NETWORKS AND HETEROGENEOUS MEDIA
Website: http://aimSciences.org

Volume **1**, Number **4**, December **2006**
pp. **601–619**

# ON THE VARIATIONAL THEORY OF TRAFFIC FLOW: WELL-POSEDNESS, DUALITY AND APPLICATIONS

Carlos F. Daganzo

Department of Civil and Environmental Engineering
416 McLaughlin Hall
Berkeley, CA 94707, USA

Abstract. This paper describes some simplifications allowed by the variational theory of traffic flow(VT). It presents general conditions guaranteeing that the solution of a VT problem with bottlenecks exists, is unique and makes physical sense; i.e., that the problem is well-posed. The requirements for well-posedness are mild and met by practical applications. They are consistent with narrower results available for kinematic wave or Hamilton-Jacobi theories. The paper also describes some duality ideas relevant to these theories. Duality and VT are used to establish the equivalence of eight traffic models. Finally, the paper discusses how its ideas can be used to model networks of multi-lane traffic streams.

## 1. Introduction.

Consider an infinite one-directional road on which vehicles cannot pass and move in the direction of increasing distance, $x$. If at some location $x = 0$ we assign consecutive integers to the vehicles we observe as time increases from $-\infty$ to $+\infty$ then the space-time trajectories of all the vehicles are completely defined by the integer contours of a surface. The idea of such a surface was proposed in [29] and further elaborated in [26]. The surface is characterized by a continuous function with contour levels $n$, $N(t, x) = n$. The floor $\lfloor n \rfloor$ is the number of the last vehicle to have advanced beyond $x$ by time $t$. Since passing is not allowed the ordering of the vehicles is preserved everywhere. Therefore we can assume without loss of generality that $N(t, x)$ is non-decreasing in $t$ for every $x$. Moreover, since vehicles move in the direction of increasing $x$, we can also assume that $N$ is non-increasing in $x$ for every $t$.

The simplest model of traffic flow further assumes that $N$ is differentiable almost everywhere (except possibly along some curves that would form ridges in the surface defined by $N$) and that the first partial derivatives of $N$ are related by a function; i.e.:

$$\partial N/\partial t = Q(-\partial N/\partial x, t, x). \tag{1}$$

This is a Hamilton-Jacobi (HJ) equation with $Q$ as the Hamiltonian. Note that $\partial N/\partial t$ (abbreviated $q$) is the flow and $-\partial N/\partial x$ (abbreviated $k$) is the density, and that meaningful solutions require flow and density to be non-negative.

2000 *Mathematics Subject Classification.* Primary: 49L25, 90B20; Secondary: 70H20, 70H30 .
*Key words and phrases.* Traffic Flow Theory, Hamilton-Jacobi Equations, Variational Principles, Duality.
Work supported by the U.C. Berkeley Center for Future Urban Transport: a Volvo Foundations' International Center of Excellence.

The function $Q$ is called the "fundamental diagram" (FD) by traffic engineers. We assume in rough agreement with experiments that $Q$ is piecewise differentiable and concave in its first argument and returns non-negative values (for every $t$ and $x$) if the first argument is in an interval $[0, \kappa(t, x)]$ such that $Q(0, t, x) = Q(\kappa, t, x) = 0$, with $\kappa(t, x) < \infty$. The parameter $\kappa$ is called the "jam density." The maximum of $Q$, $q_{max}$, is called the "capacity"; see Fig. 1(a).

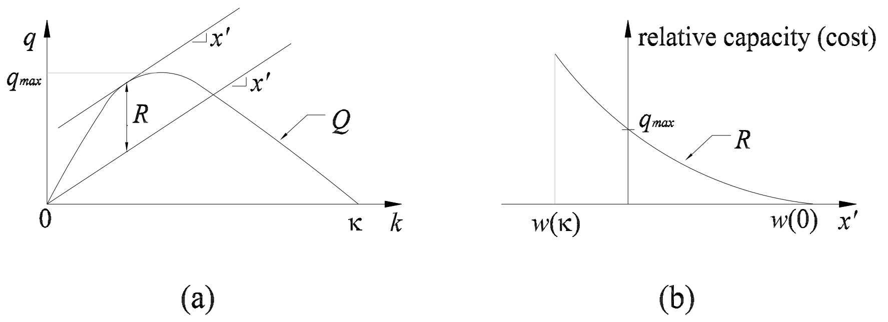


FIGURE 1. Basic concepts of Variational Theory: (a) fundamental diagram; and (b) cost function.

Note that if (1) is differentiated with respect to $x$ and expressed in terms of density it reduces to the conservation law: $\partial k/\partial t + (\partial Q/\partial k)(\partial k/\partial x) = -\partial Q/\partial x$. This is the classical kinematic wave (KW) formulation of [22, 35]. A simplified solution of some KW problems in terms of the Moskowitz function was later proposed in Newell's seminal trilogy [32]. Newell's results have been recently formalized and extended in [11, 12, 14]. These references propose a variational theory (VT) able to capture bottlenecks of all types.

1.1. **Variational theory.** Variational theory also assumes that (1) holds where $N$ is differentiable and that $Q$ is concave, but treats the problem as a capacity-constrained optimization problem. An intuitive explanation is as follows. We know that the flow at any point in space-time is bounded from above by $q_{max}$, the capacity. A similar capacity constraint should also hold if the road is viewed from a rigid frame of reference that moves with speed $x'$. In this case the capacity relative to the frame (the "relative capacity") is the maximum rate at which traffic can pass an observer attached to the frame. Since an observer that moves with speed $x'$ next to a traffic stream with density $k$ and flow $q$ is passed by traffic at a rate $q - kx'$, the FD for the moving frame is $Q(k, t, x) - kx'$ and its relative capacity is:

$$R(x', t, x) = \sup_k \{Q(k, t, x) - kx'\}. \tag{2}$$

Figure 1(a) shows these relations geometrically; note that the relative capacity $R$ is the intercept of the tangent to curve $Q$ with slope $x'$. Figure 1(b) shows the relative capacity function (also called the "cost function" in variational theory) with $x'$ as the argument. Note that $-R$ is the Legendre-Fenchel transform of $Q$, and that as a result, as shown by Fig. 1(b), $R$ is convex and strictly decreasing in the range of "valid" slopes where $Q$ is non-negative; i.e., for $x' \equiv w(k, t, x) \equiv \partial Q/\partial k \in$

$[w(\kappa,t,x), w(0,t,x)]$. The quantity $w(0,t,x)$ is called the "free-flow speed" and will be denoted $u(t,x)$. Note as well that $R \geq 0$, and $R(u) = 0$ since curve $Q$ touches the origin. Thus, in traffic flow the Legendre-Fenchel transform has an intuitive physical interpretation, which makes its application fairly intuitive as we shall now see.

Clearly, an observer traveling with a valid speed, i.e., with $x'(t) \in [w(\kappa), u]$, along a "valid" space-time path $\mathcal{P}$ from point D to point P cannot see a change in vehicle number greater than the integral with respect to time of its relative capacities; i.e., an upper bound to change is:

$$\Delta(\mathcal{P}) = \int_{t_D}^{t_P} R\left(x', t, x\right) dt, \tag{3}$$

where $t_D$ and $t_P$ are the times associated with the path endpoints. Therefore, an upper bound to the vehicle number $N_P$ observed at a point P can be written by considering the set $\boldsymbol{P}$ of all valid observer paths to P from the points of a boundary $\boldsymbol{D}$ where the vehicle numbers are known. In other words, if $\mathrm{D}(\mathcal{P}) \in \boldsymbol{D}$ is the beginning of a valid path $\mathcal{P}$, and $N_{D(\mathcal{P})}$ is the known vehicle number at $\mathrm{D}(\mathcal{P})$, then it must be true that $N_P$ must satisfy:

$$N_P \leq \inf\{N_{D(\mathcal{P})} + \Delta(\mathcal{P}) \mid \mathcal{P} \in \boldsymbol{P}\}. \tag{4}$$

Equation (4) is the capacity constraint mentioned at the outset.

In variational theory the solution domain $\boldsymbol{S}$ is the set of points P such that all infinitely long valid paths ending at P intersect the boundary. For example, the solution domain for the initial value problem (IVP) is the half plane, $t > 0$. Variational theory assumes that capacity constraint (4) is binding; i.e., that the actual value of $N_P$ for $\mathrm{P} \in \boldsymbol{S}$ is the largest possible allowed by (4):

$$N_P = \inf\{N_{D(\mathcal{P})} + \Delta(\mathcal{P}) \mid \mathcal{P} \in \boldsymbol{P}\}. \tag{5}$$

This is a calculus of variations problem. It is well known, see e.g., [2, 3], that under some regularity conditions (5) characterizes both the viscosity solution of the HJ-IVP and also the entropy solution of the KW-IVP. (If $Q$ is not concave (4) continues to be true but (5) may not match the HJ and the KW solutions because other constraints come into play.) A key advantage of VT over the HJ and KW theories is its natural framework for expressing the relatively complicated problems arising in traffic flow applications (including bottlenecks and finite roads), and the convenient way in which the "well-posedness" of such problems can be assessed; see below.

## 2. Simplifications.

### 2.1. Homogeneous problems with point bottlenecks: solution existence and uniqueness.

In traffic flow theory it is often necessary to consider "point bottlenecks". These are usually slower vehicles or fixed obstructions that reduce the maximum rate at which traffic can flow past them. A point bottleneck is defined by its space-time trajectory $x_B(t)$, assumed to be a valid path, and by its relative capacity (maximum passing rate) $r_B(t)$. In HJ theory, the relative capacity restriction is the following upper bound on the total derivative of $N$ along the bottleneck trajectory: $dN(t, x_B(t))/dt \leq r_B(t)$. This type of constraint seems not to have received much attention in the mathematics literature. The constraint is even more complicated when expressed in terms of KW theory. But the complication disappears in VT.

In VT a bottleneck reduces the original relative capacity of the road along the bottleneck trajectory. This is recognized by using $r_B(t)$ instead of $R$ as the integrand in (3) for the portion of any path that overlaps $x_B(t)$. Nothing else needs to be changed: (4) and (5) continue to apply. Hence, in VT, point bottlenecks are just shortcuts through space-time, which preserve the shortest-path character of the problem without increasing its complexity. Hence, a solution with bottlenecks is as easy to find as one without bottlenecks. The question is whether the solution with bottlenecks is continuous and varies with $t$ and with $x$ at allowable rates.

We assume that the boundary data satisfy the following Lipschitz conditions in $\boldsymbol{D}$, and look for solutions that satisfy them in $\boldsymbol{D} \cup \boldsymbol{S}$:

$$[N(t, x_1) - N(t, x_2)]/[x_2 - x_1] \in [0, \kappa] \qquad \text{if } x_1 < x_2, \tag{6}$$

$$[N(t_2, x) - N(t_1, x)]/[t_2 - t_1] \in [0, q_{max}] \qquad \text{if } t_1 < t_2. \tag{7}$$

A solution satisfying (6-7) in $\boldsymbol{D} \cup \boldsymbol{S}$ is obviously continuous; thus, vehicles have continuous trajectories. Furthermore, if (6-7) hold, vehicles can neither reverse direction nor overtake an object moving with free flow speed $u$; i.e. their average speed is always bounded in a physically meaningful way.* Therefore, solutions satisfying (6-7) will be called "valid". A VT problem whose solution is valid will be said to be "well-posed".

We examine below whether these conditions are satisfied for "homogeneous highway problems"; i.e, problems in which $Q$ and $R$ are time-independent and space-independent. Therefore, they will be expressed from now on as functions of one argument, $Q(k)$ and $R(x')$, respectively; the parameters $\kappa, q_{max}$, etc. become constants. It will be useful to keep in mind that for homogeneous highway problems without bottlenecks straight lines turn out to be optimum paths and the RHS of (3) reduces to $R(v_{DP})(t_P - t_D)$, where $v_{DP}$ is the slope of segment DP: $v_{DP} = (x_P - x_D)/(t_P - t_D)$; see [12]. In this special case, thus, the calculus of variations problem (5) reduces to an ordinary minimization for the point on the boundary (D $\in \boldsymbol{D}$) that produces the minimum cost. This is the well-known Lax integral formula for conservation laws [17]. We can now state the following.

**Theorem 1.** *A VT-IVP with any number of piecewise differentiable bottlenecks "B" is well-posed if the initial data satisfy (6) and the bottlenecks satisfy:* $r_B(t) \geq 0$ *and* $dx_B(t)/dt \geq 0$.

*Proof.* see Appendix A (Lemmas 5 and 6). □

Lemmas 5a and 6 of Appendix A (summarized as Theorem 2, below) prove a similar result for the finite highway problem (FHP) with bottlenecks. The highway extends from $x = 0$ to $x = x^o > 0$. Given are the vehicle numbers along the boundary: $N(0, x)$ for $0 \leq x \leq x^o$, and $N(t, 0)$ and $N(t, x^o)$ for $t \geq 0$. We look for the least costs $N(t, x)$ in the solution domain: $\boldsymbol{S} = \{(t, x) \mid t > 0 \text{ and } 0 < x < x^o\}$.

*This should be clear: (i) direction reversals cannot occur because otherwise $N$ could take the same value with increasing $t$ and decreasing $x$, which is incompatible with (6-7); and (ii) vehicles cannot overtake an observer traveling with free-flow speed $u$ everywhere because this would imply that $N$ increases along the observer's path, which is incompatible with the requirement $R(u) = 0$.

Because the boundary at $x = 0$ and $x = x^o$ can be reached by valid paths from the boundary, we add the necessary consistency condition for well-posedness stipulated in [14] for problems with complex boundaries: namely, that the least cost of reaching a boundary point with a valid path from the boundary starting at an earlier time, $n(t, 0)$ or $n(t, x^o)$, be no less than the cost specified for that point. For the FHP this necessary condition is: $N(t, x^b) \leq n(t, x^b)$ for $t > 0$, where $x^b = 0$ or $x^o$.

**Theorem 2.** *An FHP with bottlenecks is well posed if:* (*i*) *the* $N(t, x^b)$ *are non-decreasing in t, and* $N(0, x)$ *satisfies* (6)*;* (*ii*) *the bottlenecks satisfy* $r_B(t) \geq 0$, $dx_B(t)/dt \geq 0$ *and* $x_B(t) \geq 0$*; and* (*iii*) *the consistency condition is met:* $N(t, x^b) \leq n(t, x^b)$ *for* $t > 0$.

In applications, Theorem 2 can be expressed in terms of a competition between "upstream demand" and "available downstream capacity". Let $U(t)$ be a real function giving the cumulative upstream demand at $x = 0$ over time; i.e., the desired entrances. We assume that $U(t)$ is non-decreasing and $U(0) = N(0, 0)$. Recall that $n(t, 0)$ is the infimum of the costs of reaching point $(t, 0)$ from the boundary with valid paths in the solution domain starting at an earlier time; i.e., starting from $\boldsymbol{D}_t = \{(t', x) \mid (t', x) \in \boldsymbol{D};\ t' < t\}$ and ending at $E(\mathcal{P}) = (t, 0)$. Because the starting points are at $x \geq 0$ we call this infimum the available *downstream* capacity and introduce the superscript $D$ to emphasize the idea; i.e.: $n^D(t, 0) \equiv n(t, 0) = \inf\{N_{D(\mathcal{P})} + \Delta(\mathcal{P}) \mid \mathcal{P} \in \boldsymbol{D}_t;\ E(\mathcal{P}) = (t, 0)\}$.

Likewise, let $C(t)$ be a real function giving an upper bound on the cumulative number of exits at $x^o$; an available downstream capacity at $x^o$. We assume that $C(t)$ is non-decreasing and $C(0) = N(0, x^o)$. We also define the *upstream* demand at $x = x^o$, $n^U(t, x^o) \equiv n(t, x^o)$, as the infimum of the costs of reaching point $(t, x^o)$ from the boundary with valid paths in the solution domain starting from $\boldsymbol{D}_t$; i.e., $n^U(t, x^o) = \inf\{N_{D(\mathcal{P})} + \Delta(\mathcal{P}) \mid \mathcal{P} \in \boldsymbol{D}_t;\ E(\mathcal{P}) = (t, x^o)\}$. Then, Lemma 7 of Appendix A establishes the following:

**Theorem 3.** *An FHP with bottlenecks is well posed if:* (*i*) *the initial data* $N(0, x)$ *satisfy* (6)*;* (*ii*) *the bottlenecks satisfy* $r_B(t) \geq 0$, $dx_B(t)/dt \geq 0$ *and* $x_B(t) \geq 0$*; and* (*iii*) *the upstream and downstream data are given by* $N(t, 0) = \min\{U(t), n^D(t, 0)\}$ *and* $N(t, x^o) = \min\{C(t), n^U(t, x^o)\}$*, where* $U(t)$ *and* $C(t)$ *are non-decreasing functions such that* $U(0) = N(0, 0)$ *and* $C(0) = N(0, x^o)$.

The functions $U(t)$ and $C(t)$ can be chosen in any way. For example, if there is a highway with bottlenecks and different $Q$ and $R$ for $x < 0$, we can define $U(t)$ for the downstream problem as the demand at $x = 0$ arising from the upstream problem, $n^U(t, 0)$; and choose $C(t)$ for the upstream problem as the available capacity arising from the downstream problem at $x = 0$, $n^D(t, 0)$. To stitch together the two solutions we simply stipulate $N(t, 0) = \min\{n^U(t, 0), n^D(t, 0)\}$ for both problems. This is a natural and obviously well-posed way to treat inhomogeneous highways.

The Theorems are consistent with existing results of KW and HJ theories for problems without bottlenecks in, both, unbounded [34, 6] and bounded [4, 1] domains. The results also generalize the demand vs. capacity framework of the cell-transmission model (CTM) for networks [9] and the related results for single links in [7, 8, 20]. The demand-capacity framework allows one to construct well-posed network models such as the CTM, even if there are point bottlenecks; see Sec. 5.

The results can also be applied to time-dependent problems. Well-posedness can be checked in this case by slicing the solution space into successive time-independent problems and verifying that each time-independent slice satisfies the conditions of one of the above theorems. Unfortunately, well-posedness cannot always be tested *a priori* (before solving the problem) as in the time-independent case because for the initial data of a slice to be valid (and the theorems to hold) the solution obtained at the end of the previous slice must satisfy (6) with the jam density $\kappa$ specified a priori for the current slice.

2.2. **Linear cost functions.** In this subsection $Q$ is triangular in $k$. Now, the problem simplifies even more because the cost function (3) is linear [12]. If we use $u \equiv w(0)$ and $-w \equiv w(\kappa)$ for the slopes of the rising and dropping branches of $Q$ (in traffic flow lingo $w$ is called the "backward wave speed"), then (2) becomes:

$$R(x') = (1 - x'/u)q_{max} \qquad \text{for} \quad x' \in [-w, u]. \tag{8}$$

Note that $R(x')$ decreases. We shall abbreviate its maximum by the symbol $r$: $r = R(-w) = (1 + w/u)q_{max},$. This parameter (the maximum relative capacity) will be useful later. Experiments show that $r$ is about 15% greater than $q_{max}$.

This case is so simple because when $R$ is linear the path cost (3) is just a linear function of the path's duration and distance; i.e., if $\mathcal{P}$ goes from D to P:

$$\Delta(\mathcal{P}) = q_{max}(t_P - t_D) - (q_{max}/u)(x_P - x_D). \tag{9}$$

Hence, if as is often the case in traffic flow the boundary data (i.e., the coordinates $t_D$ and $x_D$ and the values $N_D$ for all points D$\in\boldsymbol{D}$) are given as piecewise linear functions of a parameter, $t_D = t(\alpha)$, $x_D = x(\alpha)$ and $N_D = N(\alpha)$, then (5) becomes:

$$N_P = \inf_{\alpha}\{q_{max}[t_P - t(\alpha)] - (q_{max}/u)[x_P - x(\alpha)] + N(\alpha)\}, \tag{10}$$

which is just the minimization of a piecewise linear function. Obviously, we can find its minimum by inspecting the corners of the objective function.

The solution can also be found with network algorithms; see e.g., [12]. These methods are advantageous when the solution is sought at many points in the solution domain. The networks in question are digraphs with nodes $L$ embedded in space-time, with directed arcs *LL'*. Arcs are defined only for node pairs that can be connected by a valid path. We call these "valid node pairs." Each arc is assigned a cost, $c_{LL'}$, equal to that of an optimum continuum path between its end nodes; e.g., as given by (9) when $Q$ is triangular. Of interest are networks whose shortest "walks" (network paths) between all valid node pairs are shortest continuum paths. These networks are said to be "sufficient" because by solving the shortest path problem on the network one solves the continuum problem exactly for all its valid node pairs. This is useful because if one puts nodes of a sufficient network on the corners of a piecewise linear boundary, then the network solution identifies the exact $N$ at every node. The solution can then be found with the usual dynamic programming recursion:

$$c_{L'} = \min_{L \in F(L')} \{c_L + c_{LL'}\}, \tag{11}$$

where $F(L')$ is the set of "from" nodes of $L'$.

Sparse sufficient networks with as few as two links per node can be constructed for problems with linear $R$; thus (11) can be computed fast. We will use in this paper sufficient networks of the "lopsided" type defined in [12].† A lopsided network (see

†Lopsided networks are sufficient for problems without bottlenecks and can be easily modified to retain sufficiency if there are bottlenecks; see [27].

Fig. 2) is a network with the following properties: (i) its nodes are on a rectangular lattice with space separation $\delta$ and time step $\varepsilon$, (ii) the set of links pointing to any node is translationally symmetric, (iii) this set contains two links with slopes $u$ and $-w$, and (iv) no link spans a distance greater than $\delta$. Note: since the nodes are on a rectangular lattice, $\delta/u$ and $\delta/(-w)$ must be integer multiples of $\varepsilon$, assuming $u$, $w \neq 0$. These networks will help us compare different ways of finding $N$. But before this is done we introduce some duality ideas, which will allow us to double the number of models covered under the same umbrella.

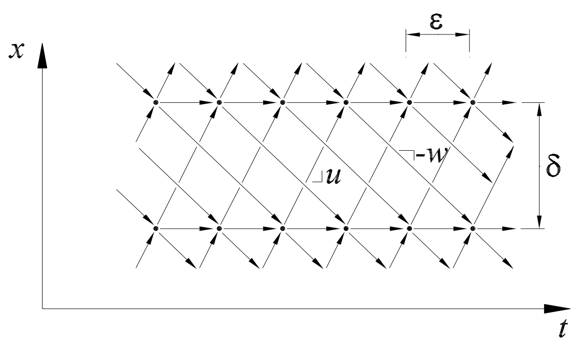


**Figure 2**: A lopsided network.

3. **Duality.** In this section $Q(k)$ is concave – not necessarily triangular. We shall show that associated with every HJ-VT problem satisfying a mild monotonicity condition there is a dual problem obtained by interchanging the $n$ and $x$ variables, and that the dual problem is also an HJ-VT problem that can be solved with the same methods. The primal and dual solutions describe the same Moskowitz surface. The results apply to problems where $N(t, x)$ strictly decreases with $x$ for every $t$ in the relevant solution domain. Lemma 5b of Appendix A shows that an IVP with strictly decreasing $N(0, x)$ satisfies this condition if it has no bottlenecks; and also if there are bottlenecks but the solution is only sought upstream of them.[‡] The monotonicity requirement on $N(0, x)$ is innocuous in practice because if there is at least 1 vehicle on the road at $t = 0$ then there is a strictly decreasing $N(0, x)$ that describes the vehice positions.

Since $N$ is continuous and declines with $x$, the relation $N(t, x) = n$ defines an implicit function for $x$ in terms of $t$ and $n$, $x = X(t, n)$. This function gives the position of vehicle $n$ at time $t$. It is also continuous and declines with $n$. Both functions describe the same Moskowitz surface. The two functions are connected by the functional relation:

$$X(t, N(t, x)) = x, \tag{12}$$

which merely expresses that the position at time $t$ of the vehicle that was at $x$ at time $t$ must be $x$. Conversely, we can also write:

$$N(t, X(t, n)) = n, \tag{13}$$

[‡]Solutions upstream of bottlenecks can be used to describe the behavior of car platoons following slow vehicles, e.g., in multi-lane traffic models with lane changes [18]; see section 5.

since the vehicle number found at time $t$ at the position of vehicle $n$ at time $t$ is $n$. Note that (13) is obtained from (12), and vice versa, by interchanging $(x, X)$ with $(n, N)$. Since the (primal) results of Secs 1 and 2 were derived with $N$ as the unknown, this symmetry suggests that similar (dual) results could be derived with $X$ as the unknown after swapping the variables and functions for position and vehicle number.

Differentiation of (12) with respect to $t$ and $x$ yields the following relation among the partial derivatives of the primal and dual functions:

$$\partial N/\partial x = 1/(\partial X/\partial n) < 0, \qquad \text{and} \tag{14}$$

$$\partial N/\partial t = -(\partial X/\partial t)/(\partial X/\partial n) \geq 0. \tag{15}$$

The same expressions with $(x, X)$ and $(n, N)$ interchanged are obtained if one differentiates (13). The quantity $v = \partial X/\partial t$ is the vehicle speed, and the quantity $s = -\partial X/\partial n$ the reciprocal of density; i.e., the continuum version of vehicular spacing. If we now insert (14) and (15) into (1) we find:

$$\partial X/\partial t = V(-\partial X/\partial n), \tag{16}$$

where $V$is related to $Q$ by the following transformation:[§]

$$V(s) = Q(1/s)/s, \qquad \text{where } \ s = -\partial X/\partial n. \tag{17}$$

Equation (16), like (1), is a HJ equation. Note as well that the transformation (12-13) is a reflection, which preserves stability; thus, if we take a stable (viscosity) solution of (1) and transform it with (12-13) the result is a stable solution of (16). The reverse is also true. Thus, for any given set of boundary conditions $[t_D = t(\alpha)$, $x_D = x(\alpha)$ and $N_D = N(\alpha)]$ the stable solutions of (16) and (1) describe the same Moskowitz surface. It can be found by solving either a primal problem (1) or a dual problem (16).[¶] Furthermore, since $V$, like $Q$, is concave in the relevant range of its argument, $s \in [1/\kappa, \infty)$, VT can be used with the dual problem too.

The dual cost function $R^d$ is given by (2) with $V$ substituted for $Q$. We find that $R^d$ is the inverse of $R$, with the roles of speed $x'$ and passing rate $n'$ reversed, and that it still is convex and decreasing in the relevant range of passing rates. In the triangular case the dual cost function is the inverse of (8), and is still linear:

$$R^d(n') = (1 - n'/q_{max})u, \qquad \text{for } n' \in [0, r]. \tag{18}$$

Therefore, the sufficient lopsided networks that one can use with (11) now have slopes equal to the cost rates of the primal (0 and $r$) and cost rates equal to the slopes of the primal ($u$ and $-w$, respectively). Variational theory in its primal and dual forms is used in the next section to examine the connection between eight different traffic models.

---

[§]Note that the transformation $Q \rightarrow V$ is an involution, which should not be surprising since the swap of $x$ and $n$ is a reflection.

[¶]Note too that one can define by differentiating (16) with respect to $n$ a conservation law, $\partial s/\partial t$ + $(\partial V(s)/\partial s)$ $\partial s/\partial n = 0$, which is the dual of $\partial k/\partial t + (\partial Q(k)/\partial k)\partial k/\partial x = 0$. The analyses and methods relevant to the primal conservation law also apply to the dual.

4. **Application: eight traffic models.** In this section the highway is homogeneous and the FD is triangular. We classify traffic models into 4 categories distinguished by the number of variables that are treated discretely: 0-*models* treat all variables continuously, as in the discussion up to this point; 1-*models* treat the non-time independent variable ($x$ or $n$) discretely; 2-*models* treat both independent variables ($t$ and $x$; or $t$ and $n$) discretely; and 3-*models* treat all variables ($t$, $x$ and $n$) discretely. Here we present a primal and dual VT model of each type (eight models in total) and see how they relate to existing ones.

**0-models:** These are fluid models. Our primal 0-model is (3 and 5). We have already seen that it has the following dual 0-model, where $\mathcal{P}$ is a dual path $n(t)$ from $(t_D, n_D)$ to $(t_P, n_P)$:

$$X_P = \sup\{\Delta(\mathcal{P}) + X_{D(\mathcal{P})}\}, \tag{19}$$

where $X_{D(\mathcal{P})}$ is the vehicle position at the beginning of path $\mathcal{P}$ and $\Delta(\mathcal{P}) = \int_{t_D}^{t_P} R^d(n')\,dt = u(t_P - t_D) - (u/q_{max})(N_P - N_D)$. Note the similarity of this formula for $\Delta(\mathcal{P})$ and (9).

**1-models:** These are queuing and car-following models. An example of a primal 1-model is Newell's queuing formula [32] which gives the cumulative curve at some point of a highway $N(t, x_M)$ from the vehicle number curves observed at its upstream and downstream ends: $N_U(t)$ and $N_D(t)$. The formula is:

$$\begin{aligned} N(t, x_M) = \min\{&N_U(t - (x_M - x_U)/u), \\ &N_D(t - (x_D - x_M)/w) + (x_D - x_M)r\}. \end{aligned} \tag{20}$$

The reader can verify that (20) is the result of applying (10) to our boundary data.[‖]

We now apply (19) to a "lead vehicle problem". This is a dual problem with boundary conditions: $X(t, 0) = x_0(t)$ for $t \geq 0$ and $X(0, n) = x_n(0)$ for $n \geq 0$. Assume that $x_n(0)$ is linear in $n$ (vehicles are uniformly spaced) and that $dx_0(t)/dt \leq u$. Then, an optimum path to reach point $(t, n)$ for some integer $n$ must begin at one of the two extreme points of the relevant part of the boundary for point $(t, n)$: either point $(0, n)$ or point $(t - n/r, 0)$. The result is:

$$X(t, n) = \min\{x_n(0) + ut,\ x_0(t - n/r) - nw/r\},$$

which is the trajectory of vehicle $n$. Note that $w/r \equiv 1/\kappa$ is the "jam spacing", which we shall denote $s_j$. The parameter $1/r$ (comparable with 1 second in practical applications) has the interpretation of a reaction time and will be denoted by $\tau$. In practice we are usually interested in the values of $X$ for all integer $n$. A recursive expression is obtained by setting $n = 1$ in the above and applying the same recipe to all consecutive vehicle pairs. The result is the car-following law in [33]:

$$X(t, n) \equiv x_n(t) = \min\{x_n(0) + ut,\ x_{n-1}(t - \tau) - s_j\}. \tag{21}$$

[‖]Since $N_U$ and $N_D$ cannot increase at a rate that exceeds $q_{max}$, an optimum path to point "P" must emanate from a point on the (upstream or downstream) boundary with the highest possible $t$. Only two such points generate valid paths. They correspond to the two arguments of (20).

**2-models:** These are numerical analysis schemes for fluid models. For the primal we use (11) with a lopsided network with two links per node. We choose $\delta = u\varepsilon$. Therefore, the links with slope $u$ (and zero cost) span one time step. The links with slope $-w$ (and cost rate $r$) span $\theta = u/w$ time steps. Hence their cost is $c = r\theta\varepsilon = r\delta/w$. Note: since $\theta$ must be integer we are assuming that $u/w$ is an integer — this ratio is comparable with 6 in practice. If we now use sub-indices $l$ and $m$ to identify the time and distance steps, i.e., so that $N_{lm} \equiv N(l\varepsilon, m\delta)$, (11) becomes:

$$N_{lm} = \min\{N_{l-1,m-1},\ N_{l-\theta,m+1} + c\}. \tag{22}$$

Equation (22) expresses the ACT (asynchronous cell transmission) model for cells of size $\delta$. Appendix B derives the ACT formula — equation (B3) — and explains its connection to conventional numerical schemes for conservation laws.**

A dual 2-model is obtained by applying (11) to a lopsided network on the $(t, n)$ plane as described above with arc slopes $(0, r)$ and arc cost rates $(u, -w)$. We choose the step for variable $n$ to be 1 and the time step, $\varepsilon = 1/r = \tau$. This achieves a rectangular lattice, since $\varepsilon r = 1$. The link costs become as a result: $u\tau$ and $-w\tau = -s_j$. Therefore, with the convention: $X_{lm} \equiv X(l\tau, m)$, recursion (11) reduces to:

$$X_{lm} = \min\{X_{l-1,m} + u\tau,\ X_{l-1,m-1} - s_j\}. \tag{23}$$

This is the CF(L) model in [15], which merely expresses (21) on a lattice.

**3-models:** Examples of 3-models are cellular automata (CA) models, where cars are assumed to jump on a lattice. Most CA models are described in dual space, but as we now show primal models can also be derived. Simply, use $\delta = s_j \equiv w/r$ in (22), which yields $c = 1$, and therefore:

$$N_{lm} = \min\{N_{l-1,m-1},\ N_{l-\theta,m+1} + 1\}. \tag{24}$$

This expression returns an integer if the input vehicle numbers are integer. Therefore it is a CA model. The expression indicates that the vehicle count at a point in space is the smaller of either the upstream count in the prior time step, or the downstream count $\theta$ time steps earlier plus 1. That is, a vehicle jumps from $m-1$ to $m$ if and only if the previous vehicle jumped from $m$ to $m+1$ at least $\theta$ time steps ago. This is the CA(M) rule in [15].

Consider now the dual formula (23) and express it in dimensionless distance, $Z = X/s_j$. It becomes:

$$\begin{aligned} Z_{lm} &= \min\{Z_{l-1,m} + u\tau/s_j,\ Z_{l-1,m-1} - 1\} \\ &= \min\{Z_{l-1,m} + \theta,\ Z_{l-1,m-1} - 1\}. \end{aligned} \tag{25}$$

We see that if vehicles are initially on the lattice (the $Z$'s are integer) and if $\theta$ is an integer, then (25) keeps vehicles on the lattice; i.e., it is a CA model. Equation (25) is the unbounded acceleration special case of the model model in [31], called the CA(L) model in [15].

**The middle term of (B3) turns out to be redundant for the homogeneous highway problem. But if we had used a lopsided network with one horizontal link of cost $\varepsilon q_{max}$, (22) would have included the middle term of (B3).

This concludes our review. Duality and variational theory provided a framework that clearly established the equivalence of eight traffic models. The best model for any given application depends on the form of the data and the requirements of the output.

5. **Composition into networks and discussion.** Primal analysis looks for the flow of vehicles from the perspective of the road; and dual analysis for the "flow of road" from the perspective of the vehicles.[††] Fixed bottlenecks such as merges and lane-drops are understood by scientists in primal space, from the perspective of the road, since this is the form in which data are available. Moving bottlenecks; e.g., those caused by slow-moving obstructions are understood from the perspective of the moving bottleneck, since data from this perspective is available. The ideas in this paper allow us to combine the effects of fixed and moving bottlenecks consistently in whatever framework is most useful (primal or dual) for practical application. Multi-lane traffic streams and networks are application domains of particular interest.

**Multi-lane streams:** A vehicle that has changed lanes from a slow to a faster lane but has not yet fully accelerated acts on that lane as a moving bottleneck with a zero passing rate. Conversely, a lane changer to a slower stream creates a momentary bottleneck on the source lane. Our ability to treat moving bottlenecks suggests that multi-lane traffic streams can be modeled as hybrid systems of (continuous) single-lane VT streams linked by (discrete) lane changes. This is mathematically trivial if the trajectories of the moving bottlenecks are given, but in practice they are unknown and have to be generated endogenously as the problem is solved.

Reference [18] proposes generating these trajectories with a probabilistic demand model that initiates the lane changes and a constrained acceleration/deceleration model that creates their discrete paths. Although well-posedness conditions for hybrid problems with endogenous linkages have not been rigorously established, the approach in [18] seems to be workable in practice. Tests in [18, 19] showed that the approach was surprisingly accurate predicting the performance of lane drops, merges and moving bottlenecks. A variant of this method has been found to reproduce HOV lane phenomena [28].

**Networks:** Although composition of links that follow fluid models is in general complicated, the competition paradigm between capacity and demand discussed in section 2 simplifies matters for KW-VT links. This paradigm was exploited in the composition rules for merges and diverges of the CTM [9], and its implementation

[††]A possible interpretation of dual VT and its constraints is as follows. Imagine uniformly spaced parked (dual) vehicles by the side of the road. Then, dual VT describes the flow of these vehicles from the perspective of a flexible frame of reference attached to the moving (primal) vehicles; i.e., where (dual) distance increases by a unit with each (primal) moving vehicle. From this frame of reference, the (dual) flow is the rate at which dual vehicles (i.e., units of primal distance) flow past fixed positions in the dual frame (i.e., moving-primal vehicles). Thus, dual flow = primal speed. Conversely, the rate at which a dual vehicle overcomes dual distance (i.e., moving vehicles) is both the dual speed and the primal flow. And the number of parked vehicles between two consecutive moving vehicles is both the dual density and, the primal spacing. Thus, dual-VT can also be interpreted in terms of flows and densities, and its constraints described in terms of relative capacities, but all from the perspective of the flexible frame of reference. Thus, the dual relative capacity is the maximum flow of parked vehicles that can be seen by an observer jumping from primal vehicle to primal vehicle with a fixed jump frequency.

in [23]. Later, [8, 5] extended the rules to networks with more general (unimodal) FD's, and [21] showed their connection with the Godunov numerical method.

The CTM composition rules have been used to describe the time-dependent behavior of traffic circles and ring roads [10], shedding light on the so-called "gridlock effect", and to investigate network control strategies [24, 25, 36]. (An animation of the gridlock effect can be found in *http://www.its.berkeley.edu/volvocenter/gridlock/*). Unfortunately the CTM rules and those of other fluid models use exogenous parameters such as "priority constants" for merges, which depend on the geometry of an intersection but cannot be estimated without observing traffic. Thus, conventional fluid models cannot always predict the effect of changes in intersection design.

Multi-lane hybrid models allow us to side-step this problem and increase realism. For example, as demonstrated in [19], we can treat the two links leading to a merge (or the two emanating from a diverge) as a single link with as many lanes as the two links combined, provided that lane changes are banned across the two legs upstream of the merge (or downstream of the diverge). The third leg of the junction can then be modeled as a continuation of this link, and the turning movements as localized lane-changes. No extra parameters have to be introduced. This approach to composition can also be used with more complex junctions, allowing us to create realistic (multi-lane) networks without introducing intersection-specific parameters. Further facilitating the task, the results of section 4 demonstrate that the continuous module of any such hybrid model can take eight different equivalent forms. The choice is one of programming convenience.[‡‡]

Unfortunately, although the hybrid approach is promising in theory, it has practical limitations. The composition rules for intersections involving multiple exit legs, such as diverges, require knowing the destinations of the vehicles making up the stream. The distribution of destinations strongly affects both, the discharge rates of diverge bottlenecks, as demonstrated by the natural experiments in [30], and the performance of intersections controlled by traffic signals. Thus, it cannot be ignored. Unfortunately, as a network grow in size, the number of possible destinations grows proportionately with the number of links, and it becomes more difficult to get the input data required by the model. Thus, the practical limit to composition is not theoretical (we could model relatively well almost anything if we knew where vehicles were going) but informational. The results in this paper can be of use for the design of small/medium networks such as small freeways and sets of complex interchanges, but other approaches should be sought for very large networks; reference [16] proposes some ideas.

**Acknowledgements.** The comments of Prof. Alexandre Bayen (University California, Berkeley), Prof. Jorge Laval (Georgia Institute of Technology) and Dr. Ludovic Leclercq (INRETS) are gratefully acknowledged. Research supported by the U. C. Berkeley Center for Future Urban Transport (a Volvo Foundations international center of excellence.)

## Appendix A: proofs.

**Definition 1.** Valid path: a continuous piecewise differentiable function, $x(t)$, such that:

[‡‡]Reference [18] used both the ACT and CA(L) methods (with similar results); Reference [28] used a variant of the CF(L) model; and the gridlock animation in http://www.its.berkeley.edu/volvocenter/gridlock/ uses the CA(M) method.

$$(x(t_2) - x(t_1))/(t_2 - t_1) \in [-w, u] \qquad \text{if } \ t_1 < t_2.$$

**Definition 2.** Cost function with bottlenecks, for $(t, x) \in \boldsymbol{S}$:

$$\begin{aligned} R_B(x', t, x) &= \min\{r_B(t), R(x')\}, \ \text{ if } \ x = x_B(t),\ x' = x'_B(t) \text{ for some } B. \\ &= R(x'), \text{ otherwise.} \end{aligned}$$

**Definition 3.** Auxiliary cost function:

$$R_A(x') = \max\{0, -x'r/w\} = \max\{0, -x'\kappa\}.$$

**Definition 4.** Auxiliary path costs, $\Delta_A(\mathcal{P})$: Costs obtained with the auxiliary cost function.

**Lemma 1.** $R_A(x') \le R(x')$ *for* $x' \ge -w$.

*Proof.* The lemma holds for $x' \ge 0$ since in this case $R_A(x') = 0 \le R(x')$. For $x' < 0$, we have: $R_A(x') = -x'r/w = -x'R(-w)/w = -x' \sup_k \{Q(k) + kw\}/w = -x' \sup_k \{Q(k)/w + k\}$. And since $0 < -x' \le w$, we can also write:

$$\begin{aligned} &-x' \sup_k \{Q(k)/w + k\} \le -x' \sup_k \{Q(k)/(-x') + k\} = \\ &= \sup_k \{Q(k) - x'k\} = R(x'). \end{aligned} \qquad \square$$

We assume for the rest of this appendix that $x'_B(t) \ge 0$ and $r_B(t) \ge 0$.

**Lemma 2.** $0 \le R_A(x') \le R_B(x', t, x)$.

*Proof.* In view of Lemma 1, we only need to prove Lemma 2 for the first case of Definition 2; i.e., it suffices to show that $R_A(x') \le R_B(x', t, x) = r_B(t)$ when $x = x_B(t)$, $x' = x'_B(t) \ge 0$, and $0 \le r_B(t) < R(x')$. This is obviously true since for $x' \ge 0$, $R_A(x') = 0$ and $R_B(x', t, x) \ge 0$. $\square$

**Lemma 3.** *If valid path* $\mathcal{P}$ *goes from point D to point P, then* $\Delta(\mathcal{P}) \ge \kappa(x_D - x_P)\}$.

*Proof.* It suffices to show that $\Delta(\mathcal{P}) \ge \Delta_A(\mathcal{P}) \ge \max\{0, \kappa(x_D - x_P)\}$. The first inequality follows from Lemma 2, since $R_B$ is the cost used to calculate $\Delta(\mathcal{P})$ and $R_A$ is the cost used to calculate $\Delta_A(\mathcal{P})$. The second inequality holds because $\max\{0, \kappa(x_D - x_P)\}$ is the auxiliary cost of the linear path from D to P, which is the least possible because $R_A(x')$ is time- and space-independent. $\square$

**Definition 5.** $\epsilon$-opt path, $\mathcal{P}^\epsilon_P$, for $\epsilon \ge 0$: a valid path from the boundary to P $\in \boldsymbol{S}$ such that:

$$N_{D(\mathcal{P})} + \Delta(\mathcal{P}^\epsilon_P) \le N_P + \epsilon.$$

Note: If P $\in \boldsymbol{S}$ the argument of (5) is bounded and the feasible set $\boldsymbol{P}$ is not empty. Thus, the infimum (5) exists, and there must be an $\epsilon$-opt path for any $\epsilon > 0$, no matter how small. If there are no bottlenecks, a straight line with $\epsilon = 0$ (i.e., optimal) always exists.

**Lemma 4.** *If C is a point of an* $\epsilon$*-opt path,* $\mathcal{P}^\epsilon_P$*, and* $\mathcal{P}^\epsilon_{CP} \subset \mathcal{P}^\epsilon_P$ *is the sub-path from C to P, then*

$$N_C + \Delta(\mathcal{P}^\epsilon_{CP}) \le N_P + \epsilon.$$

*Proof.* Let P′ be the beginning point of the $\epsilon$-opt path. The path capacity constraint ensures that $N_C \leq N_{P'} + \Delta(\mathcal{P}^\epsilon_{P'C})$. If we now add $\Delta(\mathcal{P}^\epsilon_{CP})$ to both sides, the result is: $N_C + \Delta(\mathcal{P}^\epsilon_{CP}) \leq N_{P'} + \Delta(\mathcal{P}^\epsilon_{P'C}) + \Delta(\mathcal{P}^\epsilon_{CP}) = N_{P'} + \Delta(\mathcal{P}^\epsilon_P) \leq N_P + \epsilon$. The last inequality holds because $\mathcal{P}^\epsilon_P$ is $\epsilon$-opt. □

**Lemma 5.** *If the conditions of Theorem 1 in the text hold for an IVP with bottlenecks, then its solution satisfies (6) in* $\boldsymbol{D} \cup \boldsymbol{S}$.

*Proof.* Condition (6) is obviously satisfied for $t = 0$. Thus, we only have to establish it for two arbitrary points A and B with coordinates $t_A = t_B = t > 0$ and $x_B > x_A$. We first prove that $N_B \leq N_A$. To do this we assume that this is not true, i.e., that there is an $\epsilon > 0$ such that $N_B > N_A + \epsilon$, and look for a contradiction.

Consider the path, $\mathcal{U}_B$, which reaches B with $x' = u$ from a source B′ $\in \boldsymbol{D}$. (Because its speed is maximum, we call $\mathcal{U}_B$ a "maximal path".) Consider as well an $\epsilon$-opt path from the boundary to A, $\mathcal{P}^\epsilon_A$ , which emanates from a point A′ $\in \boldsymbol{D}$. If the paths do not intersect (i.e., $x_{B'} > x_{A'}$) then $N_{B'} \leq N_{A'}$, as per (6) which applies on the boundary, and we can write: $N_B \leq N_{B'} \leq N_{A'} \leq N_A - \Delta(\mathcal{P}^\epsilon_A) + \epsilon \leq N_A + \epsilon$. (The first equality holds because the maximal path imposes a capacity constraint with zero cost, the third because our path is $\epsilon$-opt, and the fourth because $\Delta(\mathcal{P}^*_A)$ is non-negative.) But $N_B \leq N_A + \epsilon$ contradicts our hypothesis. Thus, $N_B \leq N_A$ if the paths do not intersect. If the paths intersect, there is a common point C. Clearly, $N_B \leq N_C$ since C is on a maximal path to B, which has zero cost; moreover, $N_C \leq N_A - \Delta(\mathcal{P}^\epsilon_{CA}) + \epsilon \leq N_A + \epsilon$, since Lemma 4 holds for the sub-path $\mathcal{P}^\epsilon_{CA}$ from C to A, and cost rates are non-negative. Thus, $N_B \leq N_C \leq N_A + \epsilon$, which is a contradiction. Hence, $N_B \leq N_A$ if the paths do not intersect. Since paths must either intersect or not intersect, we conclude that $N_B \leq N_A$.

To finish the proof we now show that $N_A \leq N_B + \kappa(x_B - x_A)$. Again, we assume that this is not true, i.e., that there is an $\epsilon > 0$ such that $N_A > N_B + \kappa(x_B - x_A) + \epsilon$, and find a contradiction.

Consider the minimal path, $\mathcal{W}_A$, which reaches A from the boundary with minimal speed ($x' = -w$) from a source A″ $\in \boldsymbol{D}$, and consider as well an $\epsilon$-opt path from the boundary to B, $\mathcal{P}^\epsilon_B$, which emanates from a point B″ $\in \boldsymbol{D}$. If the paths do not intersect, $x_{B''} > x_{A''}$, and we can write: $N_A \leq N_{A''} + \kappa(x_{A''} - x_A) \leq N_{B''} + \kappa(x_{B''} - x_{A''}) + \kappa(x_{A''} - x_A) = N_{B''} + \kappa(x_{B''} - x_A) \leq N_B - \Delta(\mathcal{P}^\epsilon_B) + \kappa(x_{B''} - x_A) + \epsilon \leq N_B - \kappa(x_{B''} - x_B) + \kappa(x_{B''} - x_A) + \epsilon = N_B + \kappa(x_B - x_A) + \epsilon$. (The first inequality holds because $\kappa(x_{A''} - x_A)$ is the cost of the minimal path from A″, which cannot overlap with any bottlenecks; the second one because $N_{A''} \leq N_{B''} + \kappa(x_{B''} - x_{A''})$ as per (6), which applies on the boundary; the first equality is algebraic; the third inequality because $\mathcal{P}^\epsilon_B$ is an $\epsilon$-opt path emanating from B″; the fourth because of Lemma 3; and the last equality is algebraic.) Thus, $N_A \leq N_B + \kappa(x_B - x_A) + \epsilon$ if the paths do not intersect — a contradiction. If the paths intersect, then there is a common point C and a sub-path, $\mathcal{P}^\epsilon_{CB}$, of the $\epsilon$-opt path that extends from C to B. And we see using similar logic that: $N_A \leq N_C + \kappa(x_C - x_A) \leq N_B - \Delta(\mathcal{P}^\epsilon_{CB}) + \kappa(x_C - x_A) + \epsilon \leq N_B - \kappa(x_C - x_B) + \kappa(x_C - x_A) + \epsilon = N_B + \kappa(x_B - x_A) + \epsilon$. (The second inequality is based on Lemma 4.) Thus, the contradiction $N_A \leq N_B + \kappa(x_B - x_A) + \epsilon$ also arises if the paths intersect. We must conclude that $N_A \leq N_B + \kappa(x_B - x_A)$. □

**Lemma 5a.** *If the conditions of Theorem* 2 *in the text hold for an FHP with bottlenecks, then its solution satisfies (6) in* $\boldsymbol{D} \cup \boldsymbol{S}$.

*Proof.* The proof of Lemma 5 can be repeated with only two changes. Note that point A can be on the boundary line $x = 0$ and point B on the boundary line $x = x^o$.

First change: for the proof that $N_B \leq N_A$ when the maximal and optimal paths $\mathcal{U}_B$ and $\mathcal{P}_A^\epsilon$ do not intersect we need to recognize that one or both paths may emanate from the boundary line $x = 0$, and that point A itself may be on this line. If A is on this line then the necessary condition for well-posedness implies $A' = A$ and $N_{A'} \leq N_A + \epsilon$. This inequality also holds if A is not on the boundary (see Lemma 5); thus it applies to all A. Furthermore, it is still true from condition (i) of Theorem 2 that $N_{B'} \leq N_{A'}$, even if one or both paths emanate from the line $x = 0$. Thus, we can still write $N_B \leq N_{B'} \leq N_{A'} \leq N_A + \epsilon$, and find the same contradiction.

Second change: for the proof that $N_A \leq N_B + \kappa(x_B - x_A)$ when the minimal and optimal paths $\mathcal{W}_A$ and $\mathcal{P}_B^\epsilon$ do not intersect, we need to recognize that one or both paths may emanate from the line $x = x^o$ and that point B itself may be on this line. We still assume that there is an $\epsilon > 0$ such that $N_A > N_B + \kappa(x_B - x_A) + \epsilon$, and look for a contradiction. If B is not on the boundary line we can write: $N_B + \epsilon \geq N_{B''} + \Delta(\mathcal{P}_B^\epsilon) \geq N_{B''} + \kappa(x_{B''} - x_B)$, because $\mathcal{P}_B^\epsilon$ is *e*-opt, and then invoking Lemma 3. Now add, $\kappa(x_B - x_A)$ to both sides and obtain: $N_B + \kappa(x_B - x_A) + \epsilon \geq N_{B''} + \kappa(x_{B''} - x_A) \geq N_{A''} - \kappa(x_{B''} - x_{A''}) + \kappa(x_{B''} - x_A) = N_{A''} + \kappa(x_{A''} - x_A) \geq N_A$. This is our contradiction. [The first inequality is algebraic; the second holds because condition (i) of Theorem 2, which applies on the boundary, implies $N_{B''} \geq N_{A''} - \kappa(x_{B''} - x_{A''})$ since $x_{B''} \geq x_{A''}$ and $t_{B''} > t_{A''}$; and the last inequality holds because the term $\kappa(x_{A''} - x_A)$ is the cost of the minimal path from A$''$ to A, $\mathcal{W}_A$, which imposes the indicated capacity constraint on $N_A$.] If B is on the boundary line then the necessary consistency condition implies that B$''$ = B. Hence, the first inequality in the above string becomes a pure equality, and the rest continue to hold. □

**Lemma 5b.** *(Strictly monotone problems without bottlenecks): If for an IVP without bottlenecks the conditions of Lemma 5 are satisfied and $N(0, x)$ decreases with $x$, then N(t, x) decreases with $x$ for any $t > 0$. This is also true for problems with bottlenecks in the part of the solution domain upstream of the bottlenecks.*

*Proof.* It suffices to prove that that $N_B < N_A$ for two arbitrary points A and B in the solution domain with coordinates: $t_A = t_B$ and $x_B > x_A$. To this end, let $\mathcal{U}_B$ be the maximal path to B, and $\mathcal{L}_A^*$ a linear optimal path to A. The latter exists because there are no bottlenecks. Let $A'$ and $B'$ the beginning points of $\mathcal{L}_A^*$ and $\mathcal{U}_B$, respectively, and define $v_A = (x_A - x_{A'})/(t_A - t_{A'})$. This is the slope of $\mathcal{L}_A^*$.

If the paths do not intersect, $N_B \leq N_{B'} < N_{A'} \leq N_A$; i.e., $N_B < N_A$. [The first inequality holds because a maximal (zero-cost) path goes from B$'$ to B; the second is implied by boundary condition (i) of Theorem 2; and the third because an optimal path with non-negative cost goes from A$'$ to A.] Something similar happens if the paths intersect. Then there is a common point C and a portion of the optimal linear path $\mathcal{L}_{CA}^*$ goes from C to A. Hence, $N_B \leq N_C = N_A - \Delta(\mathcal{L}_{CA}^*) = N_A - R(v_A)(t_A - t_C) < N_A$; i.e., $N_B < N_A$. [The first inequality holds because B is on a maximal (zero cost) path from C; the second by virtue of Lemma 4, with $\epsilon = 0$; the third because $\mathcal{L}_{CA}^*$ is linear; and the fourth because an intersection point can only exist if $v_A < u$ and $t_A > t_C$, which imply $R(v_A) > 0$ and $R(v_A)(t_A - t_C) > 0$.] Thus, $N_B < N_A$ whether or not the paths intersect.

If there are bottlenecks but points A and B are upstream of the bottlenecks, the previous arguments continue to hold. □

**Lemma 6.** *(Bounded flows): If the conditions of Theorem 1 (or Theorem 2) hold for an IVP (or FHP) with bottlenecks, then its solution satisfies (7) in* $\boldsymbol{S} \cup \boldsymbol{D}$.

*Proof.* Since paths with $x' = 0$ are valid and satisfy $R_B \le R = q_{max}$, it follows that $N(t_B, x) - N(t_A, x) \le (t_B - t_A)q_{max}$ if $t_A < t_B$. This is true for an FHP even if $x = 0$ or $x = x^o$. Thus, the upper bound part of (7) holds for the IVP and FHP.

We now prove the lower bound part for the IVP. If the lower bound is false there should be two points A and B with coordinates $(t_A, x)$ and $(t_B, x)$, with $t_A < t_B$, such that $N_A > N_B + \epsilon$ for some $\epsilon > 0$. Consider too an $\epsilon$-opt path to B $\mathcal{P}^{\epsilon}_B$ (rooted at B′). If we consider a maximal path to A, $\mathcal{U}_A$ (rooted at A′), we can write (for obvious reasons):

$$N_A \le N_C \le N_C + \Delta(\mathcal{P}^{\epsilon}_{CB}) \le N_B + \epsilon, \qquad \text{if } \exists \text{ a common C} \tag{A1}$$

$$N_A \le N_{A'} \le N_{B'} \le N_{B'} + \Delta(\mathcal{P}^{\epsilon}_{B}) \le N_B + \epsilon, \qquad \text{otherwise.} \tag{A2}$$

This is the contradiction. The contradiction can also be derived using a minimal path $\mathcal{W}_A$ (rooted at A″), instead of $\mathcal{U}_A$. The inequalities then are:

$$N_A \le N_C + \kappa(x_C - x) \le N_C + \Delta(\mathcal{P}^{\epsilon}_{CB}) \le N_B + \epsilon, \quad \text{if } \exists \text{ a common C} \tag{A3}$$

$$\begin{aligned} N_A &\le N_{A''} + \kappa(x_{A''} - x) \le N_{B'} + \kappa(x_{B'} - x_{A''}) + \kappa(x_{A''} - x) \\ &\le N_{B'} + \kappa(x_{B'} - x) \le N_{B'} + \Delta(\mathcal{P}^{\epsilon}_{B}) \le N_B + \epsilon, \qquad \text{otherwise.} \end{aligned} \tag{A4}$$

We now prove the monotonicity condition for the FHP. The condition is satisfied at the boundary, by construction. Thus, we only need to check it in $\boldsymbol{S}$. We again consider points A and B as above, and the contradiction involving an $\epsilon$-opt path to B. If this $\epsilon$-opt path intersects $\mathcal{U}_A$ or $\mathcal{W}_A$, we invoke (A1) or (A3) and the contradiction follows. If it intersects neither, but intersects the initial boundary, then again (A2) or (A4) establish it. If it intersects the upper boundary, $x = 0$, we can use (A2) because the second inequality involving A′ and B′ still holds, since $N(t, 0)$ is non-decreasing. If it intersects the boundary $x = x^o$ we can use (A4) – since the second equality involving A″ and B′ still holds. Thus, a contradiction arises independent of the location of B′. □

**Lemma 7.** *Theorem 3 of the text holds.*

*Proof.* It suffices to show that conditions (i), (ii), (iii) of Theorem 3 imply the conditions of Theorem 2. It is obvious that conditions (ii) and (iii) of Theorem 3 imply conditions (ii) and (iii) of Theorem 2. Thus, we only have to prove the monotonicity of $N(t, x^b)$. We prove it first for $t \le t^* \equiv \min\{x^o/u, x^o/w\}$. This restriction prevents valid paths from one boundary, $x = x^b$, to reach the other.

*Proof for the upper boundary, $x = 0$, when $t \le t^*$*: At any point on this boundary either: (a) $N(t, 0) = U(t)$ (if $U(t) \le n_D(t, 0)$); or (b) $N(t, 0) = n_D(t, 0)$ (if $U(t) > n_D(t, 0) = N(t, 0)$). Clearly, $N(t, 0)$ can only decline where (b) holds. For this to happen there must be an interval where (b) holds, with points A and B such that $t_A < t_B$ and $N_B + \epsilon < N_A$ for some $\epsilon > 0$. But this cannot happen because it leads to the usual contradiction with paths $\mathcal{W}_A$ and $\mathcal{P}^{\epsilon}_B$, as we now show.

If $\mathcal{W}_A$ and $\mathcal{P}^{\epsilon}_B$ intersect then (A3) holds, with $x = 0$, which is a contradiction. Note now that an $\epsilon$-opt path that starts at or before $t_A$ must exist. (This is true because where $N(t, 0) = n_D(t, 0)$, cost is determined from $\boldsymbol{D}_t$; therefore, the cost of any point in $[t_A, t_B]$ must be determined from $\boldsymbol{D}_{t_A}$.) Thus, if the paths do not cross $\mathcal{P}_B^\epsilon$ must start from a point A$''$ on the initial line with $x_{A''} > x_{B'}$. Therefore (A4) must hold. This is also a contradiction. Thus, $N(t, 0)$ cannot decline in $[0, t^*]$.

*Proof for the lower boundary, $x = x^o$, when $t \le t^*$*: The same logic, applied to $N(t, x^o)$ using paths $\mathcal{U}_A$ and $\mathcal{P}_B^\epsilon$, reveals a contradiction involving either (A1) or (A2) if $N(t, x^o)$ declines. Thus, $N(t, x^o)$ cannot decline in $[0, t^*]$.

*Proof for both boundaries and $t > 0$*: An induction argument is used. Assume that $N(t, x^b)$ is non-decreasing for $t \in [0, jt^*]$, where $j = 1, 2, 3...$ ; i.e., that condition (i) of Theorem 2 is satisfied up to time $jt^*$. Note that the assumption is true for $j = 1$. The monotonicity proof can now be repeated step by step for the boundary points in $(jt^*, (j+1)t^*]$. The only difference is that paths $\mathcal{W}_A$, $\mathcal{U}_A$ and $\mathcal{P}_B^\epsilon$ can now be rooted in the opposite boundary. However, these paths must start at a point with $t \le jt^*$ where $N(t, x^b)$ is already known to be non-decreasing. This ensures that the second inequalities of (A2) and (A4) continue to hold. Hence, one of the 4 inequalities (A1)-(A4) will still hold and yield the desired contradiction. □

**Comment**: For the FHP (and other problems with complex boundaries) valid paths from the boundary to a point P in the solution domain may leave $\boldsymbol{S}$ and return to it. The consistency condition (iii) implies that such paths cannot be unique optima. Hence, they can be ignored when solving (5). This is recommended for applications because sufficient networks for numerical solution then only have to be defined in $\boldsymbol{S}$. (The reader may want to verify that the proofs of Lemmas 5a and 6 still hold if paths are not allowed to leave $\boldsymbol{S}$.)

**Appendix B: asynchronous cell transmission model.** In this appendix primes do not denote derivatives. The cell transmission model (CTM) [7] with time step $\varepsilon$, cells of size $\delta = u\varepsilon$, and a triangular FD is:

$$N_{lm} - N_{l-1,m} = \min\{N_{l-1,m-1} - N_{l-1,m}\ ,\ \varepsilon q_{max}\ ,\ [\kappa - (N_{l-1,m} - N_{l-1,m+1})/\delta]w\varepsilon\}. \qquad \text{(B1)}$$

The LHS of (B1) is the flow advancing in one time step across the $m^{th}$ intercell boundary. The RHS is a function of the vehicles currently in the upstream and downstream cells. It is known [21] that the CTM formula is a streamlined version of Godunov's method. It is also known [7, 8] that the last term of the CTM formula, which expresses the available capacity of the downstream cell, introduces a first order numerical error when $w \ne \delta/\varepsilon$, and that the error vanishes (the method is exact) if $w = \delta/\varepsilon$. Fortunately, these errors can be eliminated by changing the time variable to asynchronous time as proposed in Sec. 5.2.2 of [13]. The resulting procedure — no longer in Godunov's family — was called the asynchronous cell transmission model (ACTM) in this reference.

To summarize, imagine that clocks at each location have been synchronized with the passage of a reference vehicle with negative speed, $s$, such that: $1/s = 1/u - 1/w < 0$. Thus, the new (asynchronous) time is $t' = t + x/s$, and the new lattice instants at $x = x_m$ are related to the old by: $l' = t'_l/\varepsilon = t_l/\varepsilon + x_m/s\varepsilon = t_l/\varepsilon + m\delta/s\varepsilon = t_l/\varepsilon + m(u\varepsilon)(1/u - 1/w)\varepsilon = l + m(1 - \theta)$. If $\theta$ is an integer then the lattice remains the same, since the lattice instants are displaced from the old by an integer multiple of the time step. This leaves invariant the jam density but changes speed as per: $1/v' = 1/v + 1/w - 1/u$. The advantage of the new coordinate system is that the speed of the backward wave adopts the value $-w'$ such that: $1/(-w') = 1/(-w) + 1/w - 1/u$; i.e., $w' = u$, and therefore $w' = \delta/\varepsilon$. Thus, the formula for available capacity in the new coordinate system, which is $[\kappa - (N_{l'-1,m} - N_{l'-1,m+1})/\delta]w'\varepsilon$, is exact. In terms of the old variables, $l$ and $w'$, this expression becomes $[\kappa - (N_{l'-1,m} - N_{l'-1,m+1})/\delta]w'\varepsilon = [\kappa$ - $(N_{l-1,m} - N_{l-\theta,m+1}]/\delta)u\varepsilon = \kappa\delta$ - $(N_{l-1,m} - N_{l-\theta,m+1})$. Substituting this expression for the last term of (B1) we obtain the exact ACTM recipe:

$$N_{lm} - N_{l-1,m} = \min\{N_{l-1,m-1} - N_{l-1,m}\ ,\ \varepsilon q_{max}\ ,\ \kappa\delta - (N_{l-1,m} - N_{l-\theta,m+1})\}, \tag{B2}$$

or

$$N_{lm} = \min\{N_{l-1,m-1}\ ,\ N_{l-1,m} + \varepsilon q_{max}\ ,\ \kappa\delta + N_{l-\theta,m+1}\}. \tag{B3}$$

## REFERENCES


[1] J.-P. Aubin, A. M. Bayen and P. Saint-Pierre, *Dirichlet problems for some Hamilton-Jacobi equations with inequality constraints*, Preprint di Matematica - n. 4, Scuola Normale Superiore, Pisa, Italy, May 2006 (submitted for publication to SIAM Journal on Control and Optimization).

[2] M. Bardi and L.C. Evans, *On Hopf's formulas for solution of Hamilton Jacobi equations*, Nonlinear Analysis **8**(11) (1984), 1373-1381.

[3] M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton Jacobi Bellman equations*, Birkhauser Boston, Boston, Massachusetts (1997).

[4] C. Bardos, A. Y. Leroux and J. C. Nedelec, *First order quasilinear equations with boundary conditions*, Communications in Partial Differential Equations **4**(9) (1979), 1017–1034.

[5] R. Cayford, W. H. Lin and C. F. Daganzo, *The NETCELL simulation package: Technical description*, PATH Research Report, UCB-ITS-PRR-97-23, University of California, Berkeley, CA (1997).

[6] M. G. Crandall and P. L. Lions, *Viscosity solutions of Hamilton-Jacobi equations*", Transactions of the American Mathematical Society **277**(1) (1983), 1-42.

[7] C.F. Daganzo, *The cell-transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*, Institute of Transportation Studies, Research Report, UCB-ITS-RR-93-7, Univ. of California, Berkeley, CA (1993); abridged in Transportation Research Part B, **28**(4), 269-287.

[8] C.F. Daganzo, *A finite difference approximation for the kinematic wave model*, Institute of Transportation Studies Research Report, UCB-ITS-RR-93-11, Univ. of California, Berkeley, CA (1993); abridged in Transportation Research Part B **29**(4) 261-276.

[9] C.F. Daganzo, *The cell-transmission model. Part II: Network traffic*, PATH Working Paper UCB-ITS-PWP-94-12, Univ. of California, Berkeley, CA (1994); abridged in Transportation Research Part B, **29**(2), 79-94.

[10] C. F. Daganzo, *The nature of freeway gridlock and how to prevent it*, Institute of Transportation Studies Research Report UCB-ITS-RR-95-1, Univ. of California, Berkeley, CA (1994); abridged in "Transportation and Traffic Theory" (ed. J.B. Lesort) pp. 629 646, Pergamon Elsevier, Tarrytown, N.Y.

[11] C.F. Daganzo, "A Variational Formulation for a Class of First Order PDEs," Institute of Transportation Studies Report UCB-ITS-RR-2003-03, U. of California, Berkeley, CA 2003.

[12] C.F. Daganzo, *A Variational formulation of kinematic waves: solution methods*, Institute of Transportation Studies Report UCB-ITS-RR-2003-07, U. of California, Berkeley, CA (2003); abridged in Transportation Research Part B, **39**(2) 934-950.

[13] C.F. Daganzo, "A Theory of Supply Chains," Springer, Heidelberg, Germany (2003).

[14] C.F. Daganzo, *A Variational formulation of kinematic waves: basic theory and complex boundary conditions*, Transportation Research Part B, **39**(2) (2005), 187-196.

[15] C.F. Daganzo, *In traffic flow cellular automata = kinematic waves*, Transportation Research Part B, **40**(5) (2006), 396-403.

[16] C.F. Daganzo, *The nature or urban gridlock*, U.C. Berkeley Centre for Future Urban Transport, working paper VWP-2006-2, (2006); Transportation Research Part B, **41** (1) (2007), 49-62.

[17] P. D. Lax, *Hyperbolic systems of conservation laws II*, Communications in Pure and Applied Mathematics 10,(1957), 537-566 .

[18] J. A. Laval and C. F. Daganzo, *Lane-changing in traffic streams*, Transportation Research Part B, **40**(3)(2006), 251-264.

[19] J.A. Laval, M. J. Cassicy and C. F. Daganzo, *Impacts of lane changes at on-ramp bottlenecks: a theory and strategies to maximize capacity*, Presented at Traffic and Granular Flow05, Berlin, Germany, October 10-12, (2005).

[20] J.P. Lebacque, *Les modèles macroscopiques de trafic*, Annales de Ponts (3rd trim) **67** (1993), 28-45.

[21] J.P. Lebacque, *The Godunov scheme and what it means for traffic models*, pp 647-678 in "Traffic and Transportation Theory" (J.B Lesort, editor), Proc. ISTTT, Elsevier, Tarrytown, N.Y. 1996.

[22] M. J. Lighthill and G. B. Whitham, *On kinematic waves: i flow movement in long rivers. ii a theory of traffic flow on long crowded roads*, Proceedings of the Royal Society A **229** (1955), 281-345.

[23] W. H. Lin and C. F. Daganzo, *Technical description of NETCELL: General framework and data structure*, PATH Technical Note UCB-ITS-TN-94-7, University of California, Berkeley, CA (1994).

[24] H. K. Lo, *A novel traffic signal control formulation*, Transportation Research Part A **33** (1999), 433-448.

[25] H. K. Lo, *A dynamic traffic assignment formulation that encapsulates the cell-transmission model*, pp 327-350 in "Traffic and Transportation Theory" (A. Ceder, editor), Proc. ISTTT, Elsevier, Tarrytown, N.Y. 1999.

[26] Y. Makigami, G.F. Newell, and R. Rothery, *Three-dimensional representation of traffic flow*, Transportation Science **5** (1971), 302-313.

[27] M. Menendez and C. F. Daganzo, *A variational theory of kinematic waves: bottleneck properties and examples*, pp 345-364 in "Traffic and Transportation Theory" (H. Mahmassani, editor), Proc. ISTTT, Elsevier, New York, N.Y. 2005.

[28] M. Menendez and C. F. Daganzo, *Effects of HOV lanes on freeway bottlenecks*, U.C. Berkeley Center for Future Urban Transport Working Paper UCB-ITS-VWP-06-3, (2006); submitted for publication.

[29] K. Moskowitz, *Discussion of 'freeway level of service as influenced by volume and capacity characteristics' by D.R. Drew and C. J. Keese* Highway Research Record **99** (1965), 43-44.

[30] J.C. Muñoz and C. F. Daganzo, *The bottleneck mechanism of a freeway diverge*, Transportation Research Part A **36** (2002), 483-505.

[31] K. Nagel and M. Schreckenberg, *A cellular automaton model for freeway traffic*, Journal de Physique I **2** (1992), 2221-2229.

[32] G. F. Newell *A simplified theory of kinematic waves in highway traffic: (i) general theory; (ii) queuing at freeway bottlenecks; (iii) multi-destination flows*, Transportation Research Part B **27** (1993), 281-313.

[33] G. F. Newell *A simplified car-following theory: A lower order model*, Transportation Research Part B **36** (2002), 195-205.

[34] O. A. Oleinik *Discontinuous Solutions of Nonlinear Differential Equations*, Uspekhi Mat. Nauk. **12** (1957), 3-73. (English translation: American Mathematical Society, Ser. 2 **26** (1963) 95–172.)

[35] P. I. Richards *Shockwaves on the highway*, Operations Research **4** (1956), 42-51.

[36] A. K. Ziliaskopoulos, *A linear programming model for the single destination system optimum DTA problem*, Transportation Science **34** (2000), 37-49.

Received July 2006; revised September 2006.

*E-mail address*: daganzo@ce.berkeley.edu