Title

Effects of HOV Lanes on Freeway Bottlenecks

Authors

Menendez, Monica
Daganzo, Carlos F.

Publication Date

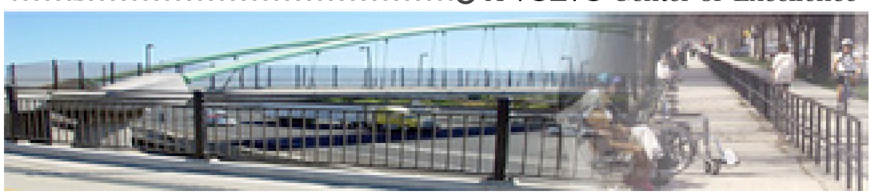
2006-09-01

Effects of HOV Lanes on Freeway Bottlenecks

Monica Menendez and Carlos F. Daganzo

WORKING PAPER
UCB-ITS-VWP-2006-3

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



September 2006

EFFECTS OF HOV LANES ON FREEWAY BOTTLENECKS

Monica Menendez and Carlos F. Daganzo
Institute of Transportation Studies
University of California, Berkeley, CA 94720

(September 5, 2006)

ABSTRACT

High occupancy vehicle (HOV) lanes can affect the capacity of freeway bottlenecks through both an under-utilization effect and a disruption effect. An under-utilized HOV lane obviously reduces its own discharge rate at the bottleneck. But lane changes in and out of the HOV lane can also disrupt the flow on the adjacent general purpose (GP) lanes, and reduce their capacity. Bottleneck capacity reductions, arising from the combination of both effects, are undesirable because they increase vehicle-hours of travel.

This paper shows with systematic simulations that the disruption effect is not significant at isolated bottlenecks. If anything, HOV lanes slightly improve GP flow at these locations. Reductions in GP-capacity were found only in highly idealized situations without bottlenecks. Thus, it appears that the total discharge rate of bottlenecks with and without HOV lanes can be conservatively analyzed assuming that the capacity of the GP lanes is constant.

Using this assumption, the paper then shows how to estimate total bottleneck capacity, and how to deploy HOV lanes without creating new bottlenecks or changing the total flow through existing ones. This milestone is sometimes sufficient to guarantee no change in the total vehicle-hours of travel and, hence, a reduction in people-hours. The paper also introduces a dynamic operating strategy for HOV lanes that significantly increases a bottleneck's total discharge rate. This suggests that dynamic strategies can be used to reduce not just people delay but also vehicle-hours and their externalities.

1. INTRODUCTION

HOV lanes are freeway lanes, usually in urban areas, reserved for vehicles carrying more than a predetermined number of occupants. As such, they allocate vehicular delay to low occupancy vehicles. Thus, if an HOV lane does not increase total vehicular delay it reduces total people delay. This is a quantifiable benefit because vehicle-hours and people-hours are easy to observe. In addition, if an HOV lane induces some drivers to share rides urban dwellers also derive environmental benefits. Because these are more difficult to quantify, we shall only study in this paper the relationship between HOV lanes and vehicular delay. There is merit in this. If we can estimate how an HOV facility changes vehicular delay we can also estimate how it affects people-delay; if the latter benefit significantly outweighs the costs, the facility should be kept.

It is well known that the total vehicular delay in any input-output system with given vehicular inputs is only determined by the system's vehicular outputs, which in turn are determined by the system capacity. Thus, a proper focus is the relation between HOV lanes and system capacity. Here, we narrow the focus to single-bottleneck systems. A follow-up study will examine the multi-bottleneck case.

Each year, significant amounts of time and money are devoted to the study, construction and improvement of HOV facilities. However, despite the efforts of both engineers and researchers, the impacts of HOV lanes are not fully understood. The empty lane syndrome is the best known problem. It obviously impacts system capacity. But other potential pitfalls have also been noted. Data from a recent study (Chen *et al.*, 2005) reveals that in many California freeways speed declines across all lanes when an HOV lane restriction starts. The lower speeds then persist while the restriction is in force. The data show that speed is reduced on both, the HOV lanes and the GP lanes, and that the drop on the latter is more severe and deleterious to travel time. But this study does not identify the locations at each site where the drops first occur, which could reveal the causal mechanism for the drops and the role that HOV lanes play in it.

One would expect drops in speed and flow to start at locations where the GP lanes are closest to saturation, since the migration of LOVs (low occupancy vehicles) toward these lanes shortly before the "restriction instant" (the time-of-day when the restriction starts) could increase the demand for GP lanes beyond their capacity and create bottlenecks. The demand/capacity imbalance could be exacerbated if the lane-changes during and after the migration were to reduce

GP lane capacity. This possibility must be considered since lane changes are known to reduce bottleneck capacity (Laval and Daganzo, 2006), and the speed drop on the HOV lanes suggests that there is considerable lane-changing interaction between the GP and HOV lanes.

This paper shows that this is not the case: HOV lanes do not reduce the capacity of GP lanes. Section 2 below presents some empirical evidence. Sections 3 and 4 complement it with simulations. Sections 5 and 6 then describe both, how to deploy HOV lanes without hindering vehicle flow and a dynamic strategy that increases it. Qualitative remarks conclude the article.

2. PRELIMINARY EMPIRICAL EVIDENCE

To gain some insight we examined an HOV site with detailed loop detector data (from the Berkeley Highway Lab) where speed drops were observed: Interstate 80 (eastbound) from Berkeley to Albany in California. This site is upstream of a major diverge; the I80/I580 split, where I80 drops from 5 to 4 lanes. Figure 1 shows the speed and flow profiles for all lanes observed at a detector station 2 miles upstream of the split from 2:00 to 8:00 pm. The data are typical. The darkest line depicts the HOV (median) lane. The HOV restriction is in force from 3:00 to 7:00 pm (shaded area). Note that after 3:00 pm the average speed on the HOV lane drops significantly, roughly from 70 to 45 mph, even though its density remains low, and the GP lanes suffer a larger drop, to 25 mph. These speed reductions on both the HOV and the GP lanes are the phenomena we wish to investigate.

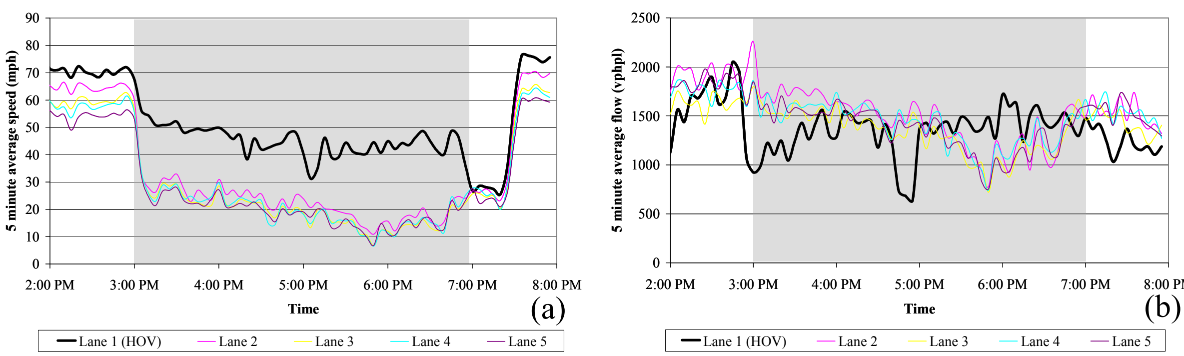


Figure 1. Loop detector data from I-80 (evening commute on August 19, 2002). Lanes are consecutively numbered from the median-HOV (1) to the shoulder (5): (a) speed profiles; (b) flow profiles.

Comparing parts (a) and (b) of the figure we see that there is a 15-min lag between the changes in flow associated with the HOV restriction and the drop in speed. This indicates that the initial 15-min of lane changing activity had little effect on speed and that the speed drop was either due to subsequent lane changes (which should be less numerous) or to something else. The most likely explanation is the arrival of the back of a queue from the I-80/I-580 split, which could easily become oversaturated when the HOV restriction starts. Unfortunately, data from the split were not available to ascertain how the bottleneck flows were affected by the HOV lane restriction, so our conclusion is not definitive.

But the lack of a speed collapse when lane changing should be at its maximum does suggest: (i) that the lane-changing effect of an HOV lane on the GP lanes should not be drastic; and (ii) that its effect on capacity should be small. Of course, this can only be firmly and generally concluded by observing active bottlenecks of many different types before and after their restriction instants. Because sites providing natural experiments of this type are scarce, we use simulations to quantify the effects. Field tests are still desirable.

3. THE SIMULATION MODEL

This section can be skipped without loss of continuity. It summarizes the simulation model and some tests; for more detail see Menendez (2006). We wished to avoid the calibration problem that plagues current simulation packages with a parsimonious model that would realistically capture the HOV/GP interactions. [Most simulation models have many parameters; some are unobservable, have unclear physical meaning and require re-calibration across sites. For an overview see Brackstone and McDonald (1999); and Hoogendoorn and Ossen, (2005).] An exception is a recent study (Laval and Daganzo, 2006), which proposed a model with only five parameters. One of these parameters is a bound on vehicular acceleration, which captures the effect of lane changes to faster lanes. Since these types of moves prevail at merge, moving and lane-drop bottlenecks the model was found to be accurate (without re-calibration) at these types of bottlenecks; see also Laval *et al.* (2005). This model, however, omitted a deceleration bound—on the belief that it would only provide marginal accuracy improvements for the abovementioned types of bottlenecks while adding another layer of complexity. However, in the case of highways with HOV facilities and diverges, vehicles often have to move to a slower lane and deceleration could be an important effect. Hence, the proposed model includes both acceleration and deceleration bounds. It also requires few parameters and is flexible enough to capture different HOV scenarios. It was shown to guarantee the same level of accuracy as existing models (Menendez, 2006). The model is driven by the absence of collisions. A brief description follows.

Although in reality drivers are idiosyncratic and strategic, the model assumes for simplicity that they are aggressive, rational, consistent, and myopic. They make decisions in discrete time, with an interval Δt. At every time *t*, each vehicle *n* knows its position, $x^n(t)$, and its average speed in the previous time interval, $v^n(t) = \left(x^n(t) - x^n(t - \Delta t)\right)/\Delta t$, as well as those of nearby vehicles. When not changing lanes a driver tries to take the most advanced position, x^n, from among those possible, p^n, constrained by the mechanical properties of his/her vehicle, safety, and comfort, i.e., the car-following model is:

$$\begin{aligned} &x^n(t+\Delta t) = max\left\{ p^n(t+\Delta t)\right\} \\ &\quad s.t.\ \Delta x_L{}^n(t+\Delta t) \le p^n(t+\Delta t) - x^n(t) \le \Delta x_U{}^n(t+\Delta t), \Delta x_S{}^n(t+\Delta t), \Delta x_C{}^n(t+\Delta t) \end{aligned} \qquad (1)$$

where $\Delta x_L{}^n(t+\Delta t)$ is the minimum distance that vehicle *n* can travel in time interval $[t, t+\Delta t]$ given its limited deceleration capabilities; and $\Delta x_U{}^n(t+\Delta t)$, $\Delta x_S{}^n(t+\Delta t)$, and $\Delta x_C{}^n(t+\Delta t)$ are upper bounds due to its limited acceleration capabilities, safety, and comfort requirements. The safety bound $\Delta \boldsymbol{x_S{}^n}$ is similar to the one in Gipps (1981) and the comfort bound $\Delta \boldsymbol{x_C{}^n}$ to the CF(L) model in Daganzo (2006).

When changing lanes, the comfort constraint is neglected, and the safety constraint is modified to enforce safe separations among vehicles in the two relevant lanes. Lane-changes are classified as mandatory-time-related (T-changes), mandatory-space-related (S-changes) and optional (O-changes). This classification covers all possible scenarios involving HOVs. T-changes are made by LOVs exiting the HOV lane shortly before the restriction instant. S-changes are made by drivers trying to exit/enter the highway. They happen upstream of diverges and lane-drops. T- and S-changes are triggered at times and locations specific to each vehicle. O-changes are made by drivers trying to gain speed by switching to a faster lane; they are only a function of the drivers' sensitivity to speed differences across lanes. A multinomial trial at the beginning of each time interval determines if a driver wishes to start a lane change. If the outcome is positive but the lane change cannot be executed right away, the driver remembers his/her decision and continues to try in future time steps -- as long as the lane-change is still advantageous or necessary. If the driver is frustrated we assume that another driver eventually cooperates. Then, either the lane-changer or the cooperating vehicle decelerates at the given fixed rate to create a feasible gap. Once a lane-change has become feasible, it is executed in a single time step.

The car-following module (without lane-changing) is parsimonious. It includes three parameters from the FD (the flow-density diagram in the steady state): the free flow speed, the maximum flow, and the jam density; these are site-specific but can be measured without running the model. The model also requires acceleration and deceleration bounds. Average values can be used. The lane-changing module involves 4 additional parameters: a time lag before cooperation; a time threshold for T-changes, a space threshold for S-changes and the sensitivity to speed for the O-changes. These can be determined by fitting the model, and potentially transferred across sites[1]. Menendez (2006) shows that the model replicates field data as well as Laval and Daganzo (2006). The transferability across sites of its key unobservable parameters were tested, as described below, with data from two continents: an on-ramp merge in Southern California (Cassidy and Rudjanakanoknad, 2005); and a lane-drop near London (Bertini and Leal, 2005)[2].

Figure 2 shows empirical and simulated cumulative curves (N-curves on an oblique coordinate system) for 4 locations at the first site -- a northbound segment of Freeway 805 in San Diego, California, on October 15, 2002. The on-ramp had a variable input rate, as low as 400 vph when metered, as shown in the figure. For this site the simulation required 7 parameters: the two acceleration bounds, three parameters connected with the FD, the time lag before cooperation and the sensitivity to speed. The first five were chosen from other data sets. Only the last two were fitted to the data. The figure (and data underlying it) shows that the simulated bottleneck deactivation and activation times are within 4 minutes of the observed, and that the errors in the simulated discharge rates in the three relevant intervals are less than 3%. Furthermore, the lane changing rates in the three periods are predicted to within 4%. Unfortunately, this site was not a good testing ground for accumulations because the three detector stations were spaced very closely (0.075 mi apart).

---

[1] Ignoring T- and S-changes, the proposed model has only 2 more parameters than the model in Laval and Daganzo (2006): the deceleration bound and the time lag before cooperation.

[2] Neither of these sites has HOV lanes. Tests with HOV lanes could not be carried out because the available HOV data belonged to a rather complex site with many on- and off-ramps. Since it was impossible to discriminate between the different types of lane-changes and one had to "fit" and O-D table, success was assured. Therefore the test would have been meaningless. A rigorous test with HOV lanes is not essential, however, because the model with HOV lanes is structurally similar to the model without and does not require any more parameters.

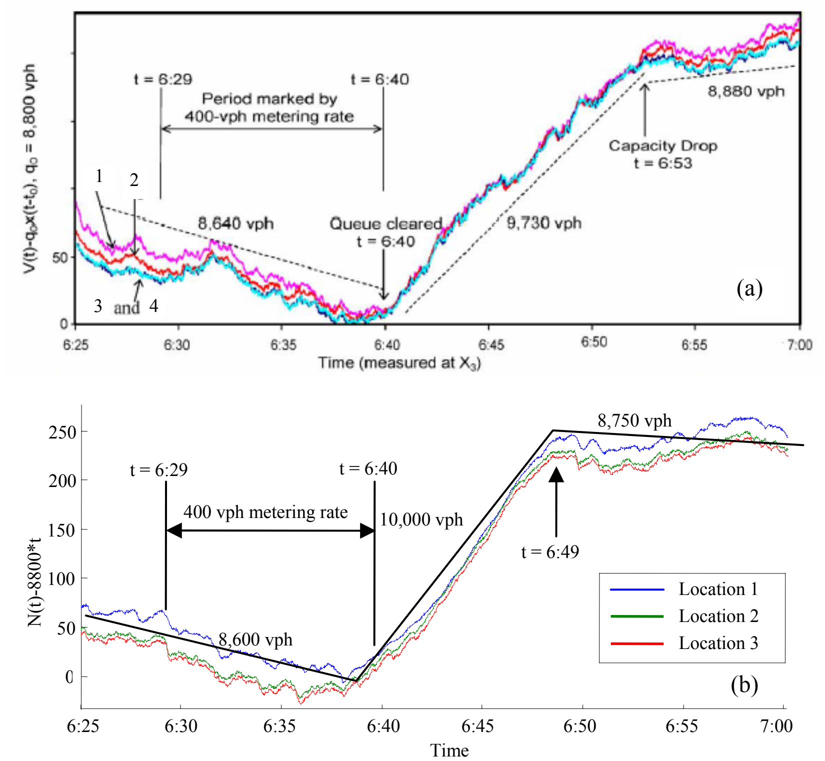


Figure 2. Oblique N-curves for 4 locations in southern California: (a) empirical data (source: Cassidy and Rudjanakanoknad, 2005); (b) simulation results.

Tests at the second site (London) were performed without fitting any parameters to the data; all seven parameters were the same as in San Diego. Remarkably, the predicted discharge rate of the bottleneck (observed for 20 min) was exactly as observed (3300 vph for two lanes). Here, the upstream detectors spanned a distance in excess of 2 miles and average accumulations were roughly predicted for distances over 1 mile. The model also predicted the period of the oscillations in the queue starting at the bottleneck and their propagation speed, but under-predicted their growth in amplitude; see Menendez (2006) for more detail.

Overall, the simulation appears to predict consistently the measures of performance that matter most for an evaluation of HOV lanes: lane changing rates and bottleneck discharge flows. Average accumulations over distances comparable with a mile are also well predicted. Therefore, the following section uses the proposed model.

4. THE EFFECT OF HOV LANES ON GP LANES AND BOTTLENECKS

Recall that the main question is: how does an HOV lane affect the capacity of the GP lanes and the overall capacity of bottlenecks? Also of interest are the effects of the HOV lane on the steady state flow-density diagram (FD) of the GP lanes. To this end we study here three different scenarios. Subsection 4.1 examines an idealized freeway without bottlenecks, where the only interference on the GP lanes is caused by the HOV lane. Vehicles enter the freeway seamlessly via the shoulder lane. This exercise uncovers the effect of an HOV lane on an idealized uninterrupted traffic stream without any damping by on- and off-ramps. Therefore, the results should be viewed as bounds. Then, sections 4.2 and 4.3 examine the effects of HOV lanes on the capacity of merge and diverge bottlenecks, taking realistically into account the interference from vehicles entering and exiting the freeway. This is important because most bottlenecks are at locations of these types.

4.1. An idealized ring without bottlenecks

We model here a 4-mile, 4-lane ring road with an HOV lane. On and off-ramps are not fixed in space. Instead, entrances and exits are assumed to take place uniformly along the whole length of the loop. In order to smear out and minimize their effect, entering vehicles are placed into gaps and given a non-disruptive following speed. Exiting vehicles are removed from the stream when they reach their exit location. Every entering vehicle is assigned a trip length, and the average trip length is adjustable. LOVs are assumed to leave the HOV lane between minutes 10 and 15 (since the restriction instant is t = 15 min). We also assume that the desired speed of the individual vehicles is uniformly distributed between 55 mph and 85 mph, and choose the remaining 6 parameters of the simulation to match those of Section 2.

For any average trip length we specify, we can maintain any desired HOV flow and any density in the GP lanes by modulating the input flows of both vehicle types. We can then record the average GP flow that can be sustained. By maintaining different GP-density/HOV-flow combinations for 20 min of real time, and recording the average flow on the GP lanes for each run, we obtain the graphs of Figure 3. Note: (a) there is a capacity reduction caused by HOV's crossing across the GP lanes when entering and leaving the freeway; (b) the largest capacity drop is 450 vphpl (from about 2800 to about 2350 vphpl) and it occurs when trip length is short and HOV demand is high; (c) the capacity exceeds 2350 vphpl in all cases; (d) the HOV lane has little or no effect when the system is undersaturated (on the rising branch of the FD); (e) in oversaturated conditions (inside queues) HOVs reduce GP density and increase GP speed for a given GP discharge flow; (f) effect (e) declines with increasing GP density and trends toward insignificance for flows below 1600 vphpl; and (g) the magnitude of effects (a) and (e) increases with the factors that increase lane-changing (more HOV flow and shorter trips).

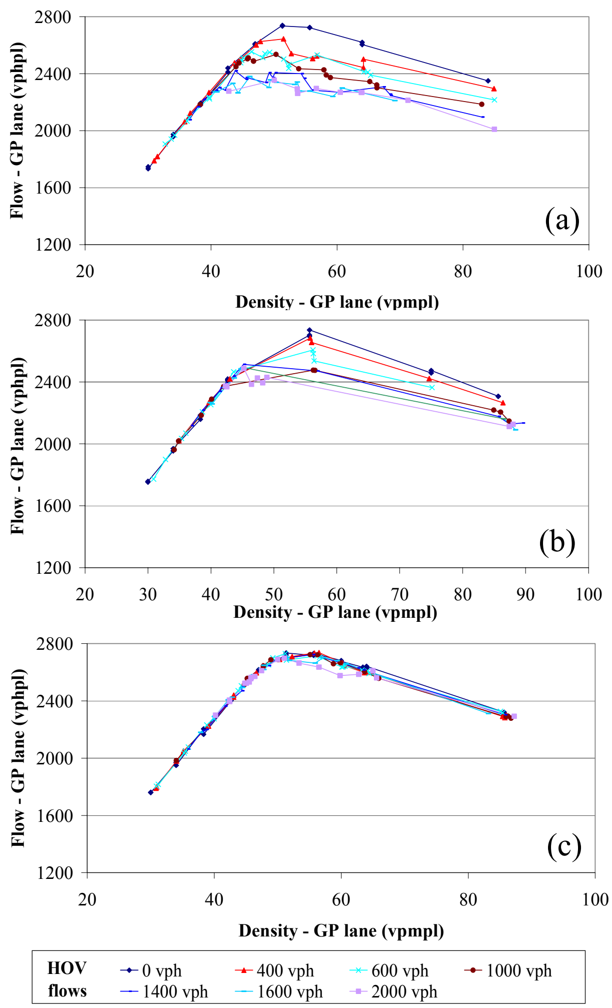


Figure 3. Fundamental diagram for a GP lane as a function of the flow in the HOV lane. Average trip length is: (a) 5-miles. (b) 10-miles. (c) infinite miles.

None of these patterns should be surprising. When traffic is near saturation or oversaturated, lane changers have to force gaps onto the GP lanes to complete their maneuvers, and this increases the average separation between cars for a given speed. Furthermore, one also expects the effects to be largest near saturation, as we found, because when the speed is high the gaps required for safe merging are wider.

We stress that the simulation exercise was designed to magnify and isolate the effects of an HOV lane on the GP lanes, and that the observed reduction in GP capacity is a "worst case". In real world situations with on- and off-ramps the capacity of the average GP lane with no HOV flow is much lower than in Fig. 3, because entering and exiting vehicles disrupt flow. These disruptions strongly mask the effects of the HOV lane, as shown in sections 4.2 and 4.3 below.

4.2. Merge bottlenecks

The worry with merge bottlenecks is that HOVs entering the freeway could create extra gaps in the GP lanes near the bottleneck -- by crossing over the GP lanes to access the HOV lane -- and in this way reduce the discharge rate of the GP lanes. The magnitude of this effect would depend on the proximity of these moves to the bottleneck, which is unknown.

The test here corresponds to the common type of merge bottleneck illustrated by Figure 4(a): an isolated merge of a 1-lane on-ramp with a 4-lane freeway with an HOV lane on the median side. The parameters of the simulation are as in Section 4.1. For any desired HOV flow, and any given on-ramp flow, we varied the percentage of HOVs and the upstream freeway demand so as to achieve a maximum sustainable discharge rate. We also simulated what would happen if the HOV restriction was eliminated. Figure 4(b) displays the results of this battery of simulations. It uses an on-ramp flow of 1000 vph, but on-ramp flows of 500 and 1500 vph yield virtually identical results. Note: (a) HOV flow has no discernible effect on the discharge rate of the GP lanes; (b) the average discharge rate is about 2050 vphpl for all HOV flows, smaller than in the idealized case of Section 4.1, but equal to the case with no HOV flow at all; and (c) if the HOV restriction is eliminated the average discharge rate of the four lanes now open for general use remains at 2050 vphpl – the HOV lane now carries about 2400 vphpl and the three other lanes average about 1950 vphpl, as shown.

These results surprised us; in particular (c). Result (b) occurs because HOVs coming from the on-ramp do not change lanes near the bottleneck any more than LOVs and because HOVs from the freeway do not interact with the GP lanes. Result (c) is explained because in our simulations lane 2 interacts with lanes 1 and 3 without the HOV lane restriction, but only with lane 3 with the restriction. Overall, the results show that the capacity of the lanes available for general use is the same, whether or not the HOV lane is on. Since median lanes can often carry considerably more traffic than 2050 vphpl, this suggests that a merge's total discharge rate could be increased by providing an HOV lane if the demand of HOVs were to exceed 2050 vph. Of course, this demand level is unlikely high but we will see in Section 6 that there are ways to artificially boost it and increase merge capacity.

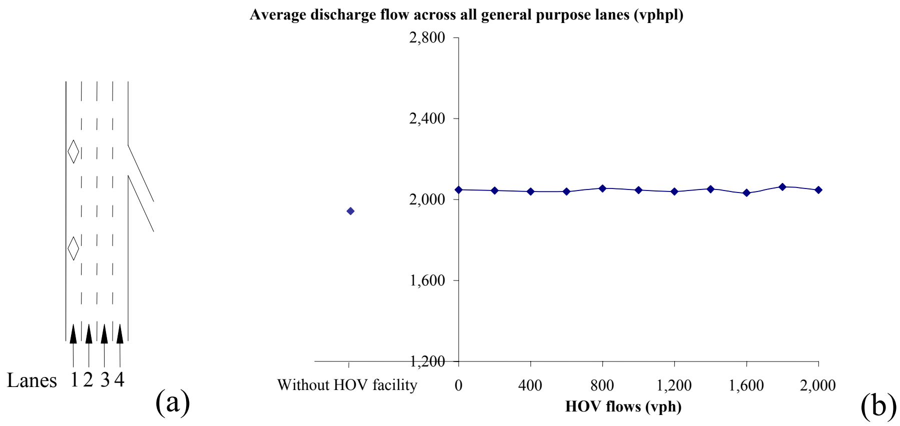


Figure 4. Effect of an HOV lane on the GP lanes at a merge bottleneck: (a) merge; (b) discharge rates

4.3. Diverge bottlenecks

Diverge bottlenecks are more complicated, because (unlike merges) their discharge rate heavily depends on the split of the demand between branches and the split of the lanes. We focus here on isolated diverges, free of weaving effects from upstream and downstream junctions, and with an arbitrary number of lanes, l and l', per branch. Let β and β' denote the fractions of traffic destined for each branch of the diverge ($\beta + \beta' = 1$), and Q the capacity of one lane. If significant queues form we expect them to be FIFO (Munoz and Daganzo, 2002); and the discharge flows per branch, μ and μ', to be as follows (Daganzo, 1994):

$$\mu/l = Q \quad \text{and} \quad \mu'/l' = [(\beta'/l')/(\beta/l)]Q \le Q\,, \qquad \text{if} \quad \beta'/l' \le \beta/l. \tag{2}$$

The branch carrying the larger flow per lane is called the critical branch. If $\beta/l < \beta'/l'$ formula (2) still applies, with the primes switched around. Thus, according to (2) the critical branch always discharges at capacity. The expression for the non-critical branch hinges on the formation of FIFO queues upstream of the diverge point.

If HOV lanes were not to change Q and if the fractions of HOVs and LOVs taking the critical branch were to be the same, then an HOV lane passing through the non-critical branch would have no effect on discharge flows. The reason is that the fraction of traffic discharging on the non-critical side is still determined by the FIFO rule, which continues to apply to HOVs and

LOVs, albeit separately. If on the other hand the HOV lane passes through the critical branch, flow would be reduced even if Q was fixed. In both cases, though, a drop in Q would imply a reduction in total discharge flow. This is a concern because vehicles leaving the HOV lane immediately upstream of the split can create gaps in traffic as they weave their way to the other side, and these gaps could reduce Q. Our tests, described below, dispel this worry.

We experimented with the drastic scenario of Figure 5(a): a 4-lane freeway that drops two lanes at an exit ($l = l' = 2$). Simulation input parameters are again as in sections 4.1 and 4.2. Control variables were the exit fraction β (assumed to be the same for HOVs and LOVs), and the total HOV flow (upstream). We recorded the flows on all parts of the bottleneck for $\beta = 0.3$ and 0.7, with total HOV flows ranging from 0 to 2000 vphpl. Figure 5 displays some details.

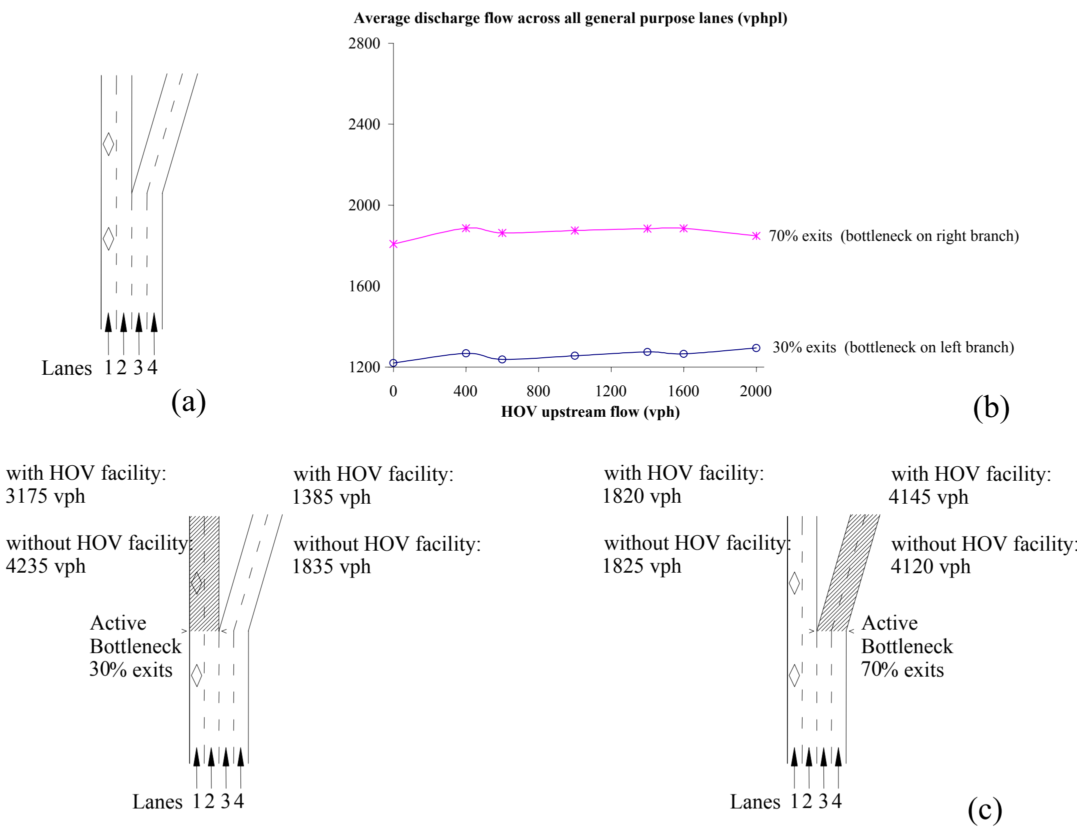


*Figure 5. Effect of an HOV lane on a diverge bottleneck: (a) diverge; (b) effect on the flow of the GP lanes; (c) effect on the total discharge flows (total HOV flow =*1000 *vphpl).*

In all cases tested total HOV flow had little effect on Q. Its value was 2400 ± 150 vphpl for the GP lane next to the HOV lane (arising when the HOV branch was critical: $\beta = 0.3$) and 2050 ± 100 vphpl, both, for the GP lanes on the exit (arising when $\beta = 0.7$) and also for all lanes when the HOV was not on. Curiously, proximity to an HOV lane increases Q. (We expect this phenomenon to be less pronounced for configurations with more freeway lanes.) In any case, this evidence suggests that we can assign fixed, lane-specific capacities to the GP lanes and that a generic value ($Q \cong 2000$ vphpl) can be used as a conservative approximation.

We also found the following: (a) when the exit is critical ($\beta = .7$) the total discharge rate of the bottleneck (or any of its branches) does not change when the HOV lane opens regardless of the HOV flow—this is illustrated by Fig. 5c; and (b) when the freeway is critical ($\beta = .3$) the changes in total discharge rate depend heavily on the HOV flow. We will see in section 5 that items (a) and (b) are predicted by a simple formula.

5. THE ENGINEERING OF HOV LANES

Let us now put these results into practice and discuss implementation issues. Since HOV flows do not significantly change the capacity of GP lanes at the isolated bottlenecks tested in Sec. 4, and since there is no reason to expect different results at other isolated bottlenecks, we shall assume that Q is fixed.

Planning: We can test whether an HOV lane would create a bottleneck at an uncongested location by comparing the total demand for the GP lanes versus their capacity with the HOV restriction in force. Let L be the number of lanes, q the average vehicle flow across all lanes before the HOV restriction, and q_H the flow on the HOV lane after the restriction. Then, the balance equation is:

$$[\text{GP-demand} \equiv Lq - q_H] \;\leq\; [\text{GP-capacity} \equiv (L\text{-}1)Q]. \qquad (3)$$

The GP demand is the flow of LOVs plus that of any HOVs on the GP lanes. For example, at an on-ramp merge where the HOV lane is on the far side from the on-ramp, all HOVs from the on-ramp are initially in the GP lanes. Thus, (3) should be evaluated at these merges assuming that q_H is the flow of HOVs upstream of the merge. Something similar occurs for off-ramp bottlenecks. Here we must assume that exiting HOVs are on the GP lanes immediately upstream of the off-ramp; hence q_H is the flow of HOVs that stay on the freeway beyond the diverging point.

The balance equation can be rewritten in terms of the difference between the total flow the freeway is required to carry and the maximum possible; i.e., as:

$$[\text{Overflow}] \;=\; [\text{Required flow} \equiv Lq] \;-\; [\text{Maximum possible flow} \equiv q_H + (L-1)Q] \;\leq\; 0. \qquad (4)$$

This leads to:

Rule 1: *If* (4) *holds everywhere along an HOV lane and traffic is uncongested, the HOV lane will not create a queue or reduce flow anywhere. Otherwise, a queue is created where* (4) *is first violated.*

We conjecture that Rule 1 is violated frequently in the real world, perhaps at the site in Sec 1.

We now show that (4) also guarantees no reduction in flow at the termination point of an HOV lane opened in queued traffic; e.g., as in Fig. 6. Since LOVs must move into the GP lanes prior to the restriction instant, density would initially increase in these lanes, ensuring that they remain queued. Thus, the flow that the HOV system would emit at its downstream end assuming that no queue blocks its release (the maximum possible) is still as in (4): $q_H + (L-1)Q$. Now, if this value is less than the *required flow* to sustain the downstream queue, which is Lq still as in (4), the back of this queue would recede forward. Otherwise it will stay pinned to the termination point and downstream flow will not change. Thus, we can still use (4) to determine whether the HOV lane would reduce flow at its downstream end.[3]

Rule 2: *If* (4) *holds at the termination point of an HOV lane that is in congested traffic at the restriction instant, the HOV lane does not change the downstream flow. Otherwise the total discharge flow is reduced by the overflow.*

We can combine Rules 1 and 2 to determine whether an HOV lane passing through an active bottleneck at the restriction instant reduces the bottleneck's total discharge rate. Since Rule 1 applies downstream and Rule 2 applies upstream, we have:

Rule 3: *The maximum of the upstream and downstream overflows* (4) *at an active bottleneck, if positive, is the reduction in the total flow caused by the HOV lane at the location of the maximum.*

For lane drops and merges the maximum always occurs downstream and $q \cong Q$. Thus, the overflow is $Q - q_H \geq 0$, and there will usually be a reduction in (downstream) discharge flow. To avoid it we can end the HOV lane upstream of the bottleneck, as in Fig. 6, where experience shows (Cassidy and Mauch, 2001) that the required flow is lower and the overflow could vanish. This still allows HOVs to bypass most of the queue.

---

[3] The interpretation of the congested case is slightly different, though. The term "maximum flow" refers to that which can be emitted, whereas in the uncongested case it refers to that which can be received. Likewise, the term "required flow" refers to that which is necessary to sustain the queue, whereas in the uncongested case it refers to that needed to satisfy the demand. Curiously, the same balance equation applies to both cases.

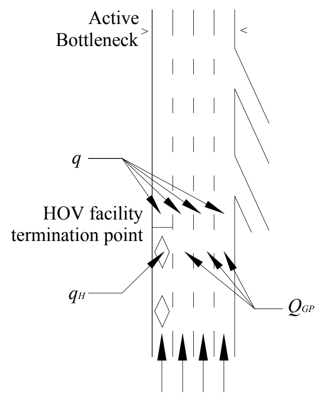


Figure 6. Termination of and HOV facility upstream of a bottleneck.

For simple off-ramp FIFO bottlenecks (without lane drops) the constraining location is always upstream. If the flow on the off-ramp (the critical branch) is q_r, then the required freeway flow is $Lq \equiv q_r/\beta$. This value can be so small (see e.g., Muñoz and Daganzo, 2002) so as to make the overflow negative. Then, the HOV lane does not change the bottleneck flows. Otherwise, total flow is reduced. The constraining mechanism is a new bottleneck created upstream, where exiting HOVs merge into the GP lanes.

Rule 3 can also be used for composite bottlenecks combining merges or diverges and lane drops, as in the example of section 4.3. Here the constraining location can be either upstream or downstream. The reader is encouraged to verify that Rule 3 predicts the changes in bottleneck flows for the two examples of Fig. 5c; i.e., that: (i) when $\beta = 0.3$ the maximum overflow is downstream and is 1060 (taking $Q = 2475$), matching the reduction in downstream flow; and (ii) when $\beta = 0.7$ both overflows are negative, predicting no reduction, as observed.

In summary, we see from all this that the main function of HOV lanes is bypassing queues. If they would reduce bottleneck flow, they should be terminated upstream of it, at a location where entering flows cutting into the downstream queue have reduced the total flow enough to eliminate the overflow.

Diagnosis of flow and speed effects: Our rules can also be used to diagnose existing HOV lanes. For example, Rule 1 implies the following: if traffic is uncongested on all freeway lanes including an HOV lane and we cancel the HOV restriction, then flows will not change anywhere. Rule 2 implies: if traffic is congested both next to an HOV lane and downstream of it, then downstream flow will not change if the HOV restriction is canceled. Rule 3 implies: if an HOV lane passes through an active bottleneck and the HOV restriction is eliminated, total bottleneck flow will increase by the maximum overflow. Note: the *maximum flows* going into the overflow calculus are observed quantities and the *required flows* are also known; the latter are: (i) LQ for merges and lane drops, and (ii) the sum of the observed upstream flows on all lanes for off-ramp bottlenecks. These diagnostics can help us determine whether an existing HOV lane is increasing total vehicle-hours.

We can also diagnose the travel time savings to HOVs as follows. Assume that our L-lane freeway is operating with an HOV lane in a steady state with average flows per lane on the two sets of lanes: q_H (uncongested) and q_{GP} (congested). The average speeds on the two sets of lanes are also known. Now assume that the HOV restriction is eliminated in a long middle section of the freeway. The resulting 3-section system will then settle in a steady state with the same aggregate flow everywhere (vehicles are not entering or leaving), and a near-uniform speed distribution across all lanes in the middle section. Since the average flow per lane in the middle section is known: $q = [q_H + q_{GP}\ (L-1)]/L$ we can estimate the common speed to all cars in this section from the conventional FD, and compare it with the known speeds on the HOV and GP lanes.

To test this idea, let us apply it to the data immediately prior to 7:00 PM in Fig 1. Using $q_H = 1400$ vph and $q_{GP} = 1500$ vph, as approximately shown by the figure, we find that the speed across all lanes without the HOV lane (i.e., after 7:00 PM) should barely change from about 22 to 22.2 mph. This is roughly as observed. The match is not perfect because total flow was increasing at the time. But in any case, since the "before" speeds are roughly 45 mph for HOVs and 22 mph for LOVs, and about 23 mph "after" (same for all vehicles), we see that the HOV lane was saving HOVs 1.3 minutes per mile traveled with only a small penalty to the LOVs. Small penalties should be generally expected whenever L is large because then $q \cong q_{GP}$.

6. DYNAMIC CONTROL

We now examine a control strategy that allows us to extend HOV lanes through merge and lane-drop bottlenecks while actually increasing their capacities. We illustrate the approach with a simulation of a lane-drop bottleneck from 3 to 2 lanes with an HOV facility in the median lane. Only the HOV lane and one GP lane continue beyond the lane drop. The input parameters are those used elsewhere in this paper. The HOV restriction starts at $t = 15$ min; demand is 1600 vph for HOVs and 3000 vph for LOVs.[4] Assume first that the bottleneck is uncontrolled. Figure 7(a) shows simulated, oblique N-curves for 5 detector stations. Stations 1 through 3 are upstream of the bottleneck; stations 4 and 5 are downstream. Figure 7(b) shows oblique N-curves for the HOV lane and its neighbor at station 5, downstream of the bottleneck.

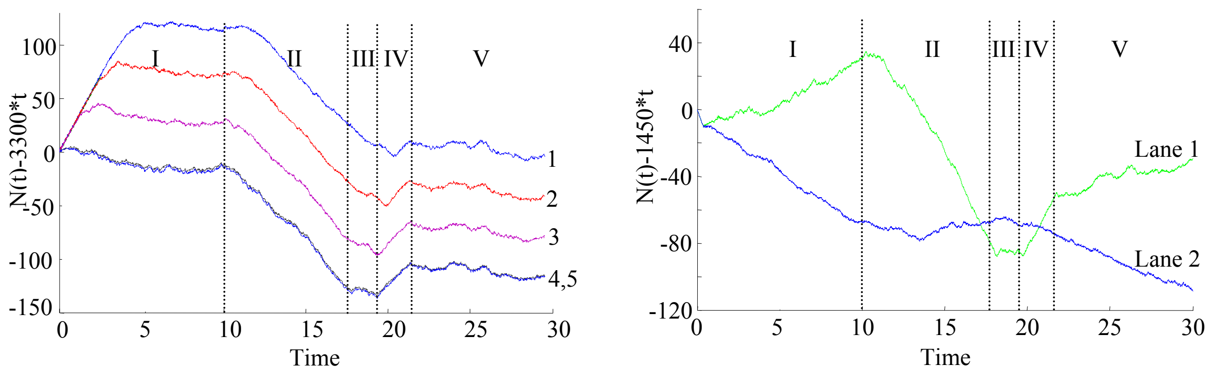


Figure 7. Simulated N-curves shown on oblique coordinates: (a) average curves across all lanes at 5 stations; (b) lane-specific curves for the HOV lane (1) and its neighbor (2) at the most downstream station.

The roman numbers in both graphs correspond to the different stages of the system: (I) shortly after $t = 0$ the queue initially appears and the bottleneck starts discharging at capacity. (II) At $t = 10$ LOVs begin to exit the HOV lane according to their individual time thresholds, and the discharge rate of lane 1 declines by almost 60%. (III) In this interval all LOVs have exited the HOV lane, but some HOVs are still moving into it. The HOV lane becomes queued and the overall discharge rate is similar to that from phase I at the beginning. (IV) Now all lane-changes into and out of lane 1 have stopped but a residual HOV queue remains close to the bottleneck; note that the discharge rate of lane 1 is about 35% higher than in phase I. (V) Finally, the HOV queue dissipates; now the discharge rate of lane 1 matches the HOV demand and this state can be sustained indefinitely. But, we can do better!

The high outflows of phase IV occur with a queued HOV lane and no lane-changes into or out of it. Our simulations show that they can be sustained with the dynamic HOV lane strategy proposed in Daganzo et al. (2002). In this strategy LOVs are allowed to use the HOV lane through the bottleneck intermittently. Either a pre-timed or an actuated set of variable message signs (or signals) are assumed to control the settings. The purpose is to create a short queue in

---

[4] The high percentage of HOVs (35%) is used to better illustrate the different phases of our bottleneck. As we shall see, a lower percentage would actually magnify the benefit of the HOVstrategy.

the HOV lane at all times that would help sustain a high discharge rate for lane 1. With this approach, HOVs would suffer a trivial penalty, but it is more than compensated by the increased discharge rates (improved capacity), and the large time savings to LOVs. Figure 8 shows the result of a simulation with a dynamic HOV section extending from 0.20 miles to 0.70 miles upstream of the bottleneck. (These dimensions help reduce the number of lane changes near the bottleneck, which could reduce the discharge rate.)

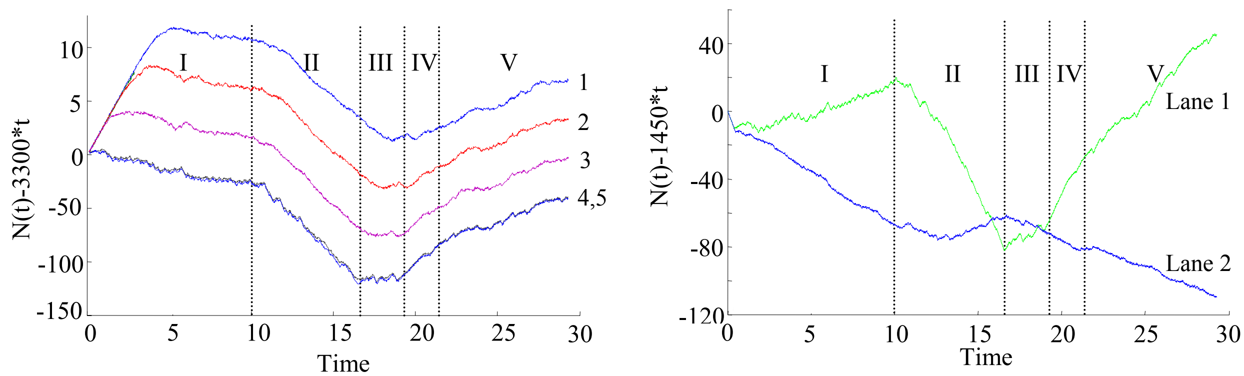


Figure 8. Oblique N-curves simulated with a dynamic HOV lane: (a) average curves across all lanes at 5 stations; (b) lane-specific curves for the HOV lane (1) and its neighbor (2) at the most downstream station.

States I through IV are the same as before, but final state V achieves and sustains a higher discharge rate than without control. We were surprised by the magnitude of the improvement. The discharge rate during state V is only slightly lower than that of "super-charged" state IV, and about 20% greater than the total capacity prior to the opening of the HOV lane (state I). This shows that it is theoretically possible, at least in some situations, to give priority to HOVs and at the same time <u>improve</u> system capacity. Note as well that the dynamic strategy would have improved the conventional operating mode even more if the percentage of HOVs had been lower; this would have lowered the flows of state V in Figure 7(a) but not those of Figure 8(a).

This example is just a simple proof of concept. An optimum control strategy would be site specific and its performance could vary with conditions. This should be explored in practice. The results of this simulation also suggest that, even if there are no HOV lanes, it could be helpful to prohibit lane-changes near lane-drops or merges.

7. DISCUSSION AND FINAL REMARKS

This paper used a simulation to study how HOV lanes affect the performance of adjacent GP lanes and nearby bottlenecks. The model assumes that drivers are aggressive, consistent, identical, myopic, and have limited information. However, despite its simplicity, it reproduces the most important features of traffic with just a few parameters. It is fail-safe and can be fitted easily. The model accurately reproduces important phenomena without a need for re-calibration. It was used to predict outcomes for a variety of scenarios concerning HOV lanes and to test new control strategies. Field tests should verify or disprove these predictions.

Our results show that non-separated HOV facilities do not significantly affect the capacity of GP lanes and that if properly engineered they do not hinder bottleneck outputs. In some cases they increase them. Figure 3 also shows that for the same queued flow (e.g., as would be observed at a fixed location upstream of a bottleneck) GP lanes are less dense next to an HOV lane than otherwise. This is undesirable because less density means less efficient storage of GP queues. Although the effect is small for GP flows of the (small) magnitude expected next to properly deployed HOV lanes, these lanes do take away one lane for queue storage and therefore spread queues more widely over the network. This is their main drawback. Longer queues reduce system output if they block busy off-ramps. This could be a problem for some systems with multiple exits or bottlenecks, but this should be assessed through further research.

Overall, we find that HOV lanes are not inherently deleterious to traffic if properly deployed. We agree with the findings in Chen *et al.* (2005) that the free-flow speed on HOV lanes can be as low as 45 mph when the GP lanes are queued, and that this is a drawback, but we also find that the drawback is minor. A 3-mile trip with the data of Fig. 1 (discussed in Sec. 5) would take 3 min at 60 mph, 4 min at 45 mph, and 8 min at 22.5 mph. Thus, an HOV experiences a 1-min delay instead of 5-min for a 3-mile trip on the HOV lane. Clearly, a properly deployed HOV will always reduce the delay to HOVs, not necessarily eliminating it, but it will not increase overall system vehicular delay either if queue storage effects are not an issue. A well-designed HOV simply allocates the invariant vehicular delay to low occupancy vehicles. This is useful to society because it reduces the number of people-hours of delay. But for this to happen, HOV facilities must be implemented properly. Otherwise, they can create bottlenecks whose damage can be enormous. Rules 1, 2 and 3 of Section 5 can be used to avoid these pitfalls. We recommend that traffic agencies take a detailed look at all bottlenecks next to HOV facilities to verify that they indeed are not exacerbated, or entirely due to the HOV facility. We suspect that the I80/I580 bottleneck is in this undesirable category. The findings in Chen *et al.* (2005) suggest that there are many similar sites spread through California.

ACKNOWLEDGEMENT

The comments of Prof. M. J. Cassidy are gratefully acknowledged. Research supported by the Berkeley Center of Excellence on Future Urban Transport.

REFERENCES

Bertini, R.L. and M.T. Leal (2005) "Empirical study of traffic features at a freeway lane drop" *Journal of Transportation Engineering*, **131**(6): 397-407.

Brackstone, M. and M. McDonald (1999) "Car-following: a historical review" *Transportation Research Part F*, **2**(4): 181-196.

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

Cassidy, M. and J. Rudjanakanoknad (2005) "Increasing capacity of an isolated merge by metering its on-ramp" *Transportation Research Part B*, **39**(10): 896-913.

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. (1994) "The cell transmission model part II: network traffic" PATH Working Paper UCB-ITS-PWP-94-12, University of California, Berkeley, CA (abridged in *Transportation Research part B,* **28**(4): 269-287, 1995).

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.

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

Gipps, P.G. (1981) "A behavioural car following model for computer simulations" *Transportation Research Part B*, 15(2):105-111.

Hoogendoorn, S. and S. Ossen (2005) "Parameter estimation and analysis of car-following models" Proceedings of the 16th International Symposium on Transportation and Traffic Theory (H. Mahmassani, editor) pp. 245-265, College Pak, Maryland, Elsevier, London, U.K.

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. (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.