# A theory of traffic flow in automated highway systems*

M. Broucke and P. Varaiya
Department of Electrical Engineering and Computer Science
University of California, Berkeley CA 94720
mire, varaiya@eclair.eecs.berkeley.edu

February 22, 1996

**Abstract**

This paper presents a theory for automated traffic flow, based on an abstraction of vehicle activities like entry, exit and cruising, derived from a vehicle's automatic control laws. An activity is represented in the flow model by the space occupied by a vehicle engaged in that activity.

The theory formulates TMC traffic plans as the specification of the activities and speed of vehicles, and the entry and exit flows for each highway section. We show that flows that achieve capacity can be realized by stationary plans that also minimize travel time. These optimum plans can be calculated by solving a linear programming problem. We illustrate these concepts by calculating the capacities of a one lane automated highway system, and compare adaptive cruise control and platooning strategies for automation.

Keywords: automated highway system, traffic flow models, capacity of automated highways, traffic management

## 1 Introduction

This paper proposes a theory of automated highway traffic. The theory predicts the performance of an automated highway system (AHS) in terms of achievable (steady state) flows and travel times. The performance predictions can be used to compare alternative AHS designs.

The theory shows how AHS steady state performance is a function of the characteristics of both the control laws that govern the movement of individual vehicles and the traffic management rules that guide the vehicle flow. This functional relationship can be used to suggest changes in vehicle control laws and traffic management rules for improving highway performance.

---

*This research is supported by the California Department of Transportation through the California PATH program and the Federal Highway Administration through the National Automated Highway System Consortium. We thank Datta Godbole, Roberto Horowitz, John Lygeros, and Jim Morehead for helpful discussions.

The theory also explains how the automated highway can become congested, and what sorts of actions need to be taken to prevent congestion from occurring and to eliminate it once it occurs. Thus the theory may be used to design vehicle control and traffic managment rules for reducing undesirable transient behavior such as congestion.

In an AHS vehicles are under automatic control: the distance a vehicle maintains from the vehicle in front, its speed, and its route from entry into the highway to exit, are all determined by the vehicle's feedback control laws. One may therefore compare the effect on the traffic of changes in vehicle control laws, and seek to calculate the "optimum" control laws. By contrast, in non-automated traffic flow theory, the driver determines a vehicle's headway, its speed, its movement during a merge, etc. Driver behavior is difficult to change significantly. One hypothesizes feedback models of driver behavior and uses real data or experiment to calibrate the model parameters.

Similarly, the traffic managment system (TMC) for the AHS can directly influence the flow by issuing orders to vehicles regarding their speed and route. Those orders will be obeyed because the vehicles are programmed to do so. The TMC for the non-automated highway also can make speed and route suggestions, but drivers may ignore these suggestions or react to them in an unexpected manner. Thus, the influence of TMC policies in the AHS is much stronger and more predictable than its influence on non-automated traffic; and so, one may again seek to determine optimum TMC policies.

Because it is possible to exercise much greater control over the movement of individual vehicles and the traffic as a whole, a theory of AHS traffic flow will tend to be *prescriptive*. Non-automated traffic flow theory is more *descriptive*, by contrast.[1]

We now introduce the main abstractions and assumptions and the structure of the proposed theory. The theory is based on an *activity* model: the movement of a vehicle is conceptualized as a sequence of activities, such as entry, cruise, and exit, which are realized by vehicle control laws; the highway is viewed as providing the space necessary to carry out each activity; the vehicle control laws and vehicle speed determine the time to complete an activity.[2] When there is insufficient space in one section of the highway, the rate of activity completion in the immediately upstream section must be reduced. Since the rate of activity completion is proportional to the speed, this causes a reduction in flow.

In this way, the interaction between the demand for space by vehicle activities and the fixed supply of space offered by the highway determines the steady state flows that can be realized, as well as the transient congestion effects that can occur. This interaction is mediated by the vehicle control policies (which determine the space needed for each activity) and the traffic management rules (which determine the activities that are to be carried out in different sections of the highway). That is how the theory relates AHS performance to characteristics of vehicle control and traffic management rules.

We now introduce the two assumptions, which we call "one activity per section" and "safety needs space," that bind together activities, vehicles and highway.

---

[1] Of course, this descriptive theory is used to design and prescribe ramp metering and other traffic management rules.

[2] This activity model is inspired by the work in [1].

To fix ideas, we assume that the AHS has a single lane, with entrances and exits. At each instant of time every (automated) vehicle is engaged in one of a finite number of activities such as cruising, changing a lane (in case of a multi-lane highway), entering the highway, exiting the highway, etc. If vehicles are organized in closely-spaced platoons, then cruising in a one-vehicle platoon is a different activity from cruising in a two-vehicle platoon, and so on. Cruising in platoons of different sizes are considered different activities because the space needed per vehicle in a cruising platoon decreases with the platoon size. (See [2].)

The highway is divided into sections, and we will assume that a vehicle executes a single activity in each section through which it travels. Consequently, the passage of a vehicle through the automated highway can be summarized by the sequence of activities that the vehicle executes, starting with the "entry" activity in the section where it enters and terminating with the "exit" activity in the section where it leaves the highway. In this model, vehicles are assumed to travel at a constant average speed within each section, and the assumption of "one activity per section" rigidly ties the spatial discretization of the highway into sections with the temporal discretization of movement into activities. Consequently, variation in speed due to interaction of activities is not captured here. Although not mathematically necessary, we adopt the one-activity-per-section assumption because it greatly simplifies the model description. (See [3] for a related modeling move to tie together spatial and temporal discretization.)

While it is engaged in a particular activity, a vehicle's motion is governed by a feedback control law which ensures that this activity is carried out safely. These feedback laws and the resulting vehicle motion can be complicated.[3] But for our purposes we will work with the assumption "safety needs space."

To motivate this assumption, consider the "cruising" activity, in which a vehicle keeps in one lane and its cruise control law guarantees safety by maintaining a minimum safe distance between its vehicle and the vehicle in front of it. This distance is an increasing function of vehicle speed.[4] We shall assume a maximum permissible speed and let $s(cruise)$ be the corresponding minimum safe distance between a cruising vehicle and the vehicle in front of it. Thus the safety-needs-space assumption says that its feedback law will guarantee that a cruising vehicle will "occupy" $s(cruise)$ meters of a highway lane. Together with the one-activity-per-section assumption, this implies that a cruising vehicle will occupy $s(cruise)$ meters in a section so long as it remains in that section.

In general, safety-needs-space says that vehicle control laws cause a vehicle engaged in activity $\alpha$ to occupy a distance $s(\alpha)$ from which all other vehicles are excluded. For activities involving vehicles in two lanes, as happens during a lane change and in some implementations of entry/exit, the vehicle occupies a minimum safety distance in both lanes.

The time the vehicle spends in a section is equal to the section length divided by the vehicle speed. When a vehicle engaged in activity $\alpha$ leaves this section, its $s(\alpha)$ space is available for use by another vehicle from the upstream section. The longer the vehicle stays in its section, the later will its space become available, and this may slow down upstream vehicles. Thus,

---

[3] Examples of such feedback laws are given in [4, 5, 6, 7].

[4] This function depends on other parameters such as maximum vehicle braking torque, road surface and tire conditions, etc.

if the activities that vehicles are executing in different sections are not well coordinated, the speed in some sections may be forced below the maximum or free flow speed, causing congestion. Traffic management rules determine the activities that vehicles undertake and their speed, and thus, ultimately, the AHS steady state performance as well as how well congestion is dissipated.

The remainder of the paper is organized as follows. In section 2 we introduce the formal activity model. This is a system of differential equations, several parameters of which are set by TMC plans, including vehicle speed and activity, and entry and exit flows.

TMC plans and achievable flows are studied in section 3. An achievable flow is any vector of flows (indexed by origin-destination pairs or other characteristics) that can be sustained in the long run. The main result of this section is that the set of achievable flows is convex.

In section 4 we define AHS capacity as the set of undominated achievable flows, and efficient TMC plans as those which minimize travel time. We show that every undominated flow, together with an efficient plan that achieves this flow, can be computed by solving a linear programming problem.

In section 5 we consider transient behavior: how congestion can develop and how TMC (feedback) rules can mitigate its effects. We will see that the information available to the TMC has a significant effect on what kind of rules can be implemented.

In section 6 we focus on two particular activities—entry and exit. These activities are likely to be the most important activities that limit AHS performance. In section 7 we compare two alternative AHS designs using the proposed theory. Section 8 collects some concluding remarks.

## 2 The activity model

We study a one-lane automated highway, divided into sections. Sections are indexed $i = 1, ..., I$; section $i$ is $L(i)$ meters in length. Section $i-1$ is upstream of section $i$. Time is indexed $t = 0, 1, ...$. Each time period is $\tau$ seconds long.

**Vehicles** Vehicles have types indexed by $\theta$ which may stand for their origin and destination and all other distinguishing characteristics of interest.

All vehicles in section $i$ at time $t$ have the same speed, denoted $v(i,t)$, and measured in meters/sec. It is required that $v(i,t) \leq V$, the maximum permissible or free flow speed. ($V$, too, may be indexed by $i$, but we don't do that to ease the notational burden.)

Let $n(i,t,\theta)$ be the number of vehicles of type $\theta$ in section $i$ at time $t$. We adopt the notational convention that $n(i,t)$ is the array indexed by $\theta$, $n(i)$ is the array indexed by $(t,\theta)$, and so on.

**Activity plan** There are finitely many activities, indexed by $\alpha$. An *activity plan* is any array of non-negative numbers $\pi = \{\pi(\alpha,i,t,\theta)\}$ such that for every $i,t,\theta$

$$\sum_{\alpha} \pi(\alpha,i,t,\theta) \equiv 1.$$

$\pi(\alpha, i, t, \theta)$ is the fraction of the $n(i, t, \theta)$ vehicles engaged in activity $\alpha$.

Associated with each activity $\alpha$ is the length (in meters) $\lambda(\alpha) > 0$ of the section occupied by each vehicle engaged in that activity. Thus $n(i, t)$ vehicles engaged in activities $\pi(i, t)$ will occupy

$$\sum_{\alpha} \sum_{\theta} \lambda(\alpha)\pi(\alpha, i, t, \theta)n(i, t, \theta)$$

meters of section $i$ in period $t$.

Two vehicles with the same $(i, t, \theta)$ index and engaged in the same activity cannot be further distinguished within the model. In that sense, this is a theory of vehicle *flow*. The theory aggregates individual vehicle movement using the one-activity-per-section and safety-needs-space assumptions.

**Speed plan** A *speed plan* is an array of nonnegative numbers $v = \{v(i, t)\}$ (in meters/sec), each less than $V$. All $n(i, t)$ vehicles move at $v(i, t)$ meters/sec to conform to the plan. This restriction in part is imposed by the single lane highway: since vehicles cannot pass each other, relative speeds cannot be too great. However, the restriction also presupposes that the sections are not so long that vehicles with significantly different speeds can coexist in the same section.

It is possible, at the cost of further notational complexity, to introduce the following features. Suppose the vehicle type $\theta$ also signifies vehicle body type: light duty, truck, bus, etc. Then we can insist that the space required depends also on vehicle type, i.e., we have $\lambda(\alpha, \theta)$. We can also insist that vehicle maximum speed is a function of $\theta$, $V(\theta)$, and require that the speed $v(i, t)$ be smaller than the maximum permissible speed, i.e., $n(i, t, \theta) > 0$ implies $v(i, t) \leq V(\theta)$. These features are very useful and easy to introduce in the simulation system, but they would make this paper difficult to read.

**Highway configuration** We have already specified parts of the highway configuration. We have a one-lane highway, divided into sections $i = 1, ..., I$ of length $L(i)$. Section $i$ is immediately downstream of section $i - 1$. It remains to specify entry and exit. Each section has at most one entrance and one exit. Vehicles can make an entry through some dedicated infrastructure that connects a non-automated highway or street to the AHS entrance. Vehicles can exit the AHS through another transitional infrastructure.[5] We can require that an entering vehicle must engage in a distinguished "entry" activity, and an exiting vehicle must engage in "exit." These activities will occupy more space than most other activities because they will involve merging from a ramp or a transition lane into the AHS main lane.

In a following paper we will extend the model to a multi-lane AHS. Such an extension then permits one to consider the "lane change" activity. It also permits the possibility of modeling entry and exit as a kind of lane change.

**Entry and exit plans** An entry plan is an array $f = \{f(i, t, \theta)\}$ of non-negative numbers. $f(i, t, \theta)$ is the number of vehicles of type $\theta$ that enter the highway in section $i$ in period $t$.

An exit plan is an array $g = \{g(i, t, \theta)\}$ of non-negative numbers. $g(i, t, \theta)$ is the number of vehicles of type $\theta$ that exit the highway in section $i$ in period $t$.

---

[5] See [8] for several transition infrastructure designs, and [9] for a similar highway configuration.

If entry or exit in a particular section, say $j$, is forbidden, one merely adds the constraint: $f(j,t,\theta) \equiv 0$ or $g(j,t,\theta) \equiv 0$, for all $t, \theta$. We will shortly impose more complex constraints on all the plans.

**Dynamics** The *state* of the system at time $t$ is $n(t) = \{n(i,t,\theta)\}$. Suppose that we are given an activity plan $\pi$, a speed plan $v$, an entry plan $f$, and an exit plan $g$. Let $n(t)$ be the state at time $t$. Then, for all $t$ and $1 \leq i \leq I$,

$$n(i,t+1,\theta) = \rho(i,t)n(i,t,\theta) + [1-\rho(i-1,t)]n(i-1,t,\theta) + f(i,t,\theta) - g(i,t,\theta). \tag{1}$$

Since the AHS sections are $i = 1, ..., I$, we also have the boundary conditions,

$$\begin{aligned} n(0,t,\theta) &= 0, \text{ for all } t, \theta, & (2) \\ n(I+1,t,\theta) &= 0, \text{ for all } t, \theta. & (3) \end{aligned}$$

Equation (1) should be interpreted as follows. First, by definition,

$$1 - \rho(i,t) := \frac{v(i,t) \times \tau}{L(i)}. \tag{4}$$

Here $\rho(i,t)$ is the fraction of vehicles in section $i$ at time $t$ that remain in that section for time $t+1$. So $[1-\rho(i,t)]$ is the fraction of vehicles in section $i$ at time $t$ that leave that section at the end of that period. By definition (4), the fraction of vehicles that leave is equal to the fraction of the section length $L(i)$ that is traveled in time $\tau$ by vehicles moving at speed $v(i,t)$. Thus this definition assumes a spatial homogeneity of the disposition of vehicles in each section. Obviously this is not the case at the level of individual vehicles. But in our model, a homogeneity assumption of this kind is necessary since we want the state simply to be the number of vehicles in each section.[6]

Thus, the first term on the right in (1) is the number of vehicles in $i$ at time $t$ that remain in $i$ at time $t+1$, and the second term is the number of vehicles in $i-1$ at time $t$ that move into $i$ at time $t+1$. The last two terms are straightforward: $f(i,t,\theta)$ is the number of vehicles of type $\theta$ that enter the AHS according to the entry plan, and $g(i,t,\theta)$ is the number that leave the AHS.

The boundary condition (3) implies that all vehicles in section $I$ leave the AHS:

$$g(I+1,t,\theta) = [1-\rho(I,t)]n(I,t,\theta), \quad f(I+1,t,\theta) = 0. \tag{5}$$

**Fact 1** $n(t)$ is indeed a state, i.e., given $n(0)$ and activity, speed, entry and exit plans $u(t) = [\pi(t), v(t), f(t), g(t)], t \geq 0$, there is a unique state trajectory $n(t)$, $t \geq 0$, that satisfies (1)-(4).

---

[6] Equation (4) also ties together the time and space discretization parameters $\tau$ and $L(i)$. Since the maximum speed is $V$, the maximum value of the right hand side of (4) is $V \times \tau / L(i)$. This ratio must be less than one.

## 3 TMC plans and achievable flows

We call $u(t) = [\pi(t), v(t), f(t), g(t)]$, $t \geq 0$, a TMC *plan*. By choice of this plan, the TMC controls the traffic flow. In this section we study the flows or throughput that TMC plans can achieve.

**Feasibility constraint** A trajectory-plan $(n(t), u(t))$ must satisfy two physical constraints

$$\begin{aligned} n(i,t,\theta) &\geq 0, && (6) \\ \sum_\alpha \sum_\theta \pi(\alpha,i,t,\theta) n(i,t,\theta) \lambda(\alpha) &\leq L(i). && (7) \end{aligned}$$

The non-negativity requirement (6) is clear. Constraint (7) expresses the requirement that there is enough space in the section safely to carry out the activities assigned by the plan.

There are, in addition, three constraints dealing with entry and exit. First, vehicles of certain types may not be allowed to enter or exit from certain sections. This constraint is of the form

$$f(i,t,\theta) \equiv 0, \text{ or } g(i,t,\theta) \equiv 0,$$

for all $t$ and for specified values of $i, \theta$.

Second, suppose that a vehicle's entry and exit is encoded in its type, i.e., $\theta$ is of the form $\theta = (\eta, j, k)$ where $j$ is the entry section and $k$ is the exit section. Then vehicles of type $(\eta, j, k)$ can enter only from section $j$. That is,

$$f(i,t,(\eta,j,k)) \equiv 0, \ i \neq j.$$

Similarly, vehicles of type $(\eta, j, k)$ exit only from section $k$. That is,

$$g(k,t,(\eta,j,k)) = [1 - \rho(k-1,t)] n(k-1,t,(\eta,j,k)),$$

or, equivalently,

$$n(k,t,(\eta,j,k)) \equiv 0.$$

Lastly, we may require that when a vehicle of type $(\eta, j, k)$ enters, it must first carry out an entry activity. If this activity is labeled $\alpha_{in}$, the requirement may be expressed as $\pi(\alpha_{in}, j, t, (\eta, j, k)) = 1$, or $\pi(\alpha, j, t, (\eta, j, k)) = 0$ for $\alpha \neq \alpha_{in}$. Other maneuver restrictions can be expressed in a similar way.[7]

All these constraints can more generally and more uniformly be expressed by specifying three subsets $T_f, T_g$ and $T_n$ of section-type pairs, and one subset $T_\pi$ of activity-section-type triples, and the requirement that for all $t$,

$$\begin{aligned} f(i,t,\theta) &= 0, \text{ for all } (i,\theta) \in T_f, && (8) \\ g(i,t,\theta) &= 0, \text{ for all } (i,\theta) \in T_g, && (9) \\ n(i,t,\theta) &= 0, \text{ for all } (i,\theta) \in T_n, && (10) \\ \pi(\alpha,i,t,\theta) &= 0, \text{ for all } (\alpha,i,\theta) \in T_\pi. && (11) \end{aligned}$$

---

[7] For example, one may require that vehicles of a particular type must execute maneuver $\alpha_1$ in section $i_1$, $\alpha_2$ in section $i_2$, and so on.

We will say that a trajectory-plan $(n, u)$ is *feasible* if the constraints (6)–(11) are satisfied. To prevent trivial cases we will not allow $f(i, t, \theta)$ and $g(i, t, \theta)$ both to be positive, by insisting that every $(i, \theta)$ is either in $T_f$ or in $T_g$.

We note some properties of feasible trajectories that will be used to define achievable flows.

**Fact 2** There is a uniform bound which applies to all feasible trajectory-plans.

Let $(n(t), u(t))$, $t = 0, 1, ...$ be a feasible trajectory-plan. Summing (1) over $i$, and cancelling some terms, gives

$$\sum_{i=1}^{I+1} n(i, t+1, \theta) = \rho(I+1, t)n(I+1, t, \theta) + \sum_{i=1}^{I+1} n(i-1, t, \theta) + \sum_{i=1}^{I+1} f(i, t, \theta) - \sum_{i=1}^{I+1} g(i, t, \theta).$$

Using the boundary conditions (2), (3) gives

$$\sum_{i=1}^{I}[n(i, t+1, \theta) - n(i, t, \theta)] = \sum_{i=1}^{I}[f(i, t, \theta) - g(i, t, \theta)].$$

Summing over $t = 0, 1, ..., T-1$ and dividing by $T$ gives

$$\frac{1}{T}\sum_{i=1}^{I}[n(i, T, \theta) - n(i, 0, \theta)] = F(T, \theta) - G(T, \theta),$$

where

$$F(T, \theta) := \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{I} f(i, t, \theta), \ G(T, \theta) := \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{I} g(i, t, \theta),$$

are, respectively, the average number of vehicles of type $\theta$ that enter and leave the AHS during $t = 0, ..., T-1$. It follows from Fact 2 that

$$\lim_{T\to\infty} F(T, \theta) - G(T, \theta) = 0. \tag{12}$$

**Definition** A vector $F = \{F(\theta)\}$ of flows is *achievable* if there is a feasible trajectory-plan and a sequence of times $T_k \to \infty$, such that

$$\lim_{k\to\infty} F(T_k, \theta) = \lim_{k\to\infty} G(T_k, \theta) = F(\theta), \quad \text{for all } \theta. \tag{13}$$

A feasible trajectory-plan $(n(t), u(t))$, $t = 0, 1, ...$ is *stationary* if the sequence $(n(t), u(t))$ does not depend on $t$.

**Theorem 1** Every achievable flow can be realized by a stationary plan which, moreover, minimizes travel time.

**Theorem 2** The set of achievable flows is convex and compact.

The next result is intuitively obvious.

**Fact 3** If $F$ is achievable and if $0 \le H(\theta) \le F(\theta)$, then $H$ is achievable.

# 4 Capacity and optimal plans

We first show that the set of achievable flows is a convex polygon.

**Fact 4** $\{F(\theta)\}$ is achievable if and only if there exist stationary flows $\phi(i, \theta)$, a trajectory $\{n(i, \theta)\}$, and plans $\{f(i, \theta), g(i, \theta), \pi(\alpha, i, \theta)\}$, all of them non-negative, such that the following linear constraints hold:

$$\begin{aligned}
\phi(i, \theta) &= \phi(i-1, \theta) + f(i, \theta) - g(i, \theta), && (14)\\
\phi(0, \theta) &= 0, && (15)\\
\phi(I+1, \theta) &= 0, && (16)\\
n(i, \theta) &= \frac{\phi(i, \theta) \times L(i)}{V \times \tau}, && (17)\\
\sum_\alpha \sum_\theta \pi(\alpha, i, \theta) n(i, \theta) \lambda(\alpha) &\leq L(i), && (18)\\
\sum_\alpha \pi(\alpha, i, \theta) &= 1, && (19)\\
f(i, t, \theta) &= 0, \text{ for all } (i, \theta) \in T_f, && (20)\\
g(i, t, \theta) &= 0, \text{ for all } (i, \theta) \in T_g, && (21)\\
n(i, t, \theta) &= 0, \text{ for all } (i, \theta) \in T_n, && (22)\\
\pi(\alpha, i, t, \theta) &= 0, \text{ for all } (\alpha, i, \theta) \in T_\pi. && (23)
\end{aligned}$$

Constraint (17) is linearized by the nonlinear transformation $p(\alpha, i, \theta) = \pi(\alpha, i, \theta) n(i, \theta)$.

**Definition** An achievable flow $F = \{F(\theta)\}$ is *undominated* if $G = F$, for any achievable flow $G$ with $G(\theta) \geq F(\theta)$ for all $\theta$. The *capacity* of the AHS is the set of all undominated flows. See Figure 1. A trajectory-plan is *efficient* if it minimizes travel time.

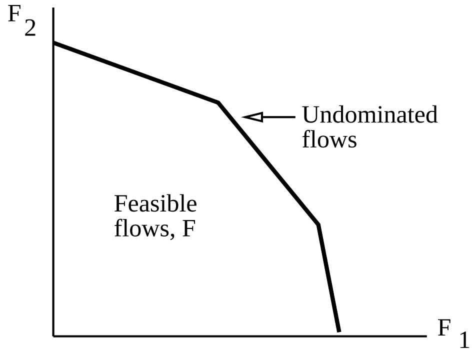

Figure 1: The convex polygon is the set of feasible flows. The bold boundary is the set of undominated flows.

**Theorem 3** A flow $F^*$ is undominated if and only if it is the optimal solution of the linear programming problem:

$$\begin{array}{ll} \max & \sum w(\theta) F(\theta) \\ \text{subject to} & \text{constraints (14)–(23)} \end{array}$$

for some weights $w(\theta) \geq 0$, not all zero. Moreover, the optimal solution yields an efficient pair that achieves $F^*$.

## 5 Transient behavior and TMC rules

A TMC plan specifies activities, speed, entry and exit flows in each section and for all times. The plan is specified ahead of time, with no measurement of the traffic state. (In control engineering, this is said to be an "open loop" specification.) Open loop specifications are very useful for analytical study but they should not be implemented in practice. This is because the state equation model (1) is an idealization which ignores the uncertainty in model parameters and the presence of random fluctuations. These departures from idealization cause the actual traffic trajectory to be different from the open loop trajectory predicted by the model.

It is, therefore, much to be preferred to design a TMC plan in the form of a (feedback) rule. The rule gives the plan values at time $t$ as a function of the state $n(t)$ at that time. A rule can be evaluated by its steady state and transient behaviors. A well-designed rule would achieve capacity and minimum travel time in the absence of fluctuations, independent of the initial state; and small fluctuations would cause small departures of the achieved flow from capacity.

Since a rule specifies the plan as a function of the state, implementation of the rule requires sensors that measure the state, and communicating measurements to appropriate locations where the plan is computed. A rule requiring fewer state measurements is, everything else equal, preferable to one that requires more measurements. A rule in which a plan for section $i$ requires state measurements in sections near $i$, is preferable to one in which it requires measurements in sections remote from $i$, because the former will require less communications hardware.

We illustrate some of the issues using the example of Figure 2. The figure shows two trajectory-plan pairs. The highway configuration is as follows. Each section is 100 m long. There is only one entry (in section 1) with flow $f$, and one exit (in section $I$) with flow $g$. There are two activities. Activity 1 must be carried out in all sections except $I$ and activity 2 (the exit activity) must be carried out in section $I$. $\lambda(1) = 10$ m, $\lambda(2) = 20$ m. The maximum speed is 10 m per unit time. Section $I$ is a "capacity bottleneck." At most 5 vehicles can be accommodated in section $I$, and so the maximum value of $g$ is 0.5. Hence the highway capacity is 0.5 vehicles per unit time.

Both trajectory-plan pairs in the Figure achieve the capacity. In the upper pair, the speed is 10, so the travel time is minimized. In the lower pair, the speed is 5, so the travel time is twice the minimum.

A rule must specify the speed in each section, and $f$, $g$ in the sections 1 and $I$ respectively. The rule for the last section $g$ is obvious: $v(t) = 10$, and $g(t) = [1 - \rho]n(I, t)$. A reasonable speed rule for all other sections is to have the maximum possible speed (up to 10). Of course, what the maximum speed in section $i$ turns out to be at any time depends on the space available in section $i + 1$. If the state $n$ is as shown in the lower part of Figure 2, the

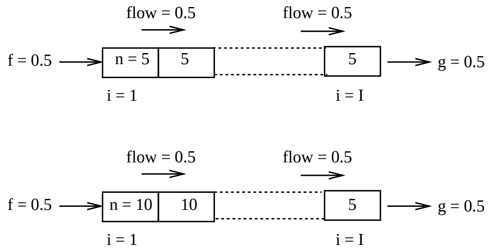

Figure 2: Both trajectory-plan pairs achieve the maximum flow of 0.5. The upper pair minimizes travel time; the lower pair doubles travel time.

maximum possible speed is 5; if it is as in the upper part, the maximum speed is 10.

**Entry and Velocity Control Policies** The example motivates the need for policies both for velocity and entry in order to achieve the maximum achievable flow, while not exceeding the space limit in each section. We will specify "greedy" policies for velocity and the entry flow $f$ and show that they achieve the maximum steady-state flow.

Let us simplify notation by eliminating indices for $\theta$ and $\alpha$. Define $n(i,t)$, the total number of vehicles in section $i$ as

$$n(i,t) = \sum_{\theta} n(i,t,\theta)$$

and $\pi(\alpha,i,t)$, the proportion of vehicles performing activity $\alpha$ as

$$\pi(\alpha,i,t) = \frac{\sum_{\theta} \pi(\alpha,i,t,\theta) n(i,t,\theta)}{\sum_{\theta} n(i,t,\theta)} \ .$$

Then $\lambda(i)$, the average space used per vehicle in section $i$, is

$$\lambda(i) = \sum_{\alpha} \lambda(\alpha)\pi(\alpha,i,t) \ .$$

$\lambda(i)n(i,t)$ is the space used by vehicles in section $i$. Also, the maximum number of vehicles in section $i$, $N(i)$ is given by

$$N(i) = \frac{L(i)}{\lambda(i)} \ .$$

Using this notation the appropriate expression for velocity in section $i-1$ is

$$v(i-1,t) = \min\{V, \ \frac{L(i)L(i-1)}{\tau n(i-1,t)\lambda(i)} - [1 - \frac{v(i,t)\tau}{L(i)}]\frac{n(i,t)L(i-1)}{\tau n(i-1,t)}\} \tag{24}$$

The flow out of section $i$ is

$$\begin{aligned} \phi(i,t) &= [1-\rho(i,t)]n(i,t) \\ &= \frac{v(i,t)\tau}{L(i)} n(i,t), \end{aligned}$$

while the maximum flow $\overline{\phi}(i)$ is

$$\overline{\phi}(i) = \frac{V\tau}{\lambda(i)} \ .$$

We will need the minimum of these flows to prove existence of an equilibrium solution of the flows; therefore, we make the following definition.

**Definition** $\underline{\phi}$ is the minimum of the maximum possible flow out of any section or

$$\underline{\phi} = \min_i \frac{V\tau}{\lambda(i)} = \min_i \overline{\phi}(i).$$

Section $I$ is the "bottleneck," i.e., $\underline{\phi} = \overline{\phi}(I)$ and $\underline{\phi} < \overline{\phi}(i)$ for $i \neq I$.

**Theorem 5** Assume the velocity policy (24) is applied and $v(I,t) \equiv V$, then for every $i$ and $t$, either $v(i,t) = V$ or $\phi(i,t) \geq \underline{\phi}$ .

It only remains to find a rule for controlling entry, i.e., $f$. As above, we propose a greedy policy for $f$ that fills the available space in section 1. We assume there is no limit on $f$ so the first section will remain filled after $t = 0$. One can easily check that the rule for $f$ is

$$f(t) = \frac{L(1)}{\lambda(1)} - n(1,t) + \frac{v(1,t)\tau}{L(1)} n(1,t) \ . \tag{25}$$

**Corollary 1** Using (25) as the rule for $f$ and (24) as the rule for $v$, $f(t) \geq \underline{\phi}$ for all $t$.

**Fact 5** If at time $t$ section $i$ is full, i.e., $n(i,t) = N(i)$, then $\phi(i,t) \geq \underline{\phi}$.

**Fact 6** If $n(i,t) = N(i)$ and $\phi(i,t) < \overline{\phi}(i)$ for all $t$, then $n(i+1,t) = N(i+1)$ and $\phi(i+1,t) \geq \underline{\phi}$ for all $t$.

**Theorem 6** Using the greedy policies (24) and (25) for $v$ and $f$, respectively, and assuming $v(I,t) = V$ for all $t$, $f$, $g$, $n$ and $v$ converge to a unique equilibrium solution for (1), i.e., as $t \to \infty$

$$\begin{aligned} f(t) &\to \underline{\phi} \\ g(t) &\to \underline{\phi} \\ \phi(i,t) &\to \underline{\phi} \\ n(i,t) &\to N(i) \\ v(i,t) &\to \frac{\underline{\phi} L(i)}{N(i)\tau} \ . \end{aligned}$$

As a final note observe that the information needed for the greedy velocity policy can be obtained from vehicle-borne sensors and requires no extra sensor information from the roadside. The policy can be implemented by a vehicle longitudinal control law that tracks $V$ while maintaining a safe distance from the vehicle ahead.

## 6 Entry and exit

An automated highway will make contact with a non-automated highway at points of entry and exit. In current design proposals [8], a "transition area" serves as interface between the two highways where vehicles undergo "check-in" and "check-out" and where vehicle control is transferred from driver to system upon entry to the AHS and from system to driver upon exit. We call these two activities "entry" and "exit."

Automation of these activities is a complex task. A vehicle entering the AHS must negotiate its passage through the transition area and coordinate its entry with vehicles on the automated lane. If this coordination is poor, there will be congestion at the entrance, slowing down upstream vehicles. A vehicle leaving the AHS may similarly disrupt traffic, thereby reducing capacity. By contrast, in between entry and exit, traffic on the automated lane should proceed very smoothly. Thus, it seems that AHS capacity and transient behavior are likely to be limited by the entry and exit activities. In this section we will formulate a micro-level queuing model for entry and show how the space occupied by the entry activity may determine the capacity of the highway. Then we show that the amount of delay incurred by upstream vehicles due to an entering vehicle depends on the sophistication of the feedback control law that implements entry.

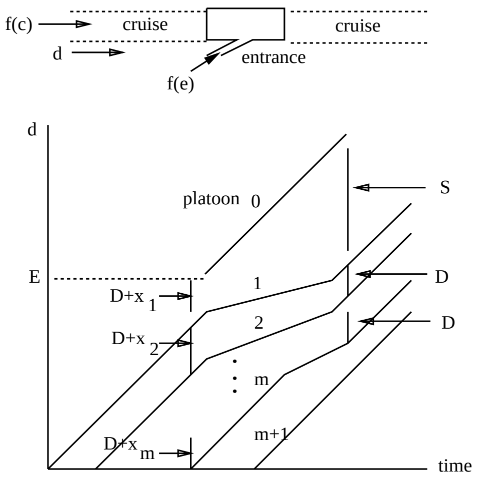

Figure 3: There is one entry in a long highway. The trajectories show how entry of platoon #0 slows down platoons $\#1, ..., \#m$.

Figure 3 shows a long automated lane, with one entrance. Distance along the highway is denoted by $d$, and the entrance is located at $d = E$. Vehicles are organized in platoons of closely spaced vehicles. (For simplicity assume that platoons have a fixed number of vehicles.) Platoons can engage in two activities: cruise and entry, with $\lambda(cruise) = D$ meters ($D$ does not include the platoon length) and $\lambda(entry) = S$ meters, with $S > D$. The maximum speed is $V$ meters/hour. Let $f(c)$ denote the number of platoons per hour that come cruising from upstream of the entrance; and let $f(e)$ be the flow of entering platoons. An entering platoon must first engage in the entry activity; it then switches to cruise.

We want to compute the achievable throughput vectors $F = (f(c), f(e))$. By Theorem 1, we may assume that a stationary trajectory-plan achieves $F$, with platoons traveling at maximum speed $V$. Let $L$ be the length of the entry section, so a platoon stays in this section for time $L/V$ hours. Hence the number of cruising platoons in this section is

$n(c) = f(c) \times L/V$, and the number of entering platoons is $n(e) = f(e) \times L/V$. The space constraint is $D \times n(c) + S \times n(e) \leq L$, or

$$D \times f(c) + S \times f(e) \leq V,$$

so the capacity of this AHS is the set of all vectors $F = (f(c), f(e)) \geq 0$ that satisfy

$$D \times f(c) + S \times f(e) = V. \tag{26}$$

This capacity estimate is optimistic. The estimate is based on our model which assumes that the inter-platoon distance among the cruising platoons is distributed in such a way that a gap of size $S$ meters appears every time a platoon is about to enter. This requires perfect coordination between the cruising platoons and the entry platoons. If this perfect coordination is lacking, then the cruising platoons will be forced to slow down in order to create the needed gap of $S$ meters for an entering platoon, resulting in an increase in total travel time. In order to estimate the total delay, we need to know the distribution of inter-platoon distances. We will assume a random distribution.

Suppose that the inter-platoon distances are iid (independent, identically distributed) random variables, denoted $z$. The cruise control law guarantees that $z \geq D$ (the safe cruising distance) with probability 1, and we assume that $x := z - D$ is an exponentially distributed random variable with mean $\mu^{-1}$, i.e., $x$ has the probability density

$$p(x) = \mu e^{-\mu x}, \;\; x \geq 0.$$

For convenience, also denote $p_1(x) \equiv p(x)$.

Suppose that a platoon enters at some time $t$ at distance $E$. This is platoon #0 in Figure 3. (Note: in the figure, platoons are indicated by points.) Number the cruising platoons that follow #0 by #1, #2, ...., and the distance between the end of platoon $\#i-1$ and the beginning of platoon $\#i$ by $z_i = D + x_i$. If $x_1 < S$, then platoon #1 will have to slow down until it creates a distance of $S$; if $x_1 + x_2 < S$, then #2 will have to slow down, too, and so on. This "shock wave" will affect a random number $M$ of platoons, where

$$M = m \Leftrightarrow \{\sum_1^m x_i \leq S < \sum_1^{m+1} x_i\}.$$

We want to calculate the statistics of $M$, and the amount of slowdown.

It will be convenient to consider the distribution of $\sum_1^n x_i$,

$$p_n(x) := p(\sum_1^n x_i = x) = \mu^n \frac{x^{n-1}}{(n-1)!} e^{-\mu x}, \;\; x \geq 0. \tag{27}$$

So the probability that $M = m$, i.e. $m$ platoons will be disturbed, is given by $P_S(m) = \text{Prob}\{\sum_1^m x_i \leq S < \sum_1^{m+1} x_i\}$. One can calculate the probabilities $P_S(m)$ from the $p_n$ by observing that

$$P_S(m) = \int_0^S p_1(x_1 \geq S - y) \times p_m(y) dy.$$

A little calculus then gives the following formula:

$$P_S(m) = e^{-\mu S}\frac{(\mu S)^m}{m!} = P_S(m-1) \times \frac{\mu S}{m},\ m = 0, 1, ... \tag{28}$$

Equation (28) is the formula for a Poisson distribution. Thus the number $M$ of platoons disturbed by the deviation $S$ has a Poisson distribution. In particular, the mean number of disturbed (or delayed) platoons is $EM = \mu S$. If we write the mean inter-platoon distance as $Z := Ez$, and recall the definition $\mu^{-1} = Ex = E(z - D)$, we conclude that

$$\text{Average number of delayed platoons} = EM = \frac{S}{Z - D}. \tag{29}$$

Observe that the average flow of cruising platoons is $f(c) = V/Z$, whose maximum value is $V/D$. As expected, (29) implies that as $Z \to D$, $EM \to \infty$, i.e., as the flow of cruising platoons increases, the shock wave from each entering platoon passes through an increasing number of platoons, on average. Another interesting point in (29) is that the average number of delayed platoons grows *linearly* with the size of the safe entry distance, $S$.

We can now calculate the total delay incurred by upstream traffic due to the entering platoon, platoon #0. The entering platoon will require $S$ meters; however, if the entering platoon encounters a free space gap, then the actual space $B$ "borrowed" from the upstream cruise platoons will be between 0 and $S$. We will consider the probability distribution of $B$ after first examining the case of a fixed space $S$.

In order to create a gap of $S$ meters, platoons #1, ..., #$M$ are slowed down, where $M$ is the random variable above. Platoon #$i$ is slowed down by a distance

$$S - \sum_{j=1}^{i}(z_j - D) = S - \sum_{j=1}^{i} x_j,\ i = 1, ..., M.$$

So the total slowdown $\delta$ (measured in platoon × meters) is the sum of these $M$ numbers,

$$\text{slowdown} := \delta = \sum_{i=1}^{M}[S - \sum_{j=1}^{i} x_j] = MS - \sum_{i=1}^{M}\sum_{j=1}^{i} x_j. \tag{30}$$

We want to calculate $E\delta$, the average slowdown.

**Fact 7** Each entering platoon on average disturbs $S/(Z - D)$ platoons and they suffer a total slowdown of $S^2/2(Z - D)$ platoon-meters.

## 7 Steady-state capacities

We consider two alternative designs. We call one design the platoon organization or PO design ([2]). We call the second the adaptive cruise control or ACC design ([10]).

**PO design** There are five activities in the PO design: merge, split, 15 vehicle platoon, entry, and exit. We will determine the steady-state capacity of an automated lane with these activities. We first specify the lane configuration. The lane consists of sections of equal length $L$. There are three types of sections. In entry sections *entry* and *platoon*15 are allowed; in exit sections *exit* and *platoon*15 are allowed; and all other sections, called cruise sections, either *platoon*15, *merge*, or *split* are allowed. (In a merge maneuver, one platoon first accelerates and then decelerates to join the platoon in front of it; in split, the rear of one platoon first decelerates and then accelerates to form two platoons.)

In order to calculate steady-state capacities, it is necessary to determine the space requirement for each activity, to specify the composition of activities in each section, and to find the section with strictest space limit which determines the maximum flow.

We specify some physical and design parameters. $D$ is the safety distance maintained by the leaders of platoons, $d$ is the inter-vehicle spacing within a platoon, $l$ is the vehicle length, $V$ is the maximum speed, $n$ is the platoon size, $Q$ is the range of the longitudinal sensor, $a_{min}$ is the maximum vehicle deceleration, $a_{max}$ is the maximum vehicle acceleration. Representative values used in the PO design are $L = 400$m, $\tau = 10$ seconds, $D = 60$m, $d = 1$m, $l = 5$m, $V = 25$m/s, $n = 15$, $Q = 60$m, $a_{max} = 2$m/s/s, and $a_{min} = -2$m/s/s.

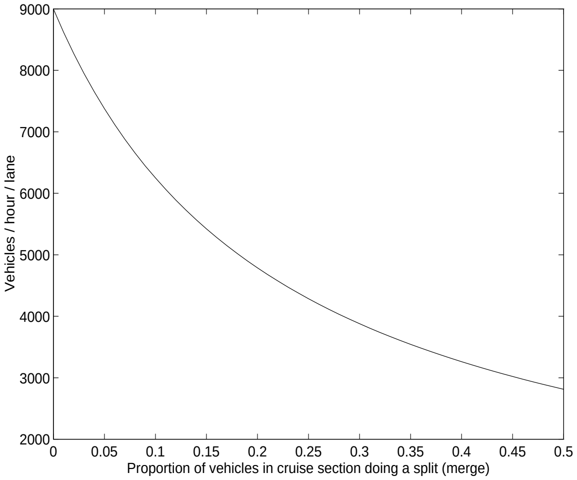

Figure 4: Maximum flow in cruise sections as a function of the proportion of vehicles doing splits (merges).

The space requirement for *entry* is $s(t) = D + l = 65$m, so $\lambda(entry) = 65$m; also, $\lambda(exit) = 65$m. The space requirement for *platoon*15 is

$$s(t) = \frac{d(n-1) + nl + D}{n}$$

or 10m, so $\lambda(platoon15) = 10$m.

The space requirement for *merge* requires some calculation. We assume that the merge is initiated by one vehicle when the platoon ahead is within the vehicle's sensor range $Q$. The relative velocity and acceleration between the two cars is zero. The merging vehicle accelerates up to $a_{max}$ while keeping a safe relative distance and velocity from the car ahead. The maneuver ends when the vehicle is within distance $d$ m of the platoon ahead. If the activity lasts for less than the duration of time that the vehicle is in the section, some extra space must be allotted. Two maneuvers constitute this activity:

$$\begin{aligned} s(t) &= f(\Delta v, \Delta a, Q) && ; \quad t_0 \le t \le t_1 \\ s(t) &= \frac{d(n-1) + nl + D}{n} && ; \quad t_1 \le t \le t_2 \, . \end{aligned}$$

$\Delta v$ is the relative velocity of the two vehicles at the beginning of the merge, and $\Delta a$ is the relative acceleration of the two vehicles at the beginning of the merge. $t_1$ is the time when the merging vehicle is within $d$ m of the vehicle ahead. $t_2$ is the time when the vehicle crosses the section.

Using the equations for the safe merge developed in [7] we obtain a space requirement of 27m with a duration of $t_1 = 16$s. To this we add the length of the vehicle $l$. Thus, $\lambda(merge) = 32$m. Also the time to cross the section is $t_2 = t_1 = 16$s. A similar exercise for *split*, which takes a vehicle from $d$ m to $D$ m from the platoon ahead and uses $a_{min}$ for deceleration yields $\lambda(split) = 32$m and takes 16s.

We must define the proportion of activities in each section. $\pi_e$ ($\pi_x$) is the proportion of vehicles doing *entry* (*exit*) in an entry (exit) section, $\pi_m$ ($\pi_s$) is the proportion of vehicles doing *merge* (*split*) in a cruise section, $\pi_c$ is the proportion of vehicles doing *platoon*15 in a cruise section, and $\pi_p$ is the proportion of vehicles doing *platoon*15 in an entry or exit section. There are some constraints on the proportions:

$$\begin{aligned} \pi_e &= \pi_x, \\ \pi_m &= \pi_s, \\ \pi_p + \pi_e &= 1, \\ \pi_c + 2\pi_s &= 1. \end{aligned}$$

Using these constraints, calling the flow $f$, and substituting values for $\lambda(\alpha)$, the space constraint for entry/exit sections is

$$[65\pi_e + 10(1 - \pi_e)]f = 25.$$

The space constraint for cruise sections is

$$[10(1 - 2\pi_s) + 32\pi_s + 32\pi_s]f = 25.$$

If we set $\pi_e = .1$ and $\pi_s = .1$, the limiting section is the entry or exit section, and the maximum flow is $f = 5806$ vehicles/hr.

Suppose we keep $\pi_e$ fixed but vary $\pi_s$ between 0 and 0.5. The constraint on the flow due to the entry (exit) sections is 5806 vehicles/hr. The constraint due to the cruise section as $\pi_s$ is varied is shown in Figure 4.

**ACC Design** In this design, some of the vehicles are manually driven, and the rest are under adaptive cruise control. So there are four activities: automatic cruise, manual cruise, manual entry, and manual exit. The lane consists of entry, exit and cruise sections. In entry (exit) sections, automatic cruise, manual cruise and entry (exit) are allowed. In cruise sections, automatic and manual cruise are allowed.

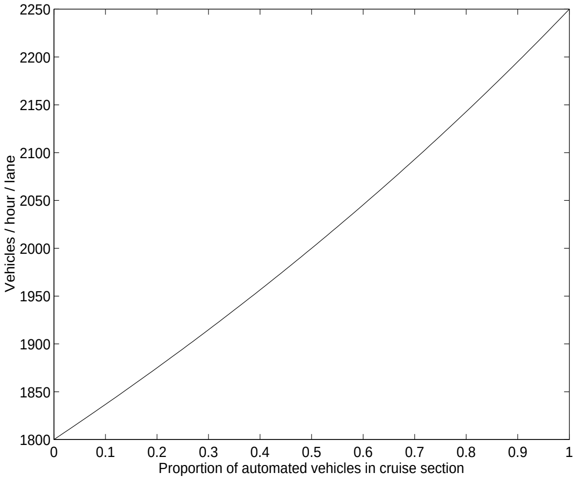

Figure 5: Maximum flow in cruise sections as a function of the proportion of automated vehicles.

The space requirement for manual entry is $\lambda(entry) = D + l = 65$m. The requirement for manual exit is $\lambda(exit) = D + l = 65$m. The requirement for manual cruise is $\lambda(mc) = 2V = 50$m. The requirement for automatic cruise is $\lambda(ac) = V + l + 10 = 40$m. The only constraint on activity proportions is $\pi_e = \pi_x$, and we set $\pi_e = .1$. Now we write the space constraint for the three types of sections. For entry (exit) sections

$$[(1 - \pi_{ac} - .1)50 + \pi_{ac}40 + 6.5]f = 25.$$

For cruise sections

$$[(1 - \pi_{ac} - .1)50 + \pi_{ac}40]f = 25.$$

We can now compute the capacity. If for example, the proportion of automated vehicles is 0.5, then the maximum flow in an entry (exit) section is 1935.5 vehicles/hr. Figure 5 shows the increase in capacity as the proportion of automated vehicles in a cruise section increases.

## 8 Conclusions

We have presented a theory for automated traffic flow, based on the notion of vehicle activities. An activity is a sequence of vehicle maneuvers executed by vehicle control laws. The space that it takes up is the abstraction used to represent an activity in the traffic flow model. A plan is defined as the proportion of activities, velocity, entry flow and exit flow in each section. The TMC controls the flow by selection of this plan. We showed that achievable flows can be realized by stationary plans, and maximal achievable flows are obtained by solving a linear programming problem.

These are results about steady-state conditions. However, since conditions may vary over time, perhaps because of incidents, one should use adaptive policies for the entry flow and velocity. We proposed one such policy: the greedy policy attempts to fill up the free space in the next section as quickly as possible. We showed that the greedy policy maximizes steady-state flow, although it does not minimize travel time.

The proposed theory can be compared with the theory of manual traffic flow. The safety-needs-space assumption makes space the crucial resource in our model, and in a one lane highway, the maximum flow is determined by the most space-constraining section. This insight holds for a network of highways, and the Ford-Fulkerson theorem can be used to relate the maximal or undominated flows with the most constraining sections. The insight is equally valuable in manual traffic. Perhaps the only important distinction is that in manual traffic the "consumption" of space by vehicles has a negative externality, because drivers interact. This interaction between vehicles is absent in our model.

The usefulness of the proposed theory must be judged by its ability to open up for investigation related questions and in application. We expect to report progress in both directions in future papers.

## References

[1] R. Hall, "Longitudinal and lateral throughput on an idealized highway," *Transportation science*, vol. 29, pp. 3499–3505, May 1995.

[2] P. Varaiya, "Smart cars on smart roads: Problems of control," *IEEE Transactions on Automatic Control*, vol. 38, pp. 195–207, February 1993.

[3] C. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research*, vol. 28B, pp. 269–287, August 1994.

[4] S. Sheikholeslam and C. A. Desoer, "Longitudinal control of a platoon of vehicles," in *Proceedings of the 1990 American Control Conference, Volume 1*, (San Diego, CA), pp. 291–296, June 1990.

[5] D. McMahon, J. Hedrick, and S. Shladover, "Vehicle modeling and control for automated highway systems," in *Proceedings of 1990 American Control Conference*, (San Diego, CA), pp. 297–303, 1990.

[6] H. Peng and M. Tomizuka, "Lateral control of front-wheel-steering rubber-tire vehicles," tech. rep., UCB-ITS-PRR-90-5, Institute of Transportation Studies, University of California, Berkeley, CA 94720, 1990.

[7] J. Frankel, L. Alvarez, R. Horowitz, and P. Li, "Safety-oriented maneuvers for IVHS," in *Proceedings of the American Control Conference*, June 1995.

[8] D. Godbole, F. Eskafi, E. Singh, and P. Varaiya, "Design of entry and exit maneuvers of IVHS," in *Proceedings of the American Control Conference*, June 1995.

[9] B.S.Y. Rao and P. Varaiya, "Roadside intelligence for flow control in an IVHS," *Transportation Research-C*, vol. 2, no. 1, pp. 49–72, 1994.

[10] P. Ioannou and Z. Xu, “Throttle and brake control systems for automatic vehicle following,” *IVHS Journal*, vol. 1, no. 4, pp. 345–377, 1994.