations which include downstream information have been proposed but they still require heuristic protections against unreasonable results (Cremer et al, 1993). Because the dynamics of higher order models have not been properly linked to any model of sensible driver behavior — as has been done for the LWR model (Newell, 1961) — and because said models have been shown to exhibit undesirable properties (del Castillo et al, 1993), this paper will focus on the LWR model.

The non-convergence problem is serious because as is well known (see Ansorge, 1990, for example) the LWR model usually has multiple solutions but only one that is physically relevant. Numerical methods must not only approximate the partial differential equations of the LWR theory but they must also identify automatically the proper solution.

Ways of addressing these difficulties for a single link exist. See the methods of Lax (1954) — tested in Michalopoulos, Beskos and Yamauchi (1984) — Luke (1972), Newell (1993), and Daganzo (1993b). **Lax's** procedure is a finite difference approximation[4], whereas Luke's and Newell's methods are exact. Lax's method converges to the physically relevant LWR solution but uses a procedure where vehicles don't always move forward; this may limit its practical appeal. Luke's method is computationally demanding but it can be streamlined in an important special case (Newell, 1993); with Newell's method solutions can be easily obtained by hand. Newell (1993) also describes a procedure for handling a single freeway. The finite difference procedure in Daganzo (1993b) applies to the special case considered in Newell (1993). It moves vehicles forward and converges rapidly. (Preliminary indications are that the procedure can be extended to the general case, Daganzo, 1993d). Although more memory intensive than Newell's method, the procedure seems easier to program and can be readily generalized to complicated networks. Here is how this can be done.

In order to limit the number of junction types to be studied, we shall restrict our attention to networks with three-legged junctions. In that case, only two junction types are needed: "merges" and "diverges". Most freeway networks can be represented in this manner. For the most **part** it will also be assumed that the best route between every origin and every destination is independent of traffic conditions, and is therefore known. Route choice issues are discussed at the end.

After a brief description of the cell-transmission results in Daganzo (1993b), the next section describes how a network with three-legged junctions should be represented. Section 3 then introduces equations for merges and diverges that define boundary conditions for every arc of the

[3] This reference also discusses methods intended to approximate higher order models (e.g. Payne, **1971).**

[4] Michalopoulos et al. **(1993)** test and describe other (more complex) finite difference approximations which require "fitting" the shocks.

network. These equations are used in an approximation procedure for networks, where it is assumed that the turning percentages at every diverge are given. Section **4** shows how the boundary conditions and the network approximation procedure should be modified when the turning percentages are determined by the destination mix in the traffic stream. The section also describes a procedure for predicting the evolution of traffic over a network. Section 5 describes the accuracy and efficiency of the procedure. Finally, section ***6*** briefly discusses route choice and equilibrium issues (items ii and iii above).

2. THE CELL-TRANSMISSION REPRESENTATION OF A NETWORK

2.1 Background

In Daganzo (1993b) it is shown that if the relationship between traffic flow (q) and density (k) is of the form depicted in figurel:

$$q = \min \{ vk , q_{max} , w (k_j - k) \}, \quad \text{for } 0 \le k \le k_j, \tag{1}$$

then the LWR equations for a single highway link can be approximated by a set of difference equations where current conditions (the state of the system) are updated with the tick of a clock. In the above expression v, q_{max} w and k_j are constants denoting respectively: the free-flow speed, the maximum flow (or capacity), the speed with which disturbances propagate backward when traffic is congested (the backward wave speed), and the maximum (or jam) density.

The method assumes that the road has been divided into homogeneous sections (cells), i, whose lengths equal the distance traveled by free-flowing traffic in one clock interval. (Although a closer approximation to the LWR results is obtained with short cell lengths (e.g. 100 meters) the procedure can be applied with cells of any length.) The state of the system at instant t is then given by the number of vehicles contained in each cell, $n_i(t)$. The following parameters are defined for each cell:

$N_i(t)$, 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.

These constants can vary with time (e.g. as per the occurrence of transient traffic incidents) but this dependence will be ignored in this paper for simplicity of notation. The first constant **is** defined to be the product of the cell's length and its jam density, and the second one the product of the clock interval and the cell's capacity.

If cells are numbered consecutively starting with the upstream end of the road from i = **1** to **I**, the recursive relationship of the cell-transmission model can be expressed as:

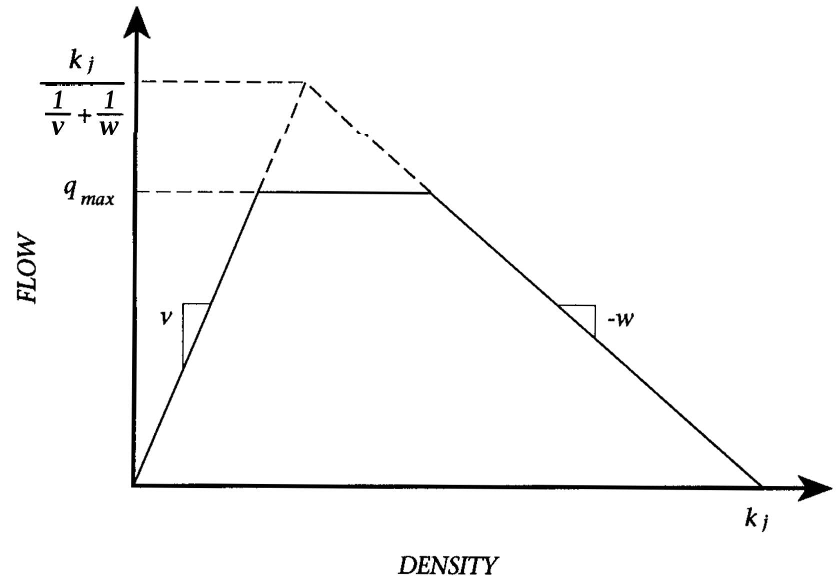


Figure 1: The equation of state of the cell-transmission model.

$$n_i(t+1) = n_i(t) + y_i(t) - y_i+1(t) \tag{2a}$$

where $y_i(t)$ is the inflow to cell i in the time interval (t,t+1), given by:

$$y_i(t) = \min \{ n_{i-1}(t) , Q_i(t) , \delta [N_i(t)-n_i(t)] \}, \tag{2b}$$

where $\delta = w/v$.[5] Note the similarity of (1) and (2b).

For finite roads, 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 should then be set equal to the desired link input flow for the corresponding time interval. 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. (Although it may be possible to eliminate gate cells in an efficient computer implementation, the program logic should preserve their effects. Gate cells ensure that any time dependent O-D table can be handled; i.e. that the LWR problem is well-posed for all O-D tables.)

[5] For reasons explained in Daganzo (1993b), the accuracy of the approximation is enhanced if a **is** redefined **as:** **6 = 1, if** $n_{i-1}(t) \le Q_i(t)$ and 6 = w/v , if $n_{i-1}(t) > Q_i(t)$.

2.2 The network representation

A general transportation network is usually described by a directed graph of nodes and arcs, including some physical data for each arc. We take this representation as a point of departure for our discussion, assuming that an arc's physical data includes its length and the parameters defining a q-k relation of the type shown in figure 1.

It should be clear that each arc of a network can be treated as in section 2.1, once appropriate boundary conditions at both of its ends have been defined. Thus, we assume that each arc of the graph has been subdivided into cells, as explained above. For general graphs, however, it is no longer possible to number the cells consecutively and specify that vehicles should always proceed to the cell numbered next.

Instead, it is convenient to describe the system of cells as the nodes {I} of a more detailed graph, and the possible vehicle transfers by a set of links {k}. (From now on we use the term "link" for the components of the detailed graph to avoid confusion with the arcs of the original graph.) Capital letters will be used to denote cells and lower case letters for links.

The topology of the detailed network is defined by specifying for each link, k, a "beginning" cell and an "ending" cell. These will be denoted by adding the prefixes "B" and "E" to the link label, so that the beginning and ending cells of k become: Bk and Ek.

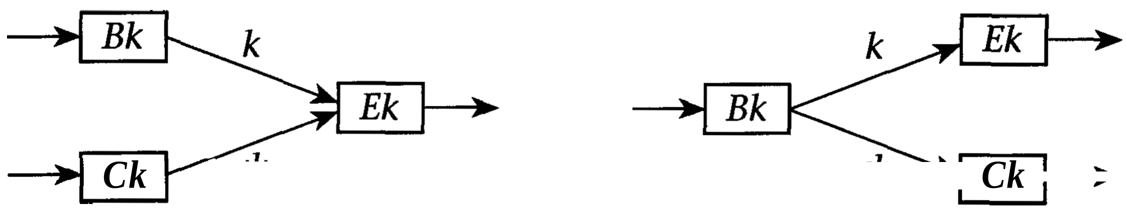
In section 2.1 Q denoted the maximum number of vehicles that could enter cell i per unit time, assuming that there was room to store them; i.e., it represented the maximum possible flow on the link from i-1 to i. Alternatively, we could have defined Q to be the maximum possible flow through cell i, as an indicator of cell "width", and then defined the minimum of Q_{i-1} and Q to be the maximum flow from i-1 to i. This revised definition is more convenient for network modeling, and will be adopted here. Thus, each cell will be characterized by a maximum occupancy N_I and a maximum throughput Q.

The above departs from the classical representation of a transportation network in that both, the state of the system (i.e. the cell vehicle occupancies at any given time) and the network performance characteristics (N_I and Q), are nodal quantities. In our formulation, links will play a minor role; they will simply define the connection of the cells, ensuring that vehicles are transferred among the proper nodes.

2.3 Network topologies allowed

We consider here networks where the maximum number of arcs (links) entering and/or leaving a node (cell) is 3. Thus, cells can be classified into three types: "diverge" if only one link enters the cell but two leave it, "merge" if two links enter and one leaves, and "ordinary" if one enters and one leaves. Origins and destinations can be modeled with ordinary cells as explained in section 2.1.

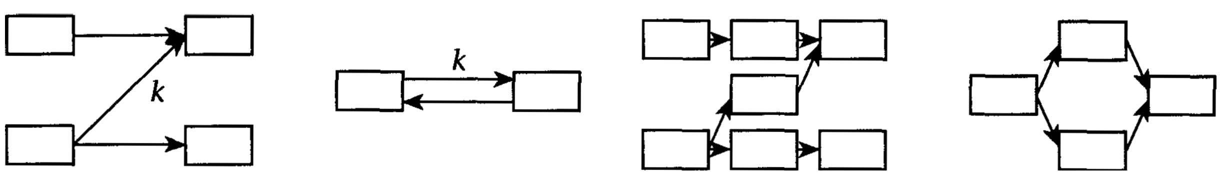
The basic modeling block, a three-legged junction, will consist of the merge (diverge) cell, the two links entering (leaving) it and the two cells at the other end of these links; see figures 2a and 2b. In order to identify all the components of the junction relative to one of its links (k), the prefixes "c" and "C" will be used: "ck" will denote the other (complementary) link in the modeling block, and "Ck" the third (complementary) cell; i.e. the cell which is neither Bk nor Ek. The flows on links k and ck will then be determined from functions similar to equation (2b), that would have as arguments the characteristics and occupancies of nodes Bk, Ek and Ck.



(a) **MERGE** (b) **DIVERGE**

Figure 2: Representation of a merge and diverge.

Before investigating the form of these equations note that a highway network with three-legged junctions might then include some links that belong to more than one junction; e.g. as part of both a diverge and a merge (see figure 3a). The presence of such links would complicate the link flow equations that are about to be presented, and will not be allowed. No generality is lost because multi-junction links can be eliminated with a faster ticking clock (and shorter cells), as shown in figure 3b. Thus, with our representation, links can only belong to one of three classes: "merge links" which belong to a merge junction, "diverge links" which belong to a diverge junction, and "ordinary links" which join ordinary cells.



(a) **INVALID** (b) **VALID**

Figure 3: Valid and invalid representations.

3. THE DIFFERENCE EQUATIONS: PERCENTAGE OF TURNS KNOWN

The general procedure for networks involves two steps for each tick of the clock

(i) Determine the flow on each link with the equivalent of equation (2b).

(ii) Update the cell occupancies by transferring the flows of step (i) from the beginning-cell to the end-cell of each link.

The procedure is now explained for ordinary links, merges and diverges.

3.1 Ordinary links

If we let $y_k(t)$ denote the flow on link k from clock tick t to clock tick t+1, and 6, the wave speed coefficient of cell I , the equivalent of (2b) for an ordinary link is:

$$y_k(t) = \min\{ n_{Bk} , \min[Q_{Bk},Q_{Ek}] , \delta_{Ek}[N_{Ek}-n_{Ek}] \}.$$

For simplicity of notation, the time variable "t" is omitted from the right side of the above and in forthcoming expressions. It should be understood that any time-dependent quantities should be valued at "t", unless explicitly noted otherwise.

A further simplification is desirable. If we define,

$$S_I(t) = \min\{Q_I, n_I\} \quad \text{and} \tag{3a}$$

$$R_I(t) = \min\{Q_I, \delta_I[N_I-n_I]\} \tag{3b}$$

as the maximum flows that can be sent and received by cell I in the interval between t and t + 1, then we can write $y_k(t)$ in the more compact form:

$$y_k(t) = \min\{ S_{Bk}, > R_{Ek} \} \tag{4}$$

That **is,** the flow on link "k" should be the maximum that can be sent by its upstream cell unless prevented to do so by its end cell. If blocked in this manner, the flow is the maximum allowed by the end cell.

Equation **(4)** is interesting because it indicates the direction of causality. During time periods when $S_{Bk} < R_{Ek}$ the flow on link k **is** dictated by upstream traffic conditions—as would be predicted from the forward moving characteristics of the LWR model. Conversely, when $S_{Bk} > R_{Ek}$, flow is dictated by downstream conditions and backward moving characteristics.

Step (ii) is then completed by taking the link flows away from the beginning cells and adding them to the ending cells:

$$n'_{Bk}(t+1) = n_{Bk}(t+1) - y_k(t) , \quad \text{for all } k^{6} \tag{5a}$$

$$n_{Ek}(t+1) = n'_{Ek}(t+1) + y_k(t) \quad \text{for all } k. \tag{5b}$$

We now explain the extensions of equation (2b) for merge and diverge links. We assume that the LWR model holds for each arc of the original network, subject to proper boundary conditions at each of its ends. In Daganzo (1993b) the boundary conditions were captured by origin and destina-

[6] The occupancies, $n'_I(t+1)$, are intermediate variables introduced for mathematical notation purposes; they can be eliminated during computer implementation.

tion cells with a time dependent capacity. For an arc ending in a merge or diverge, the boundary conditions should relate the flow at its extreme point to the flows at the extremes of the other arcs incident on the same junction.

Note that in the cell-transmission representation the boundary conditions will only involve the flows on the two links of a merge or diverge. Such conditions must include a flow conservation equation. This, however, is not sufficient to determine future flows from current conditions; other side conditions are needed. They should depend on the particular characteristics of the junction (e.g. priorities) and will be introduced separately for merges and diverges.

3.2 Merges

In the real world, a merge can be in one of three possible causality regimes:

(a) Forward. If the flow on both approaches is dictated by conditions upstream; i.e. waves move forward.

(b) Backward. If the flow on both approaches is dictated by conditions downstream; i.e. waves move backward.

(c) Mixed. If the flow is dictated by conditions upstream for one approach and down stream for the other.

Case (a) arises if both approaches are flowing freely. Case (b) arises when both approaches are congested due to the junction's lack of capacity, or to congestion downstream. Case (c), which is less common, arises when an approach with priority crowds out traffic on its complementary approach.[7] Boundary conditions should be found that properly reflect these three possibilities.[8] We specify conditions directly in cell-transmission form; the continuous model equivalences follow trivially.

Given the maximum flow that can be emitted by the two sending cells (S_{Bk} and S_{Ck}), the boundary equations should specify the advancing flows y_k and y_{ck} as a function of the maximum flow that can be received immediately downstream: R_{Ek}. (The variables S_{Bk} , S_{Ck} and R_{Ek}, which capture the current state of the system at the junction, are proxies for the traffic densities at the junction.)

Clearly, the flows must satisfy:

$$y_k(t) \le S_{Bk} \,;\; y_{ck}(t) \le S_{Ck} \quad \text{and} \tag{6a}$$

$$y_k(t) + y_{ck}(t) \le R_{Ek} \tag{6b}$$

But this is not enough to identify the flows unless additional relations are introduced. As it was done for ordinary cells and links, it will be assumed that cells Bk and Ck send the maximum traffic

[7] In this last case conditions upstream of the intersection on the congested approach are dictated by past conditions upstream of the merge on the uncongested approach.

[8] The conditions should be similar, but not identical to those one would have for two merging pipes carrying a compressible fluid.

possible if cell **Ek** can receive it—we will see in due course how the final result relates to the three possible states of causality.

Thus:

$$y_k(t) = S_{Bk} \text{ and } y_{ck}(t) = S_{Ck}, \quad \text{if } R_{Ek} > S_{Bk} + S_{Ck} \tag{7a}$$

If the condition in (7a) is not satisfied we will assume that the maximum number possible of vehicles, $R_{Ek}(t)$, advances into Ek. **As** long as the supply of vehicles from both approaches, $S_k(t)$ and $S_{Ck}(t)$, **is** not exhausted we will assume that a fraction (pk) of the vehicles come from Bk and the remainder (pck) from **Ck,** where $p_k + p_{ck} = 1$.

The constants "p" are characteristics of the intersection that capture any priorities. (Absolute priority to approach "k" for example, would imply that p_k=1 and p_{ck}=0). If the supply of vehicles on one of the approaches is exhausted before the end of the time interval between clock ticks, then the remaining vehicles to advance will come from its complementary approach.

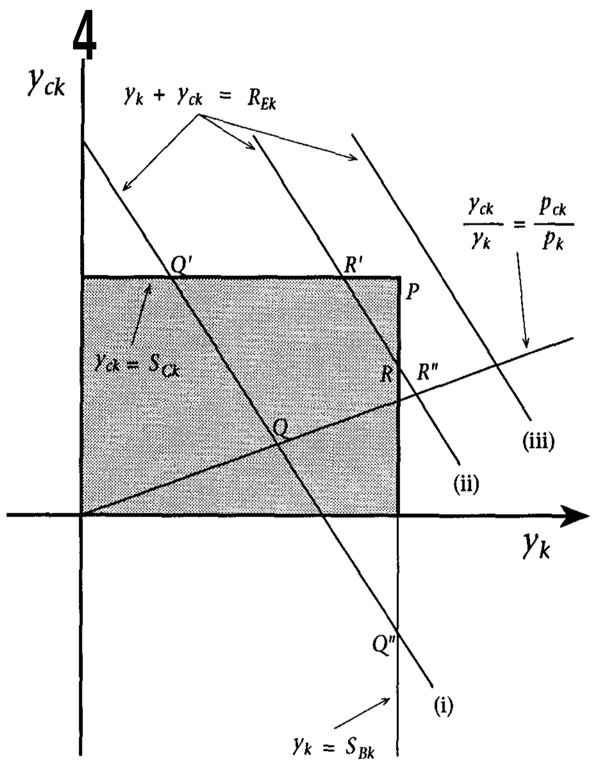


Figure **4:** Feasible **flow** diagram for a merge junction.

A simple formula for the advancing flows can be identified with the help of figure 4. The shaded rectangle in the figure represents the set of flows satisfymg equation (6a) and the (implicit) non-negativity restrictions. Three cases of equation (6b) are depicted; points below the downsloping line i (or ii, or iii) satisfy (6b). The figure also displays the line $y_{ck}/y_k = p_{ck}/p_k$, representing the flows that advance before the supply is exhausted.

Equation (7a) arises when the downsloping line does not intersect the rectangle (case iii) and the flows are the coordinates of point P. This corresponds to forward causality regime (a). When the upward and downward sloping lines intersect inside the rectangle (case i), a supply of vehicles will remain on both approaches when time expires and the solution flows are at the intersecting point (Q). This corresponds to backward causality regime (b). If the sloping lines intersect outside the rectangle but the downsloping line intersects the rectangle (case ii), then the solution must be on the downsloping line (since there is more demand than available room). Moreover, since the supply on one of the approaches is exhausted before time expires, the solution must also be on the side of the rectangle intersected by the upward sloping line; the point of intersection (R) is the solution. This corresponds to mixed causality regime (c).

We note that for both cases (i) and (ii), the solution point is the middle point of the three points intersected by the downsloping line (R, R′ and R″) or (Q, Q and Q). Thus we can write:

$$y_k(t) = \text{mid}\{ S_{Bk}, R_{Ek}\text{-}S_{Ck}, p_k R_{Ek}\} \quad \text{and} \tag{7b}$$

$$y_{ck}(t) = \text{mid}\{ S_{Ck}, R_{Ek}\text{-}S_{Bk}, p_{ck} R_{Ek}\}\ , \tag{7b}$$

$$\text{if } R_{Ek} < S_{Bk} + S_{Ck} \tag{7b}$$

Equations (7a) and (7b) — the generalization of (4) to merges — uniquely define the flows through a merge during time interval (t,t+1). These flows are then used with (5) to update the cell occupancies of the diverge.

A simulation for two highways that merge into one is easy to build using a computer spreadsheet consisting of two separate ranges—similar to the ones described in the appendix of Daganzo (1993a)—linked to a third range that contains the merge cell and equations (7). As in that reference, experiments with such a simulation confirm that the simulated behavior of the merge approximates the LWR model.

For causality regime (a), this can also be verified manually by inspecting the steady state solution to the difference equations (with very small clock ticks) for a network such as in figure 2a (consisting of two equilibrium link flows and three equilibrium cell occupancies), when the input flows to cells Bk and Ck are constant. The predictions also match if the combined input flows exceed the capacity of E_k so that a queue exists on one or both of the approaches; i.e. under causality regimes (b) and (c).

Nothing in our formulation prevents the "p" values, like the other cell characteristics (N, Q and δ), to depend on time. (Conceivably, p could also depend on the state of the end cell.) A time-dependent "Q" could be used to model ramp metering strategies, a time-dependent "N" to model temporary lane closures, and a time-dependent "p" to simulate traffic signals. For the latter, the priority constant, p_k, would alternate between 0 and 1 after a suitable number of clock ticks. Equations (7), however, need to be modified because a merge controlled by a traffic signal may not allow traffic from the blocked approach to enter the merge, even when the other approach is idle. If this is true the equations are even simpler:

$$y_k(t) = 0\ , \qquad \text{if } p_k(t) = 0 \tag{8}$$

$$= \min\{R_{Ek}, S_{Bk}\}, \qquad \text{if } p_k(t) = 1 \tag{8}$$

A variation on (8) would apply if the secondary flow is not totally interrupted when the high priority flow is low (e.g. as when right turns are allowed on red). Since equations (7 and 8) capture a broad set of conditions, we will not give here an extensive set of recipes for many forms of control; this may be attempted in the future.

Finally, note that if the p's are allowed to depend on the cell occupancies upstream of the merge, traffic actuated control strategies could also be simulated.

3.3 Model of a diverge

In deriving boundary conditions for diverges it should be recognized that the left/right turn percentages will in general depend on the mix of car destinations present in the element of flow currently upstream of the junction. This will be examined in section 4. Here we assume that the turning proportions are exogenously determined, as would occur for applications involving emergency evacuations (e.g. due to an imminent nuclear meltdown or natural disaster). In this case the logic is very simple.

In figure 2b, cell Bk can send a maximum of $S_{Bk}(t)$ vehicles during time interval (t,t+1) and cells Ek and Ck can receive a maximum of $R_{Ek}(t)$ and $R_{Ck}(t)$ respectively. These three quantities are still given by equations (3). Because part of $S_{Bk}(t)$ is destined for Ek and part for Ck, it will be assumed (as in Newell, 1993) that all the flow is restricted if either one of the diverging branches is unable to accommodate its allocation of flow. This assumes that vehicles unable to exit prevent all those behind, regardless of destination, to continue. This assumption, which should be suitable for the level of accuracy required of large scale models, implies that vehicles at the diverge (and indeed through the network) are served in a first-in-first out (FIFO) sequence.[9]

We assume here that the proportions of $S_{Bk}(t)$ going each way $\beta_{Ek}(t)$ and $\beta_{Ck}(t)$ ($\beta_{Ek}(t) + \beta_{Ck}(t) = 1$) are exogenously determined, and that traffic flows in these proportions continuously between clock ticks. Then, the (as yet unknown) number of vehicles emitted by Bk, $y_{Bk}(t)$, determines the turning flows:

$$y_k(t) = \beta_{Ek} y_{Bk} \quad \text{and} \quad y_{ck}(t) = \beta_{Ck} y_{Bk}. \tag{9a}$$

As in the other cases, the amount of traffic emitted by Bk, $y_{Bk}(t)$, should be as large as possible, without exceeding the amount that can be received by any of the exiting branches. This implies that yk must not exceed R_{Ek} and that y_{ck} must not exceed R_{Ck}. These conditions can be expressed mathematically as:

$$\max\{\, y_{Bk}(t) : y_{Bk}(t) \le S_{Bk} ,\ \beta_{Ek} y_{Bk}(t) \le R_{Ek} ,\ \beta_{Ck} y_{Bk}(t) \le R_{Ck} \,\}$$

---

[9] **In actuality, especially when the exit percentage is low, the blockage does not occur instantaneously. Furthermore, even after the conditions have settled into a stable pattern, freeway traffic is never at jam density next to the ramp. Observation of real systems reveals that the freeway traffic density upstream of a blocked diverge increases gradually toward the jam density in the u stream direction and that the density on the through lane(s) is less than on the exit lane(s). This allows &rough-vehicles to negotiate the diverge more rapidly than exiting-vehicles. A more detailed analysis of this phenomenon would require an extension of the LWR theory, involving two vehicle types. Such a theory could be useful for traffic engineering studies of freeway interchanges and weaving sections, but it is beyond the scope of this paper.**

The solution to this simple linear program is:

$$y_{Bk}(t) = \min\{ S_{Bk} , R_{Ek}/ß_{Ek} , R_{Ck}/ß_{Ck} \}, \tag{9b}$$

which together with (9a) defines the flows on a diverge.

As for ordinary and merge links, equations *(5)* complete the set of equations needed to update the state of the system.

4, THE MODEL WITH KNOWN ROUTES: TURNING PERCENTAGES NOT SPECIFIED

It is assumed in this section that the turning percentages at time t are not specified. Instead, they are derived from the destinations of the vehicles ready to advance into each diverge at time t, and information on the best paths available at the time to reach these destinations.

A FIFO discipline (consistent with the analysis in prior sections) will be used to identify the advancing vehicles. The best-path information is assumed to take the form of route choice constants defined for the two end cells of each diverge, $ß_{Ekd}(t)$ and $ß_{Ckd}(t)$. They give the proportion of vehicles with destination "d" that would advance from Bk to each of the two end cells in the time interval from t to t+1. **As** before, these constants must be between 0 and 1, and must satisfy: $ß_{Ekd}(t) + ß_{Ckd}(t) = 1$. We will discuss later how these constants can be calculated.

In order to use such route choice information within the model it is necessary to record cell occupancies, disaggregated by destination. In addition, in order to ensure the **FIFO** discipline when cells cannot emit all their contents, it will also be necessary to keep track of the time waited by the cell occupants. Thus, we characterize the state of the system by a variable, $n_{Id\tau}(t)$, representing the number of vehicles in cell I at time t bound for destination d that entered the cell in the time interval immediately after clock tick (t-τ). The parameter τ is a measure of time waited. Keeping track of τ is important because on arrival to a congested cell, vehicles with high τ must advance before those with a low τ.

Like cell occupancies, link flows must also be disaggregated by destination and wait. We use $y_{kd\tau}(t)$ to represent the number of vehicles flowing on link k during (t,t+1) that: (i) are bound for d, and (ii) had entered cell Bk in the τ^{th} interval before t.

4.1 The procedure with disaggregated occupancies and **flows**

As before, the overall algorithm will consist of a flow calculation step (i), and an occupancy revision step (ii).

With disaggregated variables, the expressions for the second step—previously equation (5)—should ensure that advancing vehicles have their τ reset to 1 while the rest have it increased by 1. The new formulas are:

$$n_{Ekd1}(t+1) = \Sigma_\tau \, y_{kd\tau} \qquad \text{for all } k,d \tag{10}$$

and

$$n_{Bkd\tau+1}(t+1) = n_{Bkd\tau} - y_{kd\tau,} \qquad \text{for all } k,d,\tau \tag{10}$$

In these equations, τ ranges from **1** to the number of time periods present in cell Bk at time t, $\tau_k(t)$.

Since the occupancies are disaggregated by destination and wait, the first step can be reduced to the identification of (real) variables $a_l(t)$ for every cell, denoting the minimum wait of the vehicles leaving each cell in interval (t,t+1). Under FIFO, the aI's readily indicate which vehicles are allowed to flow.

Accordingly, we assume that the disaggregated flows on a diverge link k are given by the following function of a_{Bk}:

$$y_{kd\tau}(t,a_{Bk}) = n_{Bkd\tau}ß_{Ekd} \quad , \quad \text{if } \tau > |a_{Bk}|+ \tag{11}$$

$$= (\tau - a_{Bk})n_{Bkd\tau}ß_{Ekd} \quad , \quad \text{if } \tau \quad |a_{Bk}|+ \tag{11}$$

$$= 0 \quad , \quad \text{if } \tau < |a_{Bk}|+, \tag{11}$$

where $|a_{Bk}|+$ denotes the smallest integer equal or greater than a_{Bk}. For ordinary or merge links, equation(11) applies with $ß_{Ekd} \equiv 1$.

Logically, equation (11) indicates that all the occupants of Bk that have entered cell Bk in a time interval prior to t-a,, (i.e. with $\tau > |a_{Bk}|+$) advance. Conversely, none of those with $\tau < |a_{Bk}|+$ advance. Of those vehicles having entered cell Bk in the time interval containing time $t-a_{Bk}$ only a fraction equal to the proportion of the interval that precedes aBk (i.e. τ - a,,) is advanced. This preserves the FIFO order to the accuracy allowed by the simulation clock tick[10].

The next subsection explains how the a, are calculated.

4.2 Minimum link waits

In order to determine the minimum link waits it is convenient to define two functions relating aggregated link flows to a,. If indexed by destination only, the link flow is:

$$y_{kd}(t,a_{Bk}) = \Sigma_\tau\, y_{kd\tau}(t,a_{Bk}). \tag{12}$$

The total aggregate flow is:

$$y_k(t,a_{Bk}) = \Sigma_d \Sigma_\tau\, y_{kd\tau}(t,a_{Bk}). \tag{13}$$

The inverse relationship between a_{Bk} and y_k will help us identify the a_l. **As** is illustrated in figure 5, expressions (11) to (13) define non-increasing, piecewise-linear functions with changes in slope at the integers. Because these functions are non-increasing, it is possible to define a non-increasing inverse function of (13), **A,,** which gives the minimum wait at the link's source cell, a_{Bk}, for a given aggregate flow, y_k:[11]

$$a_{Bk} = A_k(t,y_k) \tag{14}$$

[10] Although equation (11) ensures that vehicles entering a cell in time interval ô cannot leave the cell before those having entered in earlier intervals (τ+1, τ+2, **2+3** ...), it does not differentiate among the vehicles entering in the same clock interval. This departure from FIFO should not pose a serious problem because only a very small proportion of the total network flow is affected by this approximation. In any case, its effects can be minimized by choosing faster clocks.

[11] Relationship **(14)** will have a jump discontinuity whenever **(13)** is constant. In what follows it does not matter how aBk is valued at any such jump; we can assume for example that aBk is right-continuous, taking on the smallest possible value at each jump.

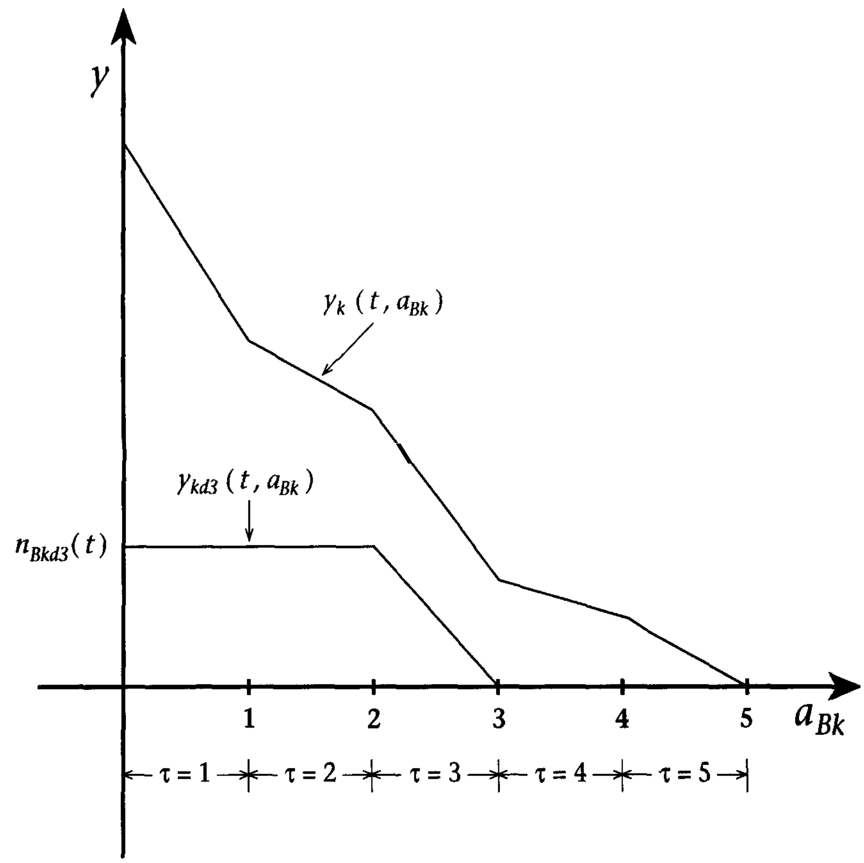


Figure 5: Graphical representation of $y_{kd\tau}\ (t, a_{Bk})$ **and** $y_k\ (t, a_{Bk})$.

Relationships **(13)** and **(14)** can be constructed concurrently in tabular form.

Relations **(13)** and **(14)** are only meaningful for a range of flows between zero and the maximum that can be sent, S_{Bk}. (For diverges, flows satisfy: $0 \le y_k(t,a_{Bk}) + y_{ck}(t,a_{Bk}) \le S_{Bk}$.) Accordingly, a„ ranges from the smallest value not violating the maximum flow constraint, which is denoted by $A_{Bk}(t)$ and may be zero, to the maximum wait present in the upstream cell, which is denoted by τ_{Bk}**(t)**. Note that for diverges $A_{Ck}(t) \equiv A_{Bk}$.

We are now in a position to identify aI(t) for the three basic types of cells.

Ordinarv and Merae Cells: Since only one link (k) leaves any such cell **(Bk),** these cells emit vehicles without a route choice. **As** a result, vehicle destinations do not influence the advancing flows and the emitted flow y_k can be obtained from the total cell occupancies as in section **3;** i.e. using (3), and either **(4)** for ordinary links or (7a) and (7b) for merge links. Then, **(14)** yields a_{Bk}.

Diverges: Although the aggregate flows at a diverge cannot be obtained directly from equation (9b), it is not difficult to generalize this result. As in section **3.3,** a diverge cell, B_k, will send as much flow as possible on both its branches, k and ck, provided the flows sent satisfy the capacity constraints for the sending cell and the two receiving cells. Thus, $a_{Bk}(t)$ should be the smallest value satisfymg:

$$y_k(t,a_{Bk}) \le R_{Ek}\ ,\ \ y_{ck}(t,a_{Bk}) \le R_{Ck}\ , \qquad \text{and}$$
$$y_k(t,a_{Bk}) + y_{ck}(t,a_{Bk}) \le S_{Bk}.$$

Because the left sides of these inequalities are non-increasing, and the last inequality is satisfied if and only if a_{Bk} ABk, the solution is:

$$a_{Bk} = \max\{\ A_k(t,R_{Ek}),\ A_{ck}(t,R_{Ck}),\ A_{Bk}\ \}. \tag{15}$$

With the a, known, an iteration of the procedure can now be completed.[12]

5. COMPUTATIONAL ISSUES

A computer program which carries out the procedure described in sections **3** and **4** for general networks has been written (Lin and Daganzo, 1993). Its results match satisfactorily the predictions of the LWR theory for special cases that can be solved by hand. Two such examples are reported in Daganzo and Lin (1993b); one of the examples illustrates the evolution of the traffic states when an incident in one of the branches of a diverge blocks traffic temporarily; the resulting queue spills onto the upstream freeway section and eventually dissipates.

As a point of reference for an evaluation of the computational complexity of our procedure we use the static traffic assignment problem. With careful definition of the data structures, the RAM memory needed to implement the procedure is a couple of times smaller than the product of the number of links L, the number of destinations D, and the average of $\tau_k(t)$, T: LDT. This can be compared with the requirements for the static traffic assignment problem, which are comparable to **AD,** where **A** is the number of arcs. Because A is perhaps an order of magnitude smaller than L, we see that the memory required by the cell-transmission model should be about ST times larger than that of the static traffic assignment. The factor T can be roughly viewed as the ratio of the average free-flow speed for the network (e.g. **64** kph) and the actual average at the most congested instant during the study period (e.g. 32 kph). For large networks T is not likely to be large, at most being comparable with 2 or 3. Thus, it appears that the memory requirements are about an order of magnitude larger than for an equivalent static traffic assignment problem.

The speed of execution with a single processor can also be estimated. Each clock tick requires a simple set of calculations to be done for every cell and every link. Because each link requires each destination and each τ to be considered, the number of calculations per tick should be on the order

[12] In addition to the cell occupancies, we should also update the maximum wait for each cell. Since by time t+l all the vehicles having entered Bk before t-$a_{Bk}(t)$ will have advanced, the maximum wait at that time by a remaining vehicle is 1+$a_{Bk}(t)$. Clearly then: $\tau_{Bk}(t+1) = |1+a_{Bk}|+$.

of LDT. For large networks, this compares well with the number of calculations for an iteration of the static equilibrium traffic assignment model, which is proportional to the number of the destinations but grows supralinearly with the number of nodes in network. On the other hand, we may need to perform more iterations than are normally done for static equilibrium models in order to simulate a rush hour period. All things considered, thus, the execution time may also be an order of magnitude greater than for static equilibrium models.

If this performance seems too cumbersome for application to large networks, the reader should recognize that the equations we have proposed are ideally suited for massively parallel computing. (Today these machines are approaching computation speeds of 1 teraflop, or 1012 floating point calculations per second, which could be used to simulate very large networks in the blink of an eye.) A large regional network can be decomposed into small subnetworks corresponding to different subareas of the study region; then, thanks to the order-free property of the cell transmission model, the algorithm can operate simultaneously and independently on the data for these subnetworks.

6. APPLICATIONS AND COMMENTS

The previous section assumed that the desired route between each origin and destination was known for all times. Thus, we were able to define turn constants $\beta_{Id}(t)$ for every downstream cell of a diverge. This is reasonable if we assume (as in Horowitz, 1984) that drivers respond to some weighted average of their past driving experiences and that such experiences determine the turn constants. The network model described in this paper could then be used to simulate a sequence of days to see if an equilibrium is reached. The model could also be used to determine the type of "real-time" information that is likely to be of most value to drivers. (Preliminary analyses with the model indicate, to not great surprise, that driver advisories should anticipate the projected evolution of the system, Daganzo and Lin, 1993a.)

Dynamic network models can also be applied to evaluate the performance of emergency evacuation plans developed in response to the possibility of a disaster such as a nuclear meltdown (Sheffi et al, 1982). Because in such an instance time is of the essence, a realistic model of network performance under a dynamic load is necessary. Evacuations are easier to study than the general dynamic traffic assignment model because traffic can be modeled as a single commodity flowing to a unique hypothetical destination, "safety".[13] An evacuation plan, consisting of a set of $\beta_I(t)$'s not differentiated by destination, could then be easily evaluated with the proposed approach.

7. Further work

Recent theoretical results (Daganzo, 1993d) indicate that the model in this paper can be extended to general equations of state with minor modifications. These new results would also allow larger cells (e.g. 1Km long) to be used on long homogeneous freeway segments and shorter

[13] The network would be represented as usual, stopping a sufficient distance from the location of the disaster. The nodes on the edge of the represented network would then be connected to the imaginary destination — the "safety" node — which all traffic would try to reach.

cells (e.g. 100 m long) on ramps and interchanges. The memory requirements of the models could then be reduced at the cost of some accuracy.

Efforts toward the development of more refined models of diverge behavior (see footnote 9) seem also worthwhile and are currently under way.

REFERENCES

Algadhi, S.A.H. and H. Mahmassani 1990. Modelling crowd behavior and movement: application to Makkah pilgrimmage. In Transportation and Traffic Theory (M. Koshi, editor), Proc 11th Int. Symp on Trans. and Traffic Theory, Yokohama, Japan, pp. 59-78, Elsevier, New York, N.Y.

Ansorge, R. 1990. What does the entropy condition mean in traffic flow theory? Trans Res 24B (2), 133-143.

del Castillo, J.M., P. Pintado, and F.G. Benitez 1993. A formulation for the reaction time of traffic flow models. In Transportation and Traffic Theory (C. Daganzo, editor), Proc 12th Int. Symp on Trans. and Traffic Theory, Berkeley, CA, pp. 387-405, Elsevier, New York, N.Y.

Chang, G.L., H. Mahmassani, and R. Herman 1985. Macroparticle traffic simulation model to investigate peak-period commuter decision dynamics. Trans. Res. Rec. 1005, 107-121.

Cremer, M., F. Meissner, and S. Schrieber 1993. On predictive control schemes in dynamic rerouting strategies. In Transportation and Traffic Theory (C. Daganzo, editor), Proc 12th Int. Symp on Trans. and Traffic Theory, Berkeley, CA, pp. 407-426, Elsevier, New York, N.Y.

Daganzo C.F. 1993a. The cell-transmission model. Part I: A simple dynamic representation of highway traffic. California PATH Research report, UCB-ITS-PRR-93-7, Inst. Trans. Studies, University California, Berkeley.

Daganzo C.F. 1993b. The cell-transmission model: A simple dynamic representation of highway traffic. Trans Res (in press)

Daganzo C.F. 1993c. Properties of link travel time functions under dynamic loads. California PATH Working paper, UCB-ITS-PWP-93-5, Inst. Trans. Studies, Univ. of California, Berkeley.

Daganzo, C.F. 1993d. A finite difference approximation for the kinematic wave model. Trans. Res. (submitted for publication).

Daganzo C.F. and W.H. Lin 1993a The spatial evolution of queues during the morning commute in a single corridor. California PATH Working paper, UCB-ITS-PWP-93-7, Inst. Trans. Studies, Univ. of California, Berkeley. (Presented at the 1994 meeting of the Trans. Res. Board).

Daganzo C.F. and W.H. Lin1993b. Moving queues in traffic networks. Proc. Eng. Foundation Conf. on Issues and Techniques in Traffic Management, St. Petersburg, FL. (In press).

Horowitz, J.L. 1984 The stability of stochastic equilibrium in a two-link transportation network. Trans. Res. 18B (1), 13-28.

Janson, B.N. and J. Robles 1993. Dynamic traffic assignment with arrival time costs. In Transportation and Traffic Theory (C. Daganzo, editor), Proc 12th Int. Symp on Trans. and Traffic Theory, Berkeley, CA, pp. 127-146, Elsevier, New York, N.Y.

Lax, P.D. 1954. Weak solutions of non-linear hyperbolic equations and theirnumerical computation. Comm. Pure App. Math. 7, 159-193.

Lighthill, M.J. and J.B. Whitham 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.

Lin, W.H. and C.F. Daganzo **1993.** Technical description of NETCELL: General framework and **data** structure. Inst. Trans. Studies, Univ. of California, Berkeley. (mimeo)

Luke J.C. **1972.** Mathematical models for landform evolution. J. Geoph. Res., **77, 2460-2464.**

Michalopoulos, P.G., D.E. Beskos, and Y. Yamauchi **1984.** Multilane traffic flow dynamics: some macroscopic considerations. Trans Res. **18B (4/5), 377-395.**

Michalopoulos, P.G., **P.** Yi, **A.S.** Lyrintzis **1993.** Continuum modelling of traffic dynamics for congested freeways. Trans. Res. 27B **(4), 315-332.**

Newell, G.F. **1961.** Non-linear effects in the dynamics of car following. Opns. Res. **9(2), 209-229.**

Newell, G.F. **1988.** Comments on traffic dynamics. Trans. Res. **23B(5), 386-389.**

Newell, G.F. **1993. A** simplified theory of kinematic waves in highway traffic, Part I: General theory; Part II: Queuing at freeway bottlenecks; Part 111: Multi-destination flows. Trans. Res. **27B (4), 281-314.**

Payne, **H.J.(1971.** Models of freeway traffic and control. Math. Models Publ. Sys., Simul. Council Proc. **28 (1), 51-61.**

Ran, B. **1993.** Dynamic transportation network models for advanced traveler information systems. PhD thesis, Univ. of Illinois, Chicago, Il.

Richards, P.I. **1956.** Shockwaves on the highway. Opns. Res. **4, 42-51.**

Sheffi, Y., H. Mahmassani, and W.B. Powell **1982. A** transportation network evacuation model. Trans Res. **16A (3), 209-218.**

Transportation Research **1991.** Special Issue on Dynamic Control of Urban Traffic Networks (Part II) (N.H. Gartner and G. Improta, eds.), Pergamon Press, New York, N.Y.