UC Berkeley
Recent Work

Title
Deploying Lanes for High Occupancy Vehicles in Urban Areas

Permalink
https://escholarship.org/uc/item/6r52d95p

Authors
Cassidy, Michael J.
Daganzo, Carlos F.

Publication Date
2007-04-11

Deploying Lanes for High Occupancy Vehicles in Urban Areas
Carlos F. Daganzo and Michael J. Cassidy

UC Berkeley Center for Future Urban Transport
A **VOLVO** Center of Excellence

DEPLOYING LANES FOR HIGH OCCUPANCY VEHICLES IN URBAN AREAS

Carlos F. Daganzo and Michael J. Cassidy
Institute of Transportation Studies
University of California, Berkeley, CA 94720

(February 14, 2007)

ABSTRACT

Simulations and field experiments in previous works suggest that a freeway's general purpose lanes (those not dedicated to high occupancy vehicles) discharge vehicles from bottlenecks at an equal or higher average rate when one of the lanes is devoted to high occupancy vehicles than when it is not. This result was used in these previous works to develop formulae for the total discharge rate of bottlenecks, with and without dedicated lanes, as a function of the percentage of high occupancy vehicles in the traffic stream.

This present paper extends these ideas by examining the effect of dedicated lanes on the density of traffic queues. We find that an underutilized dedicated lane reduces a queue's density (in vehicles per km of freeway) when the downstream flow of both high occupancy and low occupancy vehicles is the same in both scenarios and exogenously determined; e.g., as would happen if the queue's service rate is dictated by recurrent downstream congestion. A formula is given; and the reduction in density turns out to be small if the underutilization is small.

Reductions in queue density without changes in bottleneck flows or traffic demand imply spatially longer queues, and this could be problematic. The paper also shows that the extra space consumed by a queue adjacent to a dedicated lane can contribute significantly to congestion, but only if heavily traveled routes that do not go through the bottleneck pass through this extra space. To quantify this effect, the paper analyzes dedicated lanes on multi-ramp freeways and beltways. Formulae are given for the changes in the people-hours and vehicle-hours of travel due to dedicated lanes both, when there is uncongested freeway space upstream of the queue for it to expand, and when there is not. The recipes are based on readily observable data and can be used to evaluate existing and planned installations of dedicated lanes. Building on these formulae, the paper finally presents qualitative principles that can be used to plan city-wide systems of both, high occupancy vehicle lanes on freeways and dedicated bus lanes on surface streets.

1. INTRODUCTION AND BACKGROUND

High occupancy vehicle lanes (H-lanes) are supposed to give priority to HOV's (high occupancy vehicles) with as little disruption as possible to LOV's (low occupancy vehicles). Converting a general purpose lane (G-lane) into an H-lane can: (a) increase the average number of occupants in each vehicle by encouraging car-pooling; and (b) reduce person-hrs of delay for any given level of car-pooling by allocating most of it to LOV's. Effect (b) is the focus of this paper. If it is achieved, then the average distance traveled *per person* per unit time (a measure of accessibility) increases. As we shall see, a conversion that achieves (b) often does so without increasing vehicle-hours of travel. Since the latter are a proxy for the externalities of traffic, the result is increased accessibility without significant adverse side effects. This paper examines how to achieve this goal for a freeway network. The minimum number of occupants that renders vehicles eligible for an H-lane is treated as a variable because the results for a large minimum then shed light on bus-only lanes.

We will build on recent results that have clarified the interaction between H-lanes and freeway bottlenecks. Cassidy et al (2006) analyzed field data from several congested freeways in the San Francisco Bay Area, and found (surprisingly[1]) that, at these sites, H-lanes were having no discernible negative effects despite their underutilization. At one of the sites, high-resolution video data showed that although the H-lane was underutilized at the bottleneck (by about 35%) this was more than compensated by an increase in the capacity of the remaining G-lanes. This may seem counter-intuitive, but had been theoretically predicted in Menendez and Daganzo (2006). Simulations of a merge bottleneck in that reference revealed that by reserving the median lane for HOVs, fewer lane changes occur near the bottleneck, and that this has a "smoothing effect" on discharge flow. The reference also provided formulas describing how underused H-lanes affect the flow through bottlenecks of various types. To no surprise, the formulas predict that severe underutilization reduces bottleneck flows. The reference also identified a real-world example of this phenomenon.

Daganzo et al (2002), Menendez and Daganzo (2006), and Cassidy et al (2006) have recommended ways in which the problem of bottleneck underutilization can be eliminated. Those references argue that H-lanes can be installed upstream of bottlenecks without reducing

---

[1] An earlier study of the same sites (Chen et al, 2006) had conjectured from less complete data that the H-lanes were causing the congestion.

discharge flow—even in the absence of the smoothing effects—if they are terminated (i.e., opened to all vehicles) short of the bottleneck. A dynamic termination strategy has been found in simulations to enhance the smoothing effect and actually *increase* bottleneck discharge flow; see Menendez and Daganzo (2006).

Bottleneck flow, however, is not the only freeway metric of importance. If a freeway's ability to store vehicles is reduced by an H-lane, freeway queues can grow longer, blocking more off-ramps and reducing the freeway's vehicle outflow. Vehicle delay would then increase, even without a change in bottleneck flow. Longer freeway queues can also block important access points. This is of particular concern in urban areas with closely spaced interchanges. Thus, the effect of H-lanes on freeway storage density requires attention.

Unfortunately, a purely empirical understanding of storage effects would require repeated observations of freeway queues with and without an H-lane. Hence, we shall use instead the model in Menendez and Daganzo (2006), because it is simple and roughly consistent with empirical data. To be conservative, we shall choose parameters unfavorable to H-lanes when a choice has to be made. Our model ignores the smoothing effect, for example.

The paper will examine the effect of two policy scenarios (an H-scenario with an H-lane and a G-scenario without one) on freeway queues, and then generalize the ideas to transport systems on a city- or region-wide scale. Section 2 examines open-ended freeway queues, with uncongested space behind their tails. Section 3 examines close-ended queues that form a loop and have no tail, e.g. on circular beltways. Section 4 then shows how H-lanes should be systematically deployed in a city to increase people accessibility without increasing vehicle-hours of travel and the associated externalities. Section 5 generalizes the ideas and suggests how urban street space (not just freeway space) should be shared with, and/or separately partitioned among, different modes to improve accessibility.

2. OPEN-ENDED FREEWAY QUEUES

Menendez and Daganzo (2006) describes conditions to ensure that a freeway's median lane can be converted into an H-lane without creating bottlenecks or reducing the discharge rate of existing ones. The conditions are applied at every point x along the freeway in the G-scenario. For an L-lane freeway we require:

$$q(x)L - q_H(x) - Q(L-1) \leq 0, \tag{1a}$$

where Q is the capacity (veh/hr-lane) of a lane, $q(x)$ is the flow per lane at x in the G-scenario, and $q_H(x)$ the flow of HOVs that could feasibly be on the H-lane at x had the HOV restriction been activated. (One must recognize when estimating q_H that only thru-moving HOVs can be in the H-lane at any x close to exits and entrances.) The LHS of (1a) is called the "overflow". It is the difference between the flow one would have in the G-lanes in the H-scenario, $q(x)L - q_H(x)$, and the combined capacity of these lanes, $Q(L - 1)$.

The "overflow" condition should hold at the termination point of an H-lane. This ensures that if a queue exists at that location in the G-scenario, activation of the H-lane (upstream of the location only) does not starve the location for flow. Note that (1a) imposes a lower bound on q_H. Of course, a problem also arises if q_H is too high: for the H-lane to remain effective, q_H must not exceed the maximum the H-lane can carry, which is q at the termination point (due to downstream congestion). Thus, we have:

$q_H \leq q$ (at the H-lane termination point, if congested), and $q_H \leq Q$ otherwise. $\qquad$ (1b)

If (1b) were to be violated at some x, the H-lane would become oversaturated and HOVs would not find it attractive. As a result, the freeway would operate in G-mode at x despite the presence of an H-lane.

Conditions (1) guarantee that bottleneck discharge flows do not decrease, but the H-lane can still cause damaging reductions in the queued flows upstream due to the presence of on- and off-ramps. Because this is undesirable, queues should be studied in their entirety with due consideration of the effects of ramps. We start by looking at open-ended queues. Our analysis is incremental: Section 2.1 shows that introduction of an H-lane into a congested freeway stretch that contains a single interchange (an on-ramp/off-ramp pair) does not affect vehicle flows of either type (LOV or HOV) if some mild conditions are met. Sections 2.2 and 2.3 then show how the H-lane affects queue density, and its spatial growth on a stretch containing any number of interchanges. Finally, Sec 2.4 estimates the effect of the H-lane on VHT and PHT.

2.1 A building block: steady-state flows along a single interchange
As in Menendez and Daganzo (2006), we model the freeway as a set of parallel kinematic wave traffic streams on each lane with a triangular fundamental diagram (FD). The simulations in that reference showed that provision of an H-lane tended to: increase the capacity of G-lanes (the smoothing effect); reduce the free-flow speed in the H-lane when the G-lanes are congested; and

reduce slightly G-lane storage density when flows are near capacity.[2] We will thus conservatively assume that the FD of all G-lanes is the same in both scenarios, and that the free-flow speed of the H-lane is lower in the H-scenario. We shall denote the free-flow speed by v (or $v_H < v$ for the H-lane) (km/hr) and the jam-density of a lane by κ (veh/km-lane).

Assume that the freeway stretch containing our single interchange is in a queued steady state in the G-scenario; and that there is no queue on the interchange's on-ramp. Can we convert the freeway's median lane to a faster-moving H-lane, as shown in Fig. 1a, without reducing the HOV and LOV flows past any of the ten screen lines also shown in the figure? To answer this question, consider the freeway segments between consecutive screen lines one at a time.

At line $x = 0$, vehicles leaving the undifferentiated queue segregate themselves by class; between lines 1 and 2 HOV's destined for the off-ramp move onto the G-lanes; between 3 and 4 all exiting vehicles leave; between 5 and 6 on-ramp vehicles merge onto the G-lanes; between 7 and 8 HOV's from the on-ramp merge into the H-lane; and at 9, where the H-lane ends, LOV's merge into the median lane to take up its available capacity. This system can be idealized as the two-commodity network of Fig. 1b, where dotted links are reserved for HOV's and where the flow of each commodity must be conserved at every node. When operating properly the solid links would carry only LOV's and the double-lined links both commodities. There are also capacity constraints imposed by the nodes, as explained below.

Nodes c, e, g and h act as merges with capacities equal to the flow of their respective downstream links, when these links are queued. This capacity is prorated to the inbound links as per the usual priority rules (Daganzo, 1995). Since our interchange is isolated the lane changes in sections 1-2 and 7-8 can be spread over a long distance; thus, we assume that cross-over links bc and fg claim total priority over ac and bg, and have a capacity equal to the (queued) flow of cd and gh, respectively. Also, since node h does not include a lane drop or addition, the capacity of link gh should be the (queued) flow in the median lane downstream of h, which is q; i.e., the H-lane should act as an on-ramp with priority ratio $1/L$. Finally, we assume that if link ef is queued, the on-ramp will introduce into the stream no more than a fraction $\alpha < 1$ of the flow in the shoulder lane of ef; i.e., the on-ramp priority ratio is $\alpha/(L - 1)$.

---

[2] The later effect was not simulated for lower queued flows because trends in the simulated data showed that the effect would not be significant under those conditions.
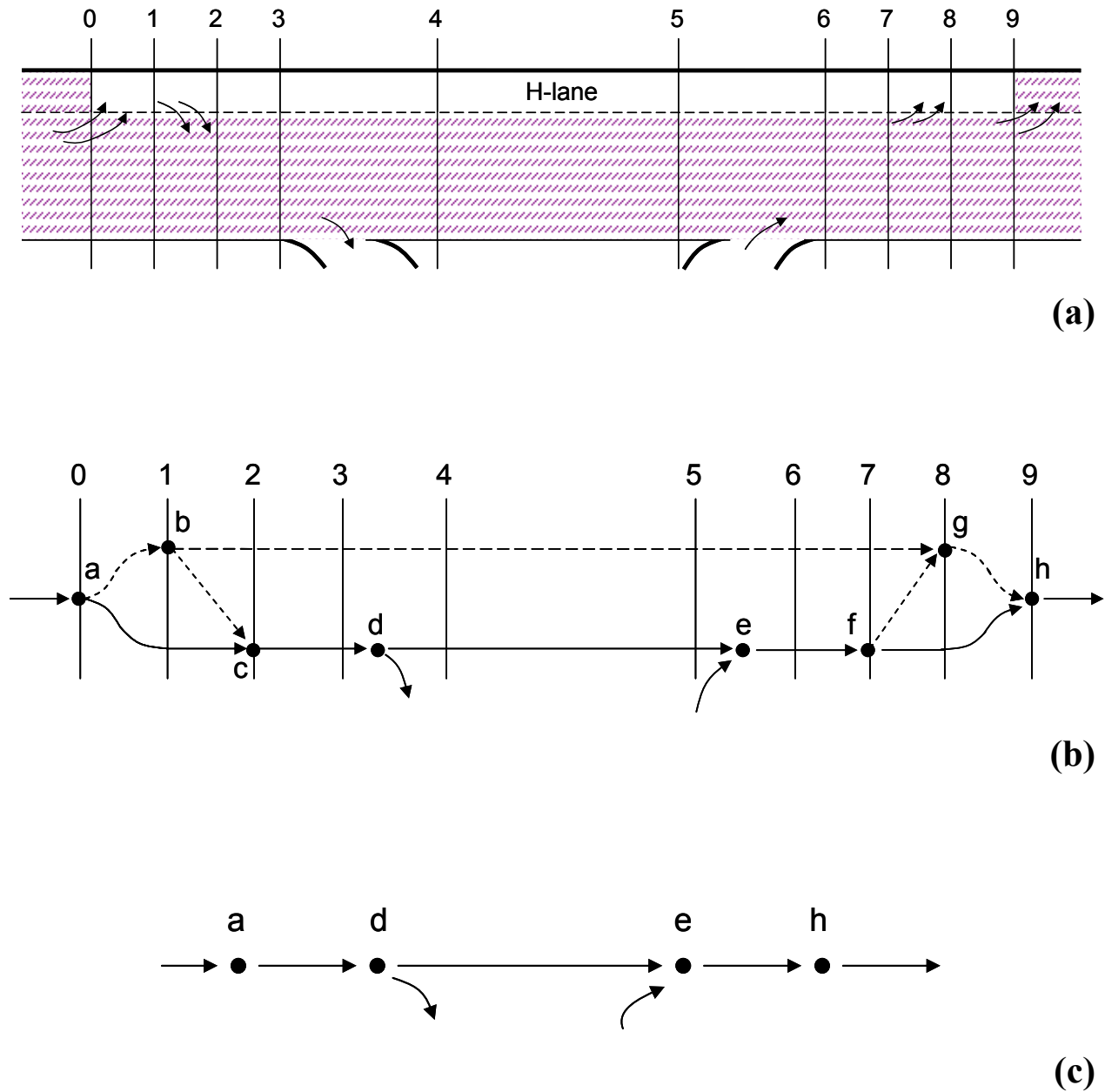
Figure 1. A freeway stretch: (a) H-scenario; (b) network representation of H-scenario; and (c) network representation of G-scenario. Arrows in part (a) denote lane changes across classes of lanes (i.e. among the G-lanes, the H-lane, and the ramps); dark shading denotes congested traffic, which takes the form of a FIFO queue upstream of screen line 0 and downstream of screen line 9; and white shading denotes the uncongested H-lane.

Nodes a, b, d and f operate as FIFO diverges with branch flows determined from their capacities in the usual way (Daganzo, 1995). Branches bc and fg are assumed to have infinite capacity (for the reasons stated above), and the off-ramp is assumed to have the same capacity as that of a freeway lane, Q.

We now check the existence conditions for an equilibrium pattern in our interchange. We start in the G-scenario (Fig. 1c) with a set of commodity flows arising from a steady state OD table: λ_{Had}, λ_{Hah}, λ_{Heh}, λ_{Lad}, λ_{Lah}, λ_{Leh}, where the three subscripts denote (from left to right) vehicle class, origin and destination. To be feasible, the OD flows passing screen line 9 must satisfy: $\lambda_{Hah} + \lambda_{Heh} + \lambda_{Lah} + \lambda_{Leh} = qL$; the off-ramp flows: $q_d \equiv \lambda_{Had} + \lambda_{Lad} \leq Q$; and the on-ramp flow: $q_e \equiv \lambda_{Heh} + \lambda_{Leh} \leq \alpha q$ (since there are no queues on the on-ramp).

To understand what happens in the H-scenario, this OD table is assigned to the shortest paths of Fig. 1b. The input and output flows are as in the G-scenario. Since there is no route choice in either scenario, the commodity flows across screen lines are also as in the G-scenario. However, there are constraints in the network of Fig. 1b that could be violated by the assigned link flows. In what follows we use q_{ij} to denote the flow on the link from node i to node j. One set of constraints ensures that the H-lane remains uncongested; they are: (i) $q_{ab} \leq Q$; (ii) $q_{bc} \leq q_{cd}$; (iii) $q_{bg} + q_{fg} \leq Q$ and (iv) $q_{gh} \leq q$. Another set ensures that the G-lanes and the ramps can carry the required flows; i.e: (v) $q_{cd} \leq (L-1)Q$; (vi) $q_e \leq \alpha q_{ef}/(L-1)$; (vii) $q_{ef} \leq (L-1)Q$; and (viii) $q_d \leq Q$.

Constraints (i), (ii) (iii) and (iv) are met if the demand for the H-lane does not exceed q at the downstream end and if, as is typical, it remains below Q elsewhere upstream; i.e. if:

$$q_H(x) \leq q \quad \text{for } x = 9 \quad \text{and} \quad q_H(x) \leq Q \quad \text{for } x < 9. \tag{2a}$$

This is the same as (1b). Constraints (v) and (vii) are met if the demand for the G-lanes immediately upstream of the off-ramp and immediately downstream of the on-ramp (a demand which would include all the LOVs and all exiting/entering HOVs) is less than $(L-1)Q$. These constraints are satisfied if the HOV-demand for the H-lane at a location ($x = 3$, 6 or 9) is large enough to relieve pressure from the G-lanes; i.e., if:

$$q_H(x) \geq q(x)L - Q(L-1) , \quad \text{for } x = 3, 6 \text{ or } 9. \tag{2b}$$

This is the same as (1a).[3]

---

[3] Note that $q_H(6) \leq q_H(9)$ and $q(6) = q(9)$; therefore the condition at $x = 9$ is redundant. This means that the overflow condition only needs to be verified downstream of on-ramps and upstream of off-ramps. As a further simplification,

We still have to check the input and output constraints (vi) and (viii). The latter is automatically satisfied since q_d and Q are the same in both scenarios, but the former needs to be checked since the shoulder lane flow on link *ef* changes from q in the G-scenario to $q_{ef}/(L\text{-}1)$ in the H-scenario; i.e., the upper bound to the on-ramp flow changes from αq to $\alpha q_{ef}/(L\text{-}1)$. We now show that the upper bound is higher in the H-scenario, i.e., $q_{ef}/(L\text{-}1) \geq q$, and thus that the H-lane cannot cause the on-ramp flow to decrease; it can only increase or stay the same.

The reason for the inequality is that the total assigned flow across screen line 6 is the same in both scenarios, and since the median lane carries less (or equal) flow in the H-scenario,[4] the shoulder lane (and the remaining G-lanes) must compensate for the deficit. If $q_{ef}/(L\text{-}1) = q$ the upper bounds are the same in both scenarios; since the assigned ramp flows are also the same, (vi) is satisfied. If $q_{ef}/(L\text{-}1) < q$ the upper bound increases. The priority rules imply in this case that on-ramp flow increases if and only if an on-ramp queue exists in the G-scenario. Absent this ramp queue, ramp inflow is invariant. Unfortunately, an increased inflow at *e* reduces the queued flow at *d*, and the off-ramp flow q_d: this flow would satisfy (viii) but the reduction is a bad thing. On the other hand, remedial action is always possible since on-ramp inflow can be reduced by metering. We summarize the above as follows:

> INSIGHT: *For both vehicle types, on-ramp and off-ramp flows, and freeway flows across any screen line, are the same in both scenarios if*: (*i*) *conditions* (2) *are satisfied; and* (*ii*) *either on-ramp queues do not exist in the G-scenario, or these queues exist but are metered in the H-scenario in such way that on-ramp inflows do not change.*

When used in the H-scenario, for example, the insight determines whether eliminating the HOV restriction would increase total system outflow. Since all the flows in (2) are observed in the H-scenario, one only has to guess whether on-ramp queues would arise in the G-scenario. To do this, check if the observed inflow q_e exceeds the reduced on-ramp capacity of the G-scenario, αq, where q is observed and α easily estimated. When applying the insight in the G-scenario, only

---

note that if L is sufficiently large (it doesn't have to be very large in typical cases) the RHS of (2b) becomes negative; then (2b) is automatically met.

[4] This is true because $q_{bg} \leq q_{gh}$ by virtue of conservation at node *g*, and $q_{gh} \leq q$ by virtue of (2a). Consequently, q_{bg} (the median lane's flow in the H-scenario) cannot exceed q (that lane's flow in the G-scenario), as claimed.

$q_H(x)$ needs to be guessed. Field experiments are planned to test if, absent on-ramp queues, total outflows from an interchange are indeed unaffected by an H-lane's activation or deactivation.

2.2 Queue density in each scenario

We now examine an open-ended queue spanning many interchanges. We consider the part of the freeway that would be queued under both scenarios and compare the two freeway densities. Density affects spatial growth, and this is examined in Sec 2.3.

The H-lane is still the median lane, but the restriction is interrupted for short sections between every pair of successive interchanges. In these sections the freeway operates in G-mode, even while the interchanges are in H-mode. In this way we can apply the insight of Sec 2.1 to each interchange separately, such that if (2) holds everywhere and the on-ramps are not queued (or metering is applied in the H-scenario), then flows of both vehicle types would remain invariant at any location that is queued in both scenarios. This is true no matter how short the G-sections are; and thus remains true even if they are eliminated.

We now prove that, at any location, the average density across all lanes in the G- and H-scenarios (k_G and k_H, respectively) satisfy the following theorem, where Q' is the capacity of the H-lane in the H-scenario and $u = 1 - q_H/Q'$ is its underutilization level.

PROPOSITION: *If the conditions of the insight in Sec.* 2.1 *apply, then at any location that is queued in both scenarios:*

$$(k_G - k_H)L = \kappa u \tag{4}$$

Proof: Displayed in Fig. 2 are the FD's for both an H- and a G-lane. Also shown are 4 points: G, H_h, H_g , and H, respectively denoting the states of: a G-lane in the G-scenario; and the H-lane, a G-lane and the average across all lanes in the H-scenario. Horizontal lines through these points identify corresponding flows on the ordinates axis: q_{GG}, q_H and q_{GH}. In light of the insight of Sec 2.1, total flow at the point of interest is the same in both scenarios and we can write: $q_H + (L - 1)q_{GH} = Lq_{GG}$. Therefore: $q_H - q_{GH} = L(q_{GG} - q_{GH})$. If we use vertical bars around the end-points of horizontal and vertical segments in Fig 2 to denote their length, this equality can be written as $|P_1P_3| = L|P_2P_3|$; thus, $1 = L|P_2P_3|/|P_1P_3|$. The geometry of Fig. 2 reveals that: $1 = L|P_2P_3|/|P_1P_3| = L|HG|/|H_hP_5| = L|HG|/(\kappa|P_0P_1|/|P_0O|) = L|HG|/[\kappa(1 - q_H/Q')]$. The second equality holds because triangles $H_hP_5H_g$ and HGH_g are

similar; the denominators of the third equality are the same because triangles OP_4P_6 and $H_hP_4P_5$ are similar; and the fourth inequality holds by substitution. Since u is defined as $(1-q_H/Q')$, the above equalities reduce to $1 = L|HG|/[\kappa u]$, and since $|HG| = (k_G - k_H)$, this concludes the proof.

We see from (4) that if the H-lane is perfectly utilized, the traffic density is the same in both scenarios. Thus, both queues would be equally long. Otherwise, the freeway is less dense in the H-scenario; it holds κu fewer vehicles per unit distance. Curiously, this storage deficit is independent of flow conditions on the freeway. The deficit will in most cases expand the H-queue and accelerate its spatial growth. This phenomenon is now examined.

2.3 Spatial growth of the H- and G-queues

Here we derive formulae for the trajectories of the backs of the H- and G-queues with due consideration of the entering and exiting ramp flows along the queues. We shall assume here and in the next sub-section that the freeway is homogeneous and long enough (length D) to include
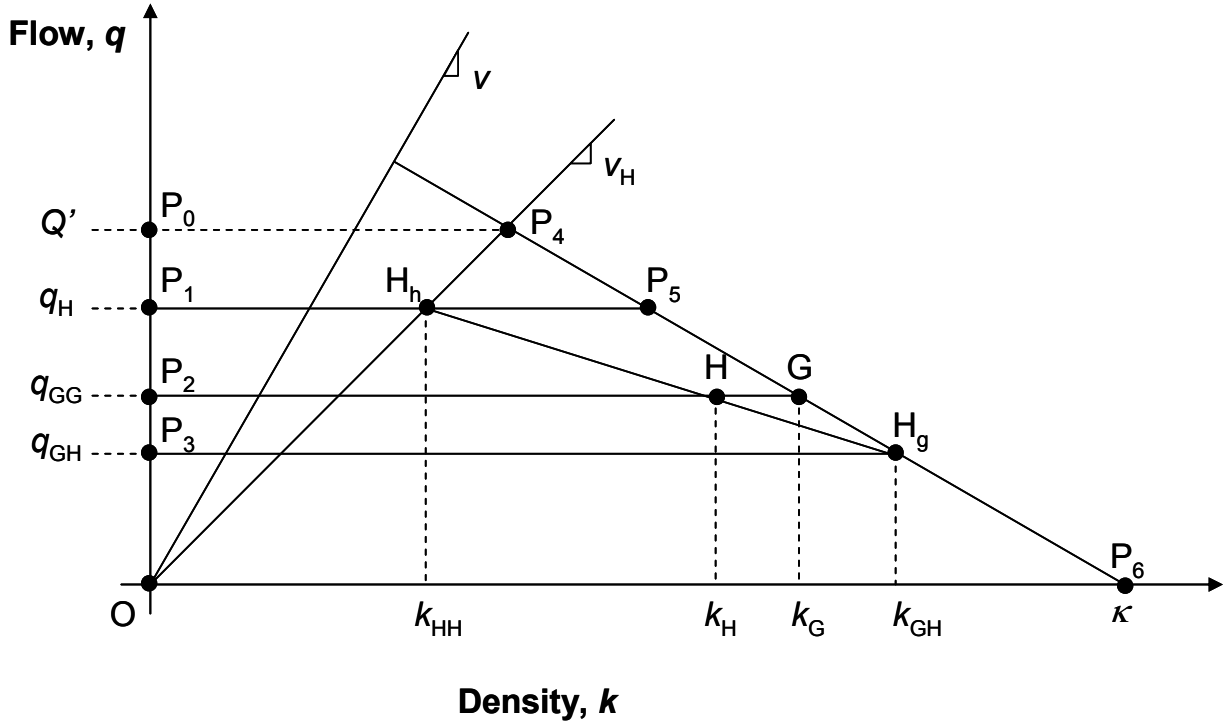

Figure 2. Fundamental diagrams for a single G-lane (triangle OP_0P_6), and a single H-lane (triangle OP_7P_6).

10

all the queues that develop; that its off-ramp/on-ramp pairs (interchanges) are spaced d distance units apart; that the demand is translationally symmetric with an inflow λ at each on-ramp; and that the distribution of destinations is Markovian in the sense that a fixed proportion of flow β exits at each off ramp. Note, that the equilibrium flow per lane Λ must satisfy $\Lambda L \beta = \lambda$. At the freeway's downstream end, a bottleneck with discharge rate per lane $q \leq \Lambda$ causes a queue to grow. We assume that on-ramp inflows along the queues are the same in both scenarios so that the insight of Sec 2.1 holds.

We assume that time starts ($t = 0$) when queues first form at the downstream bottleneck, and look for the times $T_G(y)$ and $T_H(y)$ when the G- and H-queues, respectively, grow to be y distance units long. We will develop a simple formula for the retardation factor of the G-queue, defined to be: $e(y) \equiv [T_G(y) - T_H(y)] / T_H(y)$.

Let $q(z)$ be the flow per lane in either of the queues z distance units upstream of the bottleneck, where $z \leq y \leq D$. Note that $q(z)$ is time-independent and the same for both scenarios as per the insight of Sec 2.1. To find $T_G(y)$ and $T_H(y)$, consider the vehicular accumulation growth rate, V, over the freeway's entire length D. Since vehicles are conserved, V is just the total net freeway inflow (the difference between inflows and outflows). To express this net inflow, let $\mu(y) \equiv \Sigma_{i=0,y/d} [\beta q(id)L] / (y/d)$ be the average outflow across the off-ramps that reside along the queue. Note that $\mu(y)$ is the same for both scenarios. Then, we have:

$$V(y) = (\Lambda - q)L + [\lambda - \mu(y)](y/d) \tag{5}$$

The first term of (5) is the difference between upstream and downstream mainline flows; and the second term the difference between ramp inflows and outflows along the queue. Note that there is no need to consider ramp flows upstream of the queue because entering and exiting flows are balanced there in both scenarios. Note too that $V(y)$ is the same for both scenarios.

We will show now that in either scenario, the accumulation growth rate $V(y)$ can also be expressed by considering changes in density, the motion of the back of the queue, and the fact that in any short time interval local density can only change on those points traversed by the back of the queue. Consider thus the time it takes for a queue to grow by dy length units, $dT.(y)$ -- the dot subscript shall henceforth be a place-holder for the scenario. When the queue reaches length y, it increases the density at that location from the free-flow density Λ/v_f to the queued density associated with flow $q(y)$. This density depends on the scenario and is denoted $k.(y)$. Thus, for

11

our small distance interval total accumulation increases by $[k(y) - \Lambda/v_f]\,dy$. Since this quantity must equal $V(y)dT.(y)$, the following ODE for $T.(y)$ results:

$$dT.(y)/dy \;=\; [k(y) - \Lambda/v_f]\,/V(y). \tag{6}$$

An expression for $T.(y)$, and for the difference across scenarios, $\Delta T(y) = T_G(y) - T_H(y)$, is found by integrating the right side of (6) from 0 to y for the two scenarios. The ratio of this integral and the constant $S(y) = \int_{x=0,y}[1/V(z)]dz$ is the weighted average of the changes in density $[k.(z) - \Lambda/v_f]$ for $x \in [0, y]$ with weights $1/[S(y)V(z)]$. We denote averages with these weights by the superscript "a". Thus, we can write:

$$T.(y) \;=\; [k^a.(y) - \Lambda/v_f]\,S(y),$$

$$\Delta T(y) = T_G(y) - T_H(y) = [k^a{}_G(y) - k^a{}_H(y)]\,S(y) = [\kappa u/L]\,S(y),\;\; \text{and}$$

$$e(y) \equiv \Delta T(y)/\,T_H(y) \;=\; [\kappa u/L]\,/\,[k^a{}_H(y) - \Lambda/v_f]. \tag{7}$$

Equation (7) is the sought-after retardation factor.

To simplify (7), the weighted average $k^a{}_H(y)$ appearing in the denominator can be replaced by the un-weighted average density; i.e., by the ratio of the number of vehicles in the queued freeway from 0 to y, and Ly. This simplification is (nearly) exact if the weights are (nearly) constant; i.e., if $V(y)$ is (nearly) independent of y. This will happen if the net vehicular inflow from the ramps along the queue is zero (or is small compared with the mainline net inflow). The simplification is also a good approximation if the density within the queue does not change much—since in that case weighted and un-weighted averages are similar. In most practical applications, the density within a queue does not change drastically and therefore we expect the un-weighted method to estimate e to well within a factor of 2. In our case, this is all we seek, since in typical applications e is just a few percentage points and, as we shall see, H-lanes can still reduce PHT with considerably larger e. We now examine the effect that the retardation factor has on VHT and PHT.

2.4. Effect of the H-lane on VHT and PHT

2.4.1 VHT: We show here that eliminating an existing H-lane along an open queue fractionally reduces VHT by at most: $[2u/L][\beta N]$, where N is the number of off-ramps affected by the H-queue. We also argue that for typical installations, $[2u/L][\beta N] << 1$.

Since vehicle flows entering our freeway are the same in both scenarios, to estimate changes in VHT it suffices to compare the cumulative changes in vehicular outputs. And since

12

flow at the freeway's downstream end (e.g., the discharge flow at *h* in Fig. 1b) is the same in both scenarios, we only need to consider the off-ramp flows. This estimation approach is used below.

The exit rate at an off-ramp that is *y* distance units upstream of the bottleneck is $\beta \Lambda L$ before the queue arrives and drops to $\beta q(y)L$ thereafter. These quantities are invariant across scenarios. The only difference is that in the H-scenario the off-ramp's exit rate declines $\Delta T(y)$ time units earlier, as shown by the cumulative curves of exit counts in Fig.3a. The shaded area is the total extra vehicular delay imparted by the H-lane. We can tightly bound this area from above with the product of the number of exiting vehicles delayed by the H-queue, $\beta q(y)L \tau_H(y)$, where $\tau_H(y)$ is the time elapsed while the H-queue blocks the *y*-ramp, and the maximum extra delay incurred by each vehicle, $\Delta w_{max}(y)$. Consideration of the figure shows that $\Delta w_{max}(y)$ and $\Delta T(y)$ are related by $\Delta w_{max}(y) = \Delta T(y)[\Lambda/q(y)-1]$, and therefore: $\Delta w_{max}(y) = T_H(y)e(y)[\Lambda/q(y)-1]$. Thus, the bound to the shaded area is: $\beta L \tau_H(y)T_H(y)e(y)[\Lambda-(y)]$.

We now sum these bounds across all off-ramps to obtain a tight upper bound to the total extra delay, ΔW, that occurs in the H-scenario. To this end, let *D* be the maximum length of the H-queue and $\tau_o > \tau_H(y)$ the duration of the queuing episode at the bottleneck; see Fig. 3b. Assume (neutrally) that the queue recedes steadily at the same rate it grows; thus, $\tau_H(y) = \tau_o-2T_H(y)$. In this case, $\Delta W \leq \int_{y=0,D} \beta L T_H(y)e(y)[\Lambda-q(y)]\tau_H(y)(1/d)dy = \beta L(1/d)\int_{y=0,D} T_H(y)e(y)[\Lambda-q(y)][\tau_o-2T_H(y)]dy$. If *e(y)* and *q(y)* do not vary much (as we expect in many practical applications), they can be replaced by their averages, *e* and $[\Lambda-q^a]$, in the above integrand to yield a good approximation.[5] The simplified approximate inequality is thus: $\Delta W \leq \Delta U \cong \beta Le(1/d) [\Lambda-q^a]\int_{y=0,D} T_H(y) [\tau_o-2T_H(y)]dy$. Since the queue grows (and recedes) at a steady rate such that $T_H(y) \cong \frac{1}{2}y\tau_o/D$ the integral is: $(1/12)D\tau_o^2$. Thus, the bound to the total extra delay is: $\Delta U \cong (1/12)\beta Le(1/d)[\Lambda-q^a]D\tau_o^2$.

We now develop an expression for the total vehicle delay of the H-scenario, V_H, so that we can determine the fractional contribution of the H-lane toward total delay, $\Delta U/V_H$. In this scenario, a queued freeway segment of unit length adds total delay at a rate $[k_H(y) - q(y)/v_f]L$ per

---

[5] If *e(y)* and $[\Lambda-q(y)]$ were to vary substantially, one could conservatively use high percentiles instead of averages.
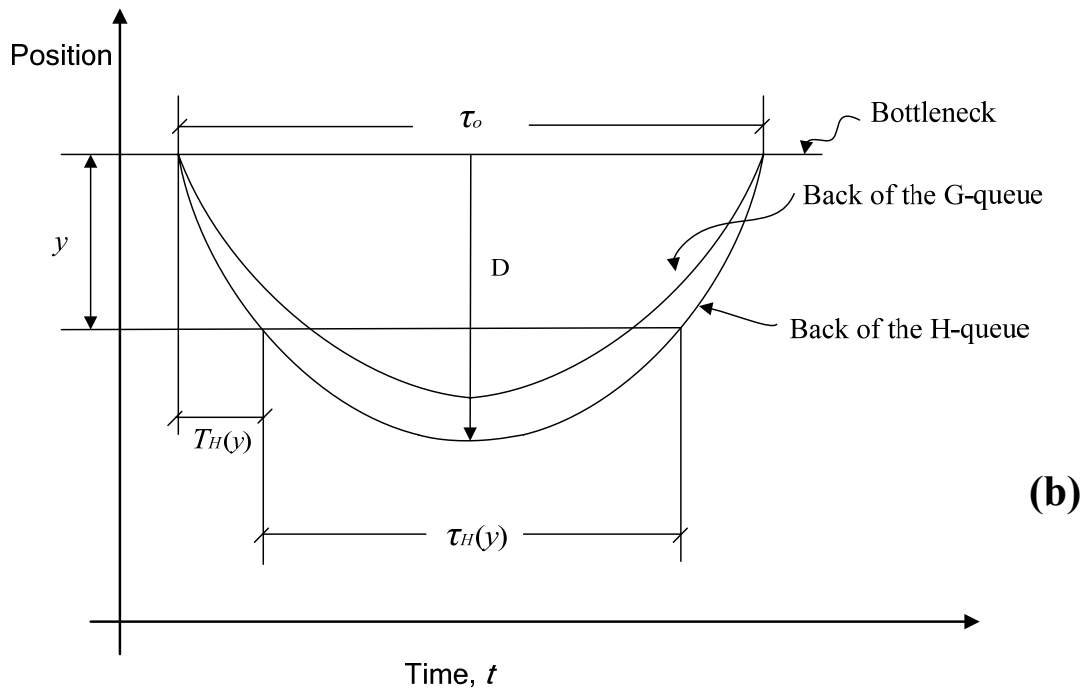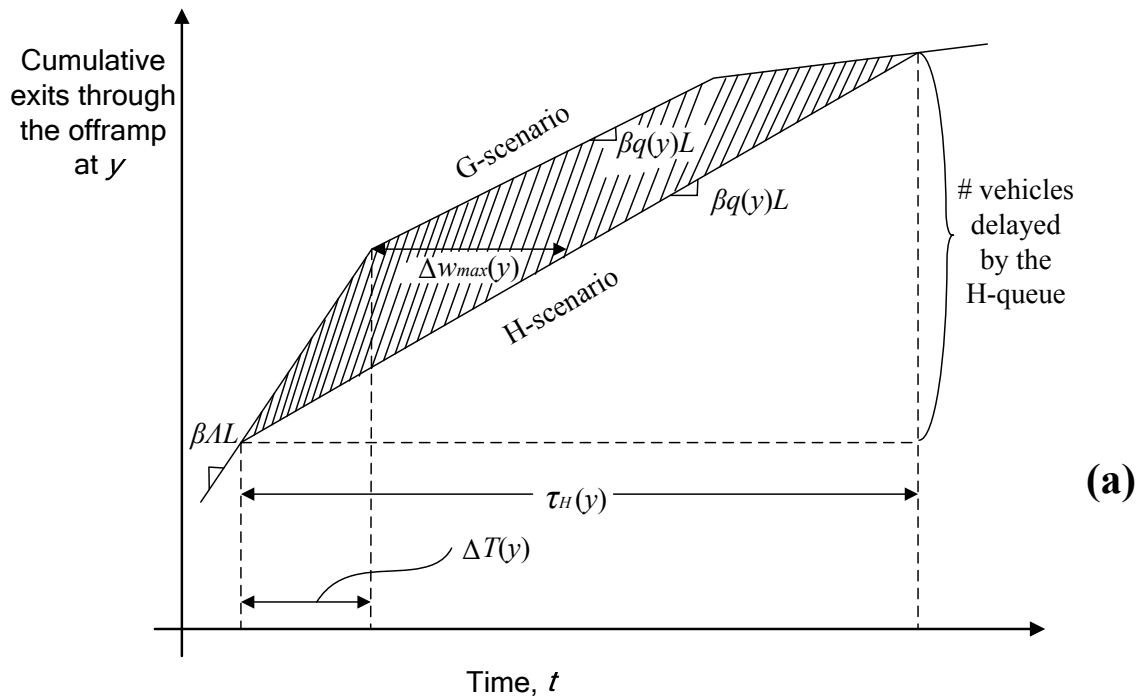
Figure 3. Off-ramp activity in the two scenarios: (a) cumulative exits at a single off-ramp; (b) spatio-temporal effects of queue.

unit time. Therefore, we have: $V_H = \int_{y=0,D} [k_H(y) - q(y)/v_f]L\tau_H(y)dy = \int_{y=0,D} [k_H(y) - q(y)/v_f]L[\tau_o - 2T_H(y)]dy \cong [k_H{}^a - q(y)/v_f]L[\frac{1}{2}D\tau_o]$. Thus, the fractional reduction in vehicle hours of delay achieved by eliminating the H-lane, $\omega = \Delta W/V_H$, satisfies: $\omega \leq \Delta U/V_H \cong (1/6)\beta e(\tau_o/d)[\Lambda - q^a]/[k_H{}^a - q(y)/v_f] \leq (1/6)\beta e(\tau_o/d)[\Lambda - q^a]/[k_H{}^a - \Lambda/v_f] \cong (1/6)\beta e(\tau_o/d)[2D/\tau_o]$. The last inequality is true because $\Lambda \geq q(y)$; and the last approximate equality because (as per the "shock equation") $[\Lambda - q^a]/[k_H{}^a - \Lambda/v_f]$ is the average speed of the back of the H-queue, which is $2D/\tau_o$. Since D/d is the number of off-ramps affected by the H-queue, N, the last of these equalities reduces to $(1/3)e\beta N$. Thus, we propose:

$$\omega \leq (1/3)e\beta N \qquad (8)$$

as a simple test to bound approximately the fractional increase in VHT imposed by an H-lane.

Since e is not readily observed (7) can be used as an estimate. And, since the ratio $[\kappa/(k^a{}_H(y) - \Lambda/v_f)]$ in (7) is typically close to 2 and rarely more than 6, we can use $6[u/L]$ as a conservative upper bound for e in (8). This yields the even simpler rule of thumb:

$$\omega \leq [2u/L][\beta N] \qquad \text{(rule-of-thumb.)} \qquad (9)$$

Reassuringly, (9) indicates that the savings in VHT obtained by switching from the H- to G-scenario are null if either the H-lane is perfectly utilized or there is no off-ramp flow, as one would expect.

Underutilizations comparable with 0.4 are typical in the San Francisco Bay Area (Cassidy et al, 2006). Thus, for freeways with 4 lanes or more − where H-lanes are normally found − the first factor in the right side of (9), $2u/L$, will rarely exceed 0.1. The second factor, βN, approximately equals the ratio of the flows leaving through the off-ramps and through the bottleneck when the queue is longest.[6] If this "off-flow ratio" is less than 30% − which is also typical − equation (9) would guarantee that $\omega \leq 3\%$. Thus, H-lanes would appear not to increase VHT significantly in typical open-ended queues.[7] Let us now see how H-lanes affect PHT.

---

[6] The two are exactly equal when the flow in the freeway queue is the same at all locations.

[7] An H-lane can add significant delays to an open ended queue if: (1) the freeway has few lanes; (2) the H-lane is severely underutilized; and (3) off-ramp flows are significant. On a 3-lane freeway with $u = 0.75$ and off-flow ratio 1 the penalty would be 25%.

2.4.2 PHT: A simple formula for the ratio of PHTs in the two scenarios is now derived. When exceptions for access to the H-lane are not made for LOVs (e.g., for low emission vehicles, on the basis of tolls, etc), the result simplifies. We show that an H-lane of this type reduces PHT if the fraction of demand that are HOVs, f, is greater than ω. This comparison only involves quantities that are readily obtained from detector data, and is therefore easy to verify in the field.

We first consider the general case, where exceptions are allowed. Let o_H be the average number of people in those vehicles allowed to use the H-lane, o_L the average in vehicles that are not, and $o = fo_H + (1 - f)o_L$ the overall average. If an H-lane is successful in allocating most of the VHT to LOVs, it will generate $P_H = o_L V_H$ person-hours of delay in the H-scenario. On the other hand, the person-hours of delay in the G-scenario would be: $P_G = oV_G = o(1-\omega)V_H$. Thus, the ratio of the PHTs in the two scenarios is:

$$P_G / P_H = (1-\omega)(o/o_L) \qquad \text{(general case).} \qquad (10)$$

When no exceptions are made, so that every vehicle on the H-lane is an HOV, all H-lane vehicles carry at least one more passenger than the rest. Hence, o_H is bounded from below as: $o_H \geq o_L + 1$. In this case, after replacing o_H by its lower bound in the equality $o = fo_H + (1 - f)o_L$, and dividing both sides by o_L, we find that $o / o_L \geq 1 + f / o_L \cong 1 + f$. The last approximation reflects current conditions in the US. Inserting it in (10) we find:

$$P_G / P_H = (1-\omega)(o/o_L) \geq (1-\omega)(1+f) \cong 1 - \omega + f \qquad \text{(if no exceptions allowed).} \quad (11)$$

The last approximation in (11) applies when both ω and f are small compared with 1, as is usually the case in the US, and implies that an H-lane improves mobility if $f > \omega$. Whereas (10) requires estimates of vehicle occupancies, (11) only involves quantities measured by detectors.

Section 2 has examined how an H-lane bypassing an open-ended queue affects freeway mobility. Metering was used, as explained in Sec. 2.1, to ensure that all vehicles entered the freeway at the same rates in both scenarios. Thus, the H-lane does not affect the surface streets: PHT and VHT outside the freeway are invariant to the scenario. We used worst-case assumptions with no smoothing effect, no modal shift, and no preferential treatment of HOVs at the metered ramps. Preferential metering should reduce on-ramp PHT, increasing overall benefits, as would the smoothing effect and modal shift (if they were to happen). This could be decisive in cases where the freeway VHT penalty is close to the freeway PHT benefit. We now examine the effect of H-lanes on closed-loop queues that form on beltways.

3. CLOSED-LOOP QUEUES

We consider here an ideal *L*-lane beltway with rotational symmetry, and use the notation of Sec. 2 except where otherwise indicated. Symmetry is useful because it can be exploited to derive general insights. Sec. 3.1 will show that an H-lane reduces the maximum outflow possible from any rotationally symmetric beltway that is susceptible to gridlock. (This maximum outflow is achieved when on-ramp inflows are as high as possible without creating beltway queues.) In this idealized case, the H-lane would increase freeway VHT and likely do more harm than good.

This result is somewhat academic, however, because, without rotational symmetry or time-invariance, maximum flow is not necessarily achieved by eliminating all queues. On a real-world busy beltway, one would expect queues to remain at isolated locations, or to persist all around the beltway, even when metering the ramps. Isolated queues can be analyzed as in Sec. 2. But we have yet to examine the case where queues persist around a (metered) beltway. This is done in Sec. 3.2; it shows that in this case an H-lane *combined with preferential metering for HOVs* can reduce PHT without increasing VHT anywhere.

3.1 Maximum outflows of an idealized beltway with and without an H-lane: time-independent case

We assume that queues exist on all the on-ramps of our congested beltway and analyze the inflows and exit flows per ramp. Since the system is closed, the exit flow per ramp is a measure of system outflow. Consider first the G-scenario.

When the (queued) flow per lane upstream of an off ramp is q the exiting flow is βqL. To avoid gridlock and maintain a steady exit flow per off-ramp μ (and circulating flow $\mu/\beta L$), on-ramp meters should allow more (μ^+) vehicles into the beltway when the circulating flow rises (i.e., queued beltway density decreases) above the target $\mu/\beta L$, and fewer (μ^-) when the circulating flow drops; see Daganzo (1996, 2007) for more details. The on-ramp flow is bounded from above by $\mu \leq \alpha Q$ and the circulating flow by $\mu/\beta L \leq Q$. Therefore, the maximum exit flow that can be achieved in the G-scenario is $\mu^G = \min\{\alpha; \beta L\}Q$. If $\alpha > \beta L$ the system is susceptible to gridlock (more vehicles can enter than leave), and the maximum outflow ($\mu^G = \beta QL$) can only be achieved by metering; if $\alpha \leq \beta L$ the system self-regulates and the maximum flow arises spontaneously as the on-ramp queues discharge. Consider now the H-scenario.

As before f is the fraction of HOVs in the on-ramp queues (the fractional demand) and q^H the average flow per lane circulating across screen lines between neighboring interchanges (e.g., each drawn downstream of an on-ramp and upstream of an off-ramp). Assume that HOVs and LOVs have the same β, so that the combined exit flow of both vehicle types is βq^H. If the on-ramps are metered so as to match and sustain this output at every off-ramp, and we do not give priority to HOVs at the on-ramps, each on-ramp's input flows of HOVs and LOVs will be: $\beta q^H f$ and $\beta q^H(1-f)$, respectively. These will create HOV and LOV flows $q^H f$ and $q^H(1-f)$ at all of our inter-ramp screen lines. For these flows to be feasible they must satisfy capacity constraints for the H-lane and the G-lanes. Thus, the maximum possible circulating flow q^{H*} is the solution of:

$$\max \{q^H\} \quad \text{s.t.:} \tag{12a}$$

$$q^H f \leq Q \qquad \text{(H-lane constraint) and} \tag{12b}$$

$$\beta q^H f + q^H(1-f) \leq Q(L-1) \qquad \text{(G-lane constraint).} \tag{12c}$$

We assume that $f < 1/L$, so that the first capacity constraint is redundant. Then, $q^{H*} = Q(L-1)/[1 + f(1 - \beta)]$. Since the maximum exit flow from an off-ramp in the H-scenario is $\mu^H = \beta q^{H*}$, we see that $\mu^H = \beta Q(L-1)/[1 + f(1 - \beta)]$. Thus, the ratio μ^H/μ^G is: $[1-1/L] / [1-f(1-\beta)]$. Since $0 < f < 1/L$ and $0 < \beta < 1$, the denominator of this ratio is always larger than the numerator. Thus, installing an H-lane always reduces an idealized beltway's maximum outflow.

3.2 A congested beltway with and without an H-lane: time-dependent case

Start again with the G-scenario. To prevent on-ramp queues from spilling back to city streets the beltway is operated during the rush in the congested regime (where it can store many vehicles) with a circulating flow per lane q inferior to the maximum possible in the H-scenario, as determined from (12). The exit flow at each off-ramp would be βqL and this would be achieved with queues on all the on-ramps.

The same outflows can be achieved in the H-scenario with κu fewer vehicles per mile of beltway, as per the proposition of Sec 2.2, but to achieve this smaller density one would have to meter the on-ramps more aggressively at the beginning of the rush, since demand is invariant to the scenario. Thus, the $\kappa u d$ vehicles removed from each inter-ramp beltway segment would have to be transferred to the on-ramps. This is undesirable, since we are trying to control spillbacks. Therefore, we do not pursue this strategy and ask instead a more relevant question: how is

outflow impaired by an H-lane with a given underutilization level u if we maintain the same beltway accumulation in both scenarios?

The proposition of Sec. 2.2 implies that the congested branch of a beltway's FD in the H-scenario must be parallel to the congested branch in the G-scenario; see Fig. 4. One can see that when density is the same in both scenarios, the circulating flow in the G-scenario is greater by $\kappa uw \cong uQ$ units, where w is the backward wave speed. Thus, an off-ramp's exit flow would increase in the G-scenario by βuQ. Since it was originally βqL, it rises by a factor $[uQ/qL]$. This is the real penalty imposed by the H-lane. It reduces the outflows that can be sustained with the same beltway accumulation. The effect is undesirable because it extends the length of the rush, negatively affecting all vehicles and lengthening on-ramp queues. Thus, H-lanes should not be active when a beltway is congested everywhere.

The only exception to this rule arises if $u \cong 0$. This, of course, would require a fortuitous value of f that cannot be expected to happen frequently. But it can artificially. Simply, modulate f by metering HOVs and LOVs differently (enriching the entering stream with HOVs) but keeping
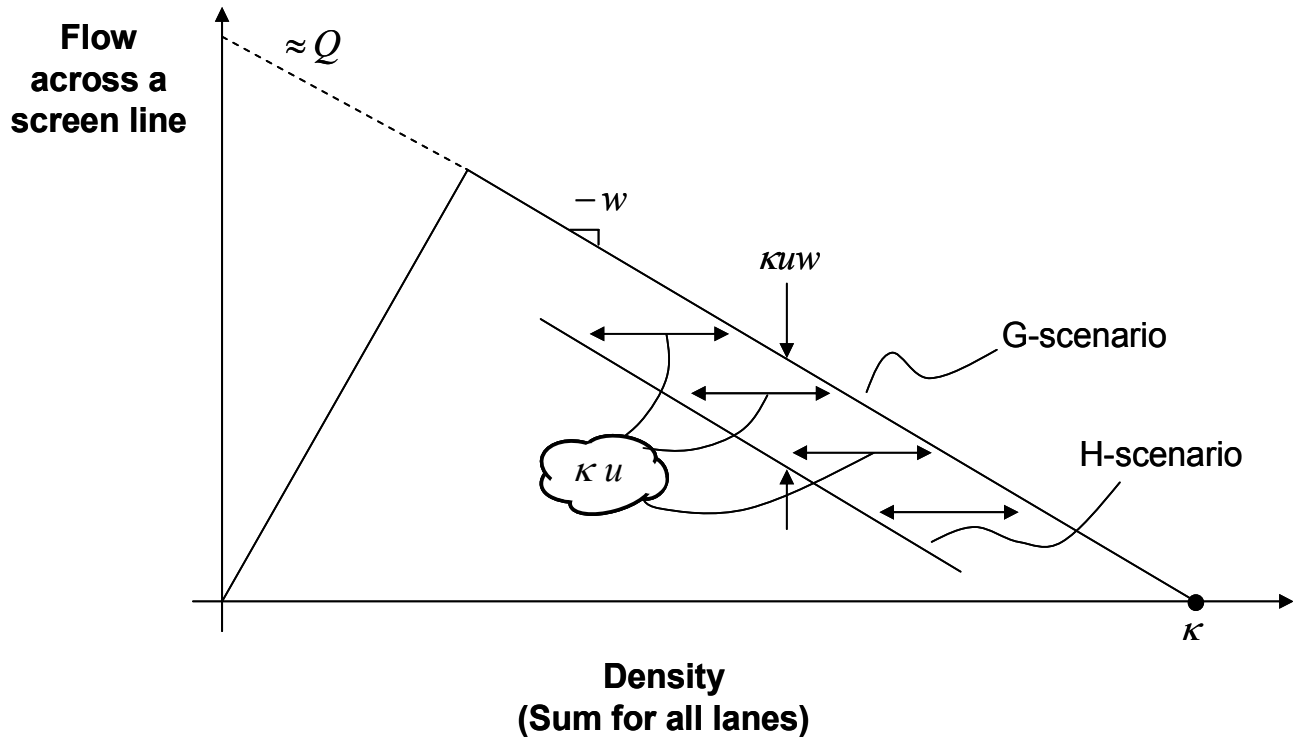

Figure 4. Fundamental diagram for a whole freeway in two scenarios (underutilization = u).

the total metering rate for the two flows combined the same as in the G-scenario. If queues of both vehicle types exist on the on-ramps one can increase f as much as necessary to saturate the H-lane, and achieve $u \cong 0$. Call this approach differential-metering; "D-metering" for short.

With D-metering in the H-scenario the beltway suffers no loss in storage compared with the G-scenario, and the same outflows are achieved. In this way we avoid transferring VHTs from the beltway to its on-ramps. Since total vehicular inputs into the beltway are the same, the beltway VHT also remains invariant. D-metering in the H-scenario, however, allocates all the beltway delay to LOVs, unlike the G-scenario. Thus, (10) and (11) apply with $\omega = 0$ and we see that D-metering reduces PHT relative to the G-scenario.

One can also D-meter in the G-scenario without changing total on-ramp flows. This reduces on-ramp PHT, same as in the H-scenario, but without the positive effects on the beltway. In either scenario, D-metering cannot be sustained for the duration of the rush, but only while HOV queues exist on the on-ramps. At other times, and if the beltway is congested everywhere, the H-lane should be deactivated.

The above illustrates that dynamic strategies can be of value even in seemingly hopeless situations, like a fully congested, symmetric beltway. The strategies should be even more beneficial in realistic cases with asymmetric demand where they can be adjusted to deal with local spatiotemporal phenomena such as fragmented queues. But an important point of this subsection is that closed-loop queues respond to H-lanes in a very different way from open ended queues and that management strategies should recognize this dichotomy.

4. DEPLOYMENT

Here we summarize those of the above findings that can be applied macroscopically on an urban scale, and then discuss what they say about H-lane deployment.

We found that underutilized H-lanes reduce traffic density for the same flow, relative to the G-scenario, albeit only slightly if the underutilization is slight. The density reduction forces queues to grow longer and faster, i.e., to expand more, in the H-scenario. This phenomenon is critical on congested beltways with closed-ended queues because such queues can only expand on the on-ramps; and on-ramp queues are problematic if they spill back onto the surface streets and interfere with local traffic. The density reduction effect, on the other hand, is not very

20

significant for open-ended freeway queues, because they can usually expand on the freeway with little harm, even if the H-lane is underutilized.

The above suggests a general deployment and design principle: since traffic is stored less densely [κu fewer vehicles per mile] on an H-system than a G-system, the former requires more storage buffers, e.g., city streets, and these should be located where the stored vehicles would not interfere significantly with traffic that does not use the freeway.

As an example, consider a city with a ring and radial system of thoroughfares where most trips during the morning (evening) rush are to (from) the center. For the morning commute, our findings suggest that H-lanes can be used on the inbound radial freeways, since queues on these facilities would typically be open-ended. Our findings also mean that H-lanes can be used on the outer rings where queues are likely to be fragmented (and therefore open-ended). But H-lanes should not be used on uniformly congested inner rings.

The evening commute is slightly different. Although rings should be treated as described above, radial links should not. Outbound queues close to the center would spillback onto the congested city streets and interfere with local traffic. Thus H-lanes should not be installed on outbound freeway links that are close to the center. If at locations further from the center outbound queues fragment, H-lanes can be used to bypass these queues.

H-lanes may be deployed even more generally than described above if they induce higher bottleneck discharge flows by means of the smoothing effect.

5. GENERALIZATION FOR OTHER MODES IN THE URBAN LANDSCAPE

Our findings about H-lanes suggest ways to allocate crowded street space to transportation modes with different passenger occupancies and performance characteristics. We found that reserving freeway lanes for HOVs reduces PHT without significantly increasing vehicle-kilometers of travel if: (i) one does not create new bottlenecks or reduce the capacity of existing ones; and (ii) LOVs removed from the reserved areas can be stored harmlessly in buffers. Thus, proper deployment increases the distance that people can travel in a given time and the number of opportunities available to them; i.e., it increases accessibility. Equation (10) reveals that accessibility increases with o/o_L. Thus, in agreement with intuition, lane reservation strategies should yield even greater benefits in urban street contexts, when applied to vehicle classes in which passenger occupancies are more markedly different, such as buses (or trams) vs. cars.

When vehicles of different classes have different performance characteristics, as occurs with buses (trams) and cars, lane reservation schemes should smooth flow more markedly than in the case of H-lanes for cars, increasing outflows and reducing PHT. With segregated lanes, buses are prevented from delaying cars (when making stops, changing lanes or accelerating slowly) and cars from delaying buses by forming queues next to bus stops. Although we do not know yet how to predict the magnitude of this smoothing effect for surface streets with cars and buses (this is an important research topic) we can predict its impact on system PHT parametrically. Of particular interest is its effect on the worst-case scenario of Sec 3.1, because a beltway is a metaphor for severely congested city centers. Therefore, the analysis of Sec. 3.1 generalized below.

5.1. Performance of bus-only lanes in uncongested beltways

We consider a beltway on which there is one (or more) H-lane(s) reserved for a bus (or tram) route. We assume that the transit agency supplies enough vehicles and drivers to sustain the same fixed frequency, q_H, in both scenarios. In this way, bus passengers experience the same out-of vehicle delay in both scenarios; and we can focus on their in-vehicle travel time: the PHT. We also assume that our HOVs (buses from now on) do not enter or leave the loop in significant numbers during the study period. We now compare the two scenarios.

As before, Q is the capacity of a lane carrying only cars in the H-scenario (cars/hr). The capacity of a lane carrying only buses will in general be a smaller number (buses/hr), since buses make stops and are less maneuverable than cars. Hence, we introduce a passenger-car-equivalent (pce) constant, p that converts buses into "car-equivalents" for the purposes of determining maximum flows. We estimate $p \cong 2$. This means that the number of lanes allocated to HOVs in the H-scenario, $l < L$, must satisfy: $pq_H < Ql$. In the G-scenario flow should be less efficient, and the capacity of a lane (in cars/hr, counting each bus as p cars) significantly smaller. It shall be denoted rQ, where $r < 1$ is a positive parameter.[8]

Consider now the H-scenario, and look for the best l. Since buses generally stay in the beltway without entering or exiting, they do not affect the LOV lanes if $lQ \geq pq_H$. Under these conditions the maximum LOV flow (arising downstream of our beltway's merges) is $Q(L-l)$. To

---

[8] This parameter should depend on the mix of buses vs. cars; and be smallest when the stream includes significant numbers of both.

maximize it, choose the smallest integer l that satisfies: $lQ \geq pq_H$. The result, l^*, should leave a gap in the inequality smaller than the capacity of one lane; i.e., such that $l^*Q - pq_H = uQ$, where $u \in (0, 1)$ is the underutilization level of the bus-lanes. Thus, in the H-scenario, the maximum combined flow of both vehicle types on the two sets of lanes, $pq_H + Q(L-l^*)$, becomes: $Q(L-u)$. In the G-scenario, the maximum combined flow is $(1-r)QL$. Thus, the extra flow circulating on the beltway in the H-scenario is:

$$\text{Extra flow on beltway in H-scenario} = Q(Lr - u). \tag{13}$$

This extra flow is composed of LOVs only, since bus flow is fixed. In the H-scenario, this extra flow produces $\beta Q(Lr - u)$ extra units of LOV exit flow per off ramp. Thus, the H-scenario is better for LOVs if $Lr > u$. Models to predict r are not available but we expect values comparable with 0.2 to arise when the stream contains a significant fraction of buses that make many stops. (This is common in cities that rely heavily on buses and para-transit to meet their transportation needs.) With r this large, segregation should improve LOV outflows on beltways with four lanes or more, even if the bus-only lanes are significantly underutilized. Recall that our segregation strategy does not affect the PHT of bus users because it keeps invariant both, the bus service frequency and the bus speed on the beltway. Thus, the strategy is Pareto efficient if $Lr > u$. This shows that segregation of street space when modes are very different improves mobility, even in the worst-case situation of a symmetric beltway.

5.2 Bus-only lanes in congested beltways and general deployment issues

The above assumed that the beltway can be maintained congestion-free, and this may not always be possible. To relax this assumption, the logic of Sec. 3.2 is now qualitatively applied with two modifications: (i) no D-metering, since buses are permanently on the beltway; and (ii) not using the proposition of Sec. 2.2 to estimate density changes because it only applies to two identically performing vehicle classes. We conjecture that a mixed stream of buses and cars is considerably less dense than if the vehicle fleet was homogeneous;[9] and that if the fraction of buses is large, the segregated stream could contain more LOVs than the mixed stream. In this case, the H-scenario would exhibit smaller VHT and PHT for LOVs.

---

[9] Research is planned to substantiate this conjecture. Our rationale is that lane-changing interactions in a G-scenario are both more numerous and onerous (they may involve mandatory stopping) in bus vs. car systems than in car vs. car systems. These interactions may introduce gaps in the traffic stream that reduce density.

In addition, because the beltway is congested, buses travel faster in the H-scenario than in the G-scenario, and this helps the bus side of the system in two ways: (i) by reducing passengers' in-vehicle PHT; and (ii) by allowing the bus agency to maintain the stipulated service frequency with fewer vehicles and drivers. This shows that bus-only lanes can offer copious benefits even in situations, such as our symmetric beltway, where conventional HOVs would not.

We conclude that the recipe for deployment of special bus-only lanes in cities (including bus rapid transit applications) should be more liberal than for conventional HOVs, but similar in character. The following are tentative rules, pending research on the behavior of mixed traffic streams: (i) devote special lanes to buses in congested central areas, but only where they can be nearly fully utilized; (ii) restrict LOV traffic in the remaining lanes (by metering access points, pricing or some other means) to ensure they flow as close to capacity as possible; (iii) use rules (i) and (ii) also for the outbound radii close to the center; and (iv) use special lanes liberally for inbound radii, and wherever streets are not uniformly congested.

The above assumes that transit routes and schedules are given, but more improvements are possible if one is allowed to change them. For example by focusing the layout of parallel transit lines on a few streets, instead of spreading them over many, we can ensure that bus frequencies are high enough for rule (i) to apply. This would increase bus speeds, but could also reduce connecting times and increase walking distances. Holistic study of these matters is necessary. To this end, a comprehensive theory of transit system design on an urban scale would be useful, particularly if it could be linked to the emerging theory of urban traffic dynamics. By combining the two theories, urban decision-makers could choose with scientific information how best to allocate the urban pavement to its various possible uses, and even to decide how much pavement should be provided per city dweller. Much research is still needed.

ACKNOWLEDGEMENT

REFERENCES

Cassidy, M.J., Daganzo, C.F., Jang, K. and Chung, K. (2006) "Empirical reassessment of traffic operations: freeway bottlenecks and the case for HOV lanes" Institute of Transportation Studies Report UCB-ITS-RR2006-6, University of California, Berkeley, CA.

Cassidy, M.J. and M. Mauch (2001) "An observed traffic pattern in long freeway queues" *Transportation Research Part A*, **35**: 143-156

Chen, C., P. Varaiya, and J. Kwon. (2005) "An empirical assessment of traffic operations" Proceedings of the 16th International Symposium on Transportation and Traffic Theory (H. Mahmassani, editor) pp. 105-123*, College Pak, Maryland, Elsevier, London, U.K.

Daganzo, C.F. (1995) "The cell transmission model. Part II: Network traffic" *Transportation Research PartB* **29**(2): 79-93.

Daganzo, C.F. (1996) "The nature of freeway gridlock and how to prevent it" in *Transportation and Traffic Theory*, Proc. 13th Int. Symp. Transportation and Traffic Theory (J.B. Lesort, ed) pp. 629-646, Pergamon-Elsevier, Tarrytown, N.Y. (1996).

Daganzo, C.F. (2007) "Urban Gridlock: macroscopic modeling and mitigation approaches" *Transportation Research Part B* **41**: 49-62.

Daganzo, C.F., J.A. Laval and J.C. Muñoz (2002) "Ten strategies for freeway congestion mitigation with advanced technologies" *Traffic Engineering and Control,* **43**: 397-403.

Laval, J.A., M.J. Cassidy and C.F. Daganzo (2005) "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 (proceedings, in press).

Laval, J.A. and C.F. Daganzo (2006). Lane-Changing in Traffic Streams. *Transportation Research Part B*, **40**(1):251-264.

Menendez, M. and Daganzo, C.F. (2006) "Effects of HOV lanes on freeway bottlenecks" U.C. Berkeley Center for Future Urban Transport Working Paper UCB-ITS-VWP-2006-3.

Menendez, M. (2006) "An analysis of HOV lanes: their impact on traffic" PhD dissertation, Department of Civil Engineering, University of California, Berkeley, CA.

Muñoz, J.C. and C.F. Daganzo (2002) "The bottleneck mechanism of a freeway diverge" *Transportation Research Part A*, **36**: 483-508.