

UC Berkeley

Research Reports

Title

Empirical Reassessment of Traffic Operations: Freeway Bottlenecks and the Case for HOV Lanes

Permalink

<https://escholarship.org/uc/item/31h8z81t>

Authors

Cassidy, Michael J.
Daganzo, Carlos F.
Jang, Kitae
et al.

Publication Date

2006-12-01

Institute of Transportation Studies
University of California at Berkeley

Empirical Reassessment of Traffic Operations: Freeway Bottlenecks and the Case for HOV Lanes

Michael J. Cassidy, Carlos F. Daganzo, Kitae Jang

Department of Civil and Environmental Engineering and the Institute of
Transportation Studies, University of California, Berkeley

Koohong Chung

California Department of Transportation, Highway Operations Special
Studies, District 4, Oakland California

RESEARCH REPORT
UCB-ITS-RR-2006-6

December 2006
ISSN 0192 4095

EXECUTIVE SUMMARY

An earlier empirical study of San Francisco Bay Area freeways concluded that HOV lanes unfavorably affect freeway traffic by creating congestion. That study attributed the observed congestion to HOV lanes and tentatively recommended their elimination over the full lengths of the freeways it examined; and even from all Bay Area freeways. It recognized, however, that its analysis is fragmentary and recommended further work to solidify its conclusions. This is logical since the study lacks a spatiotemporal analysis to pinpoint where and how congestion first forms (at bottlenecks).

The present report re-examines the same set of freeway sites in spatiotemporal detail to understand more deeply how HOV lanes are affecting traffic. It enriches the data from the original study with data from neighboring detector stations, to identify: first the locations (bottlenecks) where queues are triggered; and second the role that HOV lanes play in this phenomenon. This study includes an even more detailed analysis of high-resolution video data from a bottleneck where the HOV lane initially seemed to be having an unfavorable effect.

To our surprise, we found no compelling evidence that the HOV lanes were triggering delays and queues on the freeway sites in the earlier study. In all cases queues formed first at bottlenecks and, save for one questionable case, formed for reasons unrelated to the HOV lanes. This was true even on additional days that we studied. Moreover, data did not conclusively show that HOV lanes were reducing bottleneck flows or prolonging the queues; no adverse effects could be confirmed. To the contrary, and quite remarkably, the HOV lane seemed to *increase* the capacity of the bottleneck that was videotaped, even though that lane was underutilized. (The video data show that higher than normal discharge flows arose in the remaining lanes when the HOV lane was underutilized – enough even to compensate for that lane’s underutilization. Reassuringly, this effect had been predicted in earlier simulations.)

Although the sites assessed in this and the original study may not contain bottlenecks where the HOV lane is contributing to problems, the present study recognizes that such bottlenecks exist. They are just less prevalent than originally suspected. Fortunately for society, an HOV lane can be useful in most of these cases: theory and simulations indicate that if an HOV lane is rescinded near a problematic bottleneck *but is preserved on the entire queued freeway stretch upstream*, the lane will neither affect the bottleneck’s discharge flow nor the delay to vehicles that pass through the bottleneck. For typical freeways with four or more lanes, delay to

all vehicles would change little. The HOV lane would allow HOVs to bypass most of the congestion without significantly increasing total vehicle-hours of travel.

This means that even in problematic cases, HOV lanes can usually be preserved and allowed to perform their intended societal function: reducing people-hours of travel without significantly increasing vehicle-hours of travel by favoring HOVs where freeway queues arise. To avoid increased vehicle delays, however, and perhaps even to increase bottleneck capacities, some HOV-lane installations should be modestly altered near bottlenecks. This report also describes field studies necessary to determine where and how such alterations should be deployed.

1. INTRODUCTION

High Occupancy Vehicle (HOV) lanes are deployed on urban freeways for the exclusive use of vehicles that carry more than a predetermined number of occupants, and are often activated during rush hours only. Much attention has been devoted over the years to the planning, design and operation of these facilities. The efforts, though numerous and widespread, seem not to have settled debate on the subject. While advocates may cite, for example, an HOV lane's potential to reduce person-hours of commuter delay, critics will often counter by citing traffic problems presumably caused by the underutilization of these lanes. The persistence of the debate suggests that the impacts of HOV lanes are not fully understood.

In California, the debate has intensified. This has been fueled in part by a study titled "*Empirical Assessment of Traffic Operations*" (Chen, et al, 2005), henceforth referred to as "EATO". The EATO study evaluates the impact of HOV lanes on overall freeway congestion, including the congestion in General Purpose (GP) lanes. It concludes that in the San Francisco Bay Area, non-separated HOV lanes increase overall congestion, and that a better way to manage freeways includes the elimination of these lanes.¹

The main evidence in EATO is a reduction in freeway speeds observed at all sites between 15:00 and 19:00 hrs; i.e., when non-HOVs (Low Occupancy Vehicles, henceforth called LOVs) are prevented from using the median lane. The evidence can appear persuasive because the pattern occurred at all the EATO sites and, in some cases, the changes in speed were large and nearly synchronous with the 15:00 and 19:00 times. The evidence was (to us) unexpected and therefore scientifically intriguing. Since the EATO study also cautions that its analysis is fragmentary and needs to be completed with a more comprehensive data set, another look was in order. After all, a deeper understanding of the issue is important not just for science, but for

society: if the EATO findings could be corroborated, cities would have at their disposal a silver bullet (removing HOV lanes) for reducing congestion.

In view of the above, this report reassesses the EATO sites in detail. It complements the EATO data with data from nearby detectors (and video data in one instance) to provide a more complete spatiotemporal picture of the traffic evolution at these sites. This approach unveils the relative contributions of bottlenecks and HOV lanes to the observed congestion.

In the end, we could not confirm that HOV lanes negatively affected the EATO sites; we found no silver bullet. There is something positive in our findings, though. Data show that in some cases, HOV lanes are *enhancing* bottleneck flows; and simulations indicate that this effect can be magnified with new control strategies. In other cases, where HOV lanes do create problems, modest changes in installation can preserve the benefits of these lanes without significant adverse effects. The nuanced changes proposed in this report are not only easier and less expensive to implement than the wholesale elimination of HOV lanes; they are also more effective. The changes would enable HOVs to travel at nearly all times and almost everywhere on a freeway at reasonable speeds (usually in excess of 40 mph and faster than LOVs) without significantly increasing (and possibly even decreasing) vehicle-hours of delay.

Section 2 of this report provides background on the EATO study: it explains how, by ignoring the spatial component of the system, the EATO analysis is incomplete. (The EATO analysis also uses questionable logic, and this is discussed in the appendix.) Section 3 re-examines the EATO sites in spatiotemporal detail and shows that, in all cases, queues first formed at bottlenecks; and that, save for two possible exceptions, no unfavorable contributions from the HOV lanes were detected. Section 4 takes a closer look at one of the bottlenecks where, on the surface, the HOV lane seemed to be having an adverse effect; and shows instead that the effect may have been favorable. Section 5 offers practical prescriptions.

2. BACKGROUND

The EATO study used time series of vehicle speeds and flows (by lane) at single loop detector stations as metrics to assess HOV-lane impacts on freeway traffic. Although these data revealed periods when queues persisted atop a detector, similar data were not collected at neighboring detectors. As a result, the EATO study could not, and did not, determine whether the queues grew simultaneously everywhere along a freeway, or instead formed first at distinct bottlenecks.² To see how this omission colored the analyses, refer to Figs. 1(a) – (f). These charts reproduce the

speed time-series data in Fig. 8 of EATO and characterize all of the sites it studied. Our charts include additional annotations to aid in their interpretation.

On all the freeways of Figs. 1(a) – (f), the median lane is reserved for HOVs on weekdays from 15:00 to 19:00. This period is demarcated by vertical lines in each chart. In all cases, speeds are lower during that period than outside it, both in the HOV lanes (dark curves) and in the adjacent GP lanes. Because speeds drop more in the latter lanes, we focus on these first. (We will discuss HOV-lane speeds in Sec. 5 and in the appendix.)

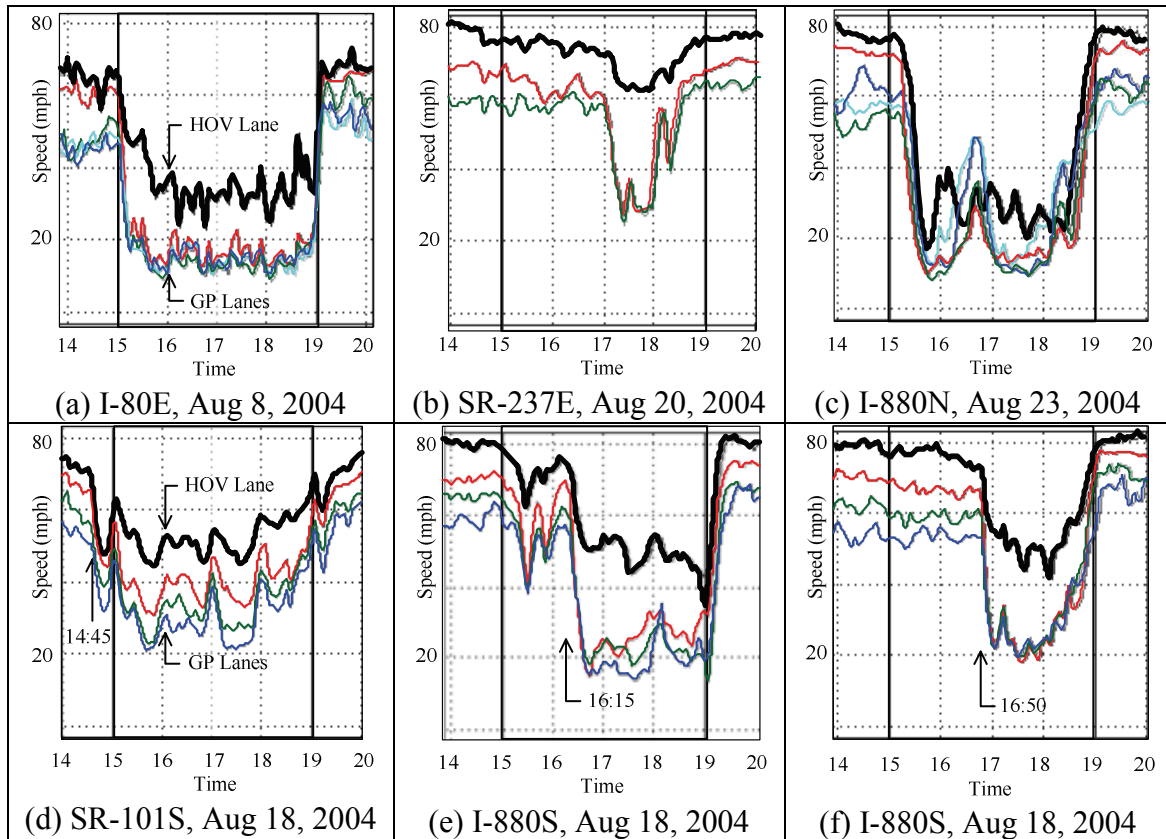


Fig. 1. Time-Series Diagrams of Speed Furnished in EATO

The EATO study claims that queues arose in the GP lanes because these lanes were eventually in short supply; i.e., demand for LOVs grew while the median lane was unavailable for general use, pushing the GP lanes into the congested regime. Note how this could leave a reader with the impression that long freeway queues can suddenly appear at the start of a rush.

The evidence of this mechanism is said in EATO to come after 19:00 hrs because by this time, when each HOV restriction had been lifted, the speeds reportedly increased. EATO's conclusion is that speeds rose because the median lane was no longer squandered on HOVs.

The evidence, however, is not as stated in EATO. In four of the cases (Figs. 1(b), (c), (d) and (f)) speeds began to recover *before* the HOV restriction expired at 19:00, and in three of these (Figs 1(b), (d) and (f)) recoveries began around 18:00 -- an hour before the restriction's expiration time. We cannot conclude that lifting the HOV restriction increased the speeds (removed the queues) in these cases: an effect cannot precede its cause.

Of course, lack of evidence around 19:00 does not mean that HOV lanes can not be causing other problems at other times. Therefore, a detailed spatiotemporal examination of the events of Fig. 1 is furnished in the following section. The section will show that: (i) the speed recoveries evident in Figs. 1(b) – (f) were not triggered by the expiration of HOV lanes, but by reductions in traffic demand;³ (ii) the speed reductions at the beginning of the rush were not induced by HOV lanes, but by other factors, including recurring bottlenecks, a traffic accident and freeway construction work; and (iii) the reductions always began at definite locations (bottlenecks), meaning that long freeway queues did not suddenly appear.

3. REASSESSMENT

Let us now consider the cases of Fig. 1 in reverse order. The reader wishing to skip over details can read the last paragraph in each subsection without loss of continuity.

3.1 The case of I-880S: Figs. 1(e) and (f).

The last two charts of Fig. 1 come from neighboring detectors on the same freeway (I-880S) during the same day and time, though the EATO study did not analyze their relationship. (It mistakenly states that the data were from different freeways.) Note how the speed curves of the two figures trace different patterns, even though they come from the same collection of drivers. Note in particular that speed dropped much earlier at one location than at the other, belying the notion that the HOV lane suddenly triggered a long queue on this link. The asynchronous drops suggest instead that the problem first started at a specific location. We now find that location, and then try to determine if the HOV lane is the culprit.

Figure 2(a) reveals the location. It displays a 6-hour long time-space diagram of detector occupancies (a dimensionless measure of density) for a 19-mile freeway stretch that includes the measurement locations of Figs 1(e) and (f). Darker shades denote higher occupancies.⁴ The speeds of these two figures and the occupancies of Fig. 2(a) all span the same 6-hour observation period.

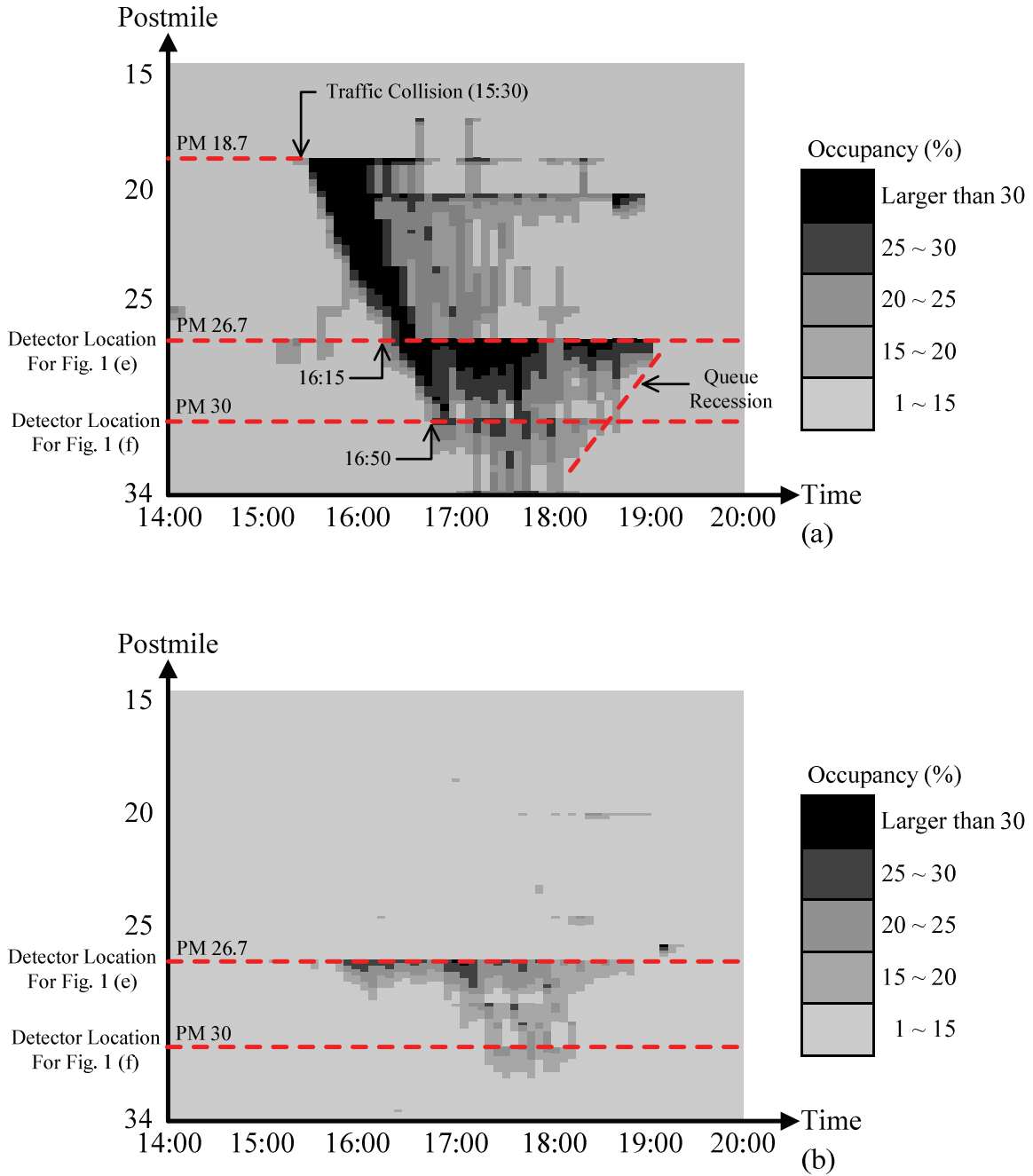


Fig. 2. Time-Space-Occupancy Plots, I-880S (a) Aug 18, 2004, (b) July 13, 2004

Figure 2(a) clearly shows that the queue did not form or dissipate simultaneously over the freeway. Instead, the slanted, diagonal patterns of the darker shadings show that the queue started locally at Post Mile (PM) 18.7, around 15:30 hrs. The queue then grew, eventually causing the speed reductions visible in Fig. 1(e) around 16:15, and then the reductions of Fig. 1(f) around 16:50.

The accident log of the California Highway Patrol indicates that this queue was triggered by a vehicle collision and not by the HOV lane. A record of the collision is archived in the PeMS website; see <http://pems.eecs.berkeley.edu>.

Figure 2(a) also shows that, after this collision was removed, a second bottleneck emerged on the freeway from 17:30 onward at PM 26.7. This is the location of a merge. Note the high occupancies upstream of this location, with freely flowing conditions downstream. Later in the rush, but still prior to 19:00, the back of the queue gradually receded forward toward this second bottleneck for lack of demand, and eventually dissipated. Thus, the gradual speed recovery seen in Fig. 1(f) was due to a reduction in traffic demand, and not to the expiration of the HOV restriction.

The speed recovery nearer this bottleneck (Fig. 1(e)) accompanied the expiration of the HOV restriction. This was a coincidence, although the elimination of the HOV restriction could have slightly accelerated the queue's dissipation: Fig. 2(a) shows that the queue was already well on its way to dissipating and apparently would have done so at around 19:00 hrs -- even if the lane-use restriction had not expired.

We examined this freeway for nine additional weekdays (in July and August, 2004). On four of these days, a queue did not arise at all. On each of the five other days when a queue did form, it did so locally at the merge bottleneck near PM 26.7. Figure 2(b) is an occupancy plot for one of these five days. (While the bottleneck was active this day, flow in the HOV lane never reached 1,500 vph.)

Summary: Although it may be tempting to blame the HOV lane for the bottleneck at this site since the lane was under-utilized, this is only a possibility. The data needed to determine how the bottleneck would have performed without the HOV lane are not available for this site. Yet, even if this lane is the culprit, the recurrent damage it creates is much less than the non-recurrent damage observed in EATO, much of which was caused by the traffic accident on that day.

3.2 The case of SR-101S: Fig. 1(d).

A look at Fig. 1(d) reveals that speeds were indeed low between 15:00 to 19:00 hrs. However, the chart also reveals that low speeds prevailed well before and after these times: speeds were already declining at 14:30 and did not fully recover until 20:00 hrs. One cannot conclude on the basis of this evidence that the HOV lane is causing the problem.

A spatiotemporal analysis was repeated for this site to unveil the causes. Figure 3 summarizes the result. It displays a 6-hr time-space-occupancy diagram for a 4-mile freeway stretch that includes the detector station of Fig. 1(d). The occupancies reveal that: (i) as in the

previous case, the queue first formed at a definite bottleneck location (a merge near PM 390) and not simultaneously everywhere; (ii) the detectors of Fig. 1(d) were close to the bottleneck; and (iii) high occupancies were reached at this location by 14:45. (Occupancy started to increase earlier, in conjunction with the speed changes of Fig. 1(d), but the color resolution of Fig. 3 is insufficient to reveal these initial changes.) With this timeline in mind, and assuming that the detectors of Fig. 1(d) are just upstream of the on-ramp merge,⁵ the sequence of events in Fig. 1(d) is now explained.

The initial speed reductions (beginning by 14:30) are most likely due to insufficient freeway capacity to accommodate the growing demand from both the freeway and the on-ramp; this is common for merge bottlenecks. (Confirmation: upstream freeway demand increased -- by 36% -- to typical capacity levels at this time.)

The sharper speed reduction at 14:45 occurred too early for HOVs to have had much of an impact. Therefore, its most likely cause is the shockwave created by a “capacity drop”.⁶ (Confirmation: flow diminished by 7%, as is typical for capacity drops, and did so only in the right-hand-lanes; flows in the median lane did not change.)

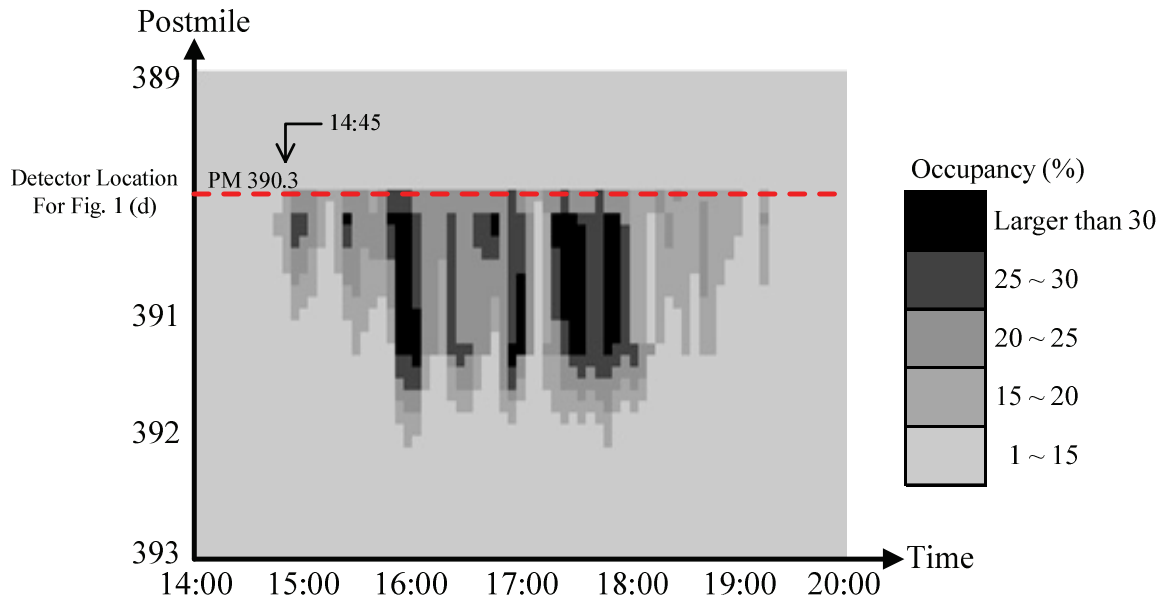


Fig. 3. Time-Space-Occupancy Plot, SR-101S (Aug 18, 2004)

The temporary but sharp rise in speed just before 15:00 was accompanied by a small increase in total flow (160 vph) despite a large reduction in the HOV-lane flow (630 vph), as LOVs vacated in response to the lane-use restriction. Thus, the vehicle migration out of the HOV

lane may have increased the bottleneck discharge rate. (This is not completely surprising: a similar “smoothing” effect has been observed in simulations; see *Menendez and Daganzo, 2006*.) In the present case, the rise in flow could also have been caused by a flow reduction from the downstream on-ramp. This is unlikely, however, given that the rush had begun. The smoothing effect is the more likely cause. We say this, in part, because at the one site studied with sufficient resolution to detect the effect, it did arise; see Sec 3.3.

The fluctuations in speed (Fig. 1(d)) and in occupancy (Fig. 3) that occurred throughout the rush are probably due to changes in the on-ramp’s flow. Fluctuations occurring late in the rush (near 19:00) are of little consequence since the queue had nearly vanished by then.

Summary: There is no evidence to suggest that the HOV lane on this site contributed significantly to the observed congestion. To the contrary, the lane may have reduced congestion slightly by increasing discharge flow in the bottleneck’s GP lanes.

3.3 The case of I-880N: Fig. 1(c).

We examine the next site with the aid of Fig. 4(a), a time-space-occupancy plot for a long stretch of freeway containing the detector station used in Fig. 1(c) and spanning the same time period. Once again, we see from the occupancies that a long queue did not form all at once along the freeway. Rather, it started at a definite point in time and space (at 15:00 and PM 26).⁷ The spatial location is not surprising because there is a merge near PM 26. The time of queue formation is quite suggestive because it coincided with the time when the HOV restriction took effect. This synchronous pattern could indicate that the HOV lane was triggering the queue. We therefore took and analyzed video data of the site to unveil the mechanism. Section 4 describes the results. To our surprise, we found that the HOV lane was not impairing bottleneck flow, but appeared instead to be enhancing it.

3.4 The case of SR-237E: Fig. 1(b).

Figure 5 is a time-space-occupancy plot for a stretch of SR-237E encompassing the data of Fig. 1(b). The occupancies reveal some light queuing. The pattern is strange (and interesting) because the initial congestion boundary does not propagate upstream as in the other cases. Could this be related to the usual workings of the HOV lane? The answer is no. On this day, the HOV lane was closed downstream near PM 9 due to construction work (the event is documented in *Caltrans, 2004*) and non-recurrent signage could have affected drivers along this site. Since queues were not detected at the site on incident-free days, there is no basis for concluding that the HOV lane has a negative effect on this freeway stretch.

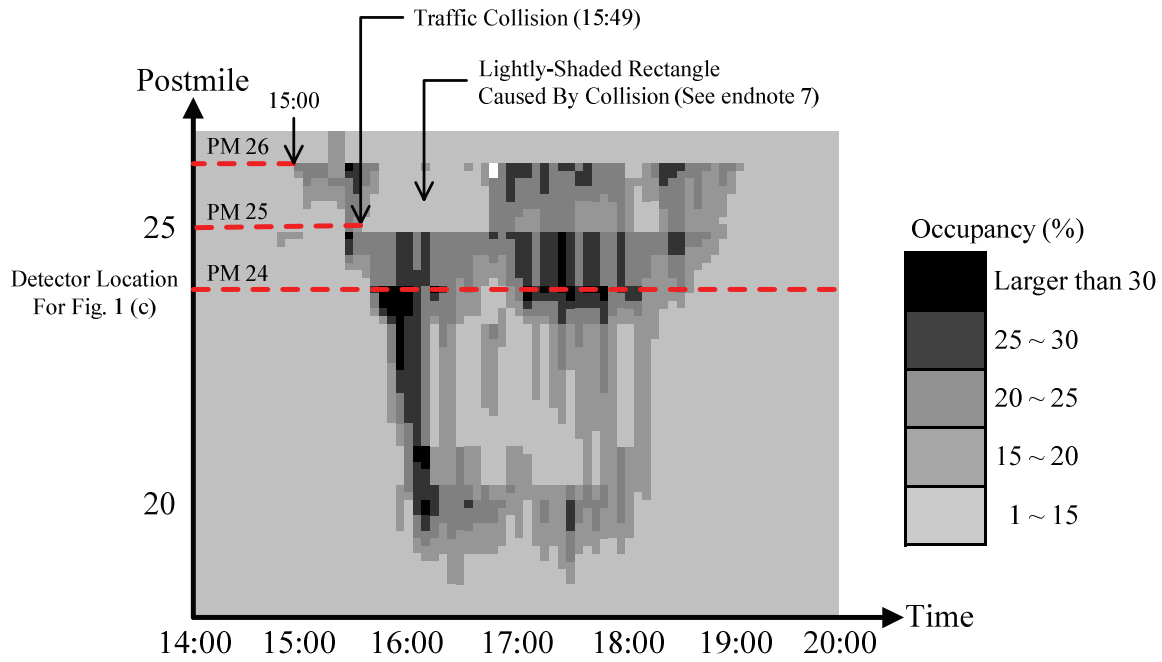


Fig. 4. Time-Space-Occupancy Plot, I-880N (Aug 23, 2004)

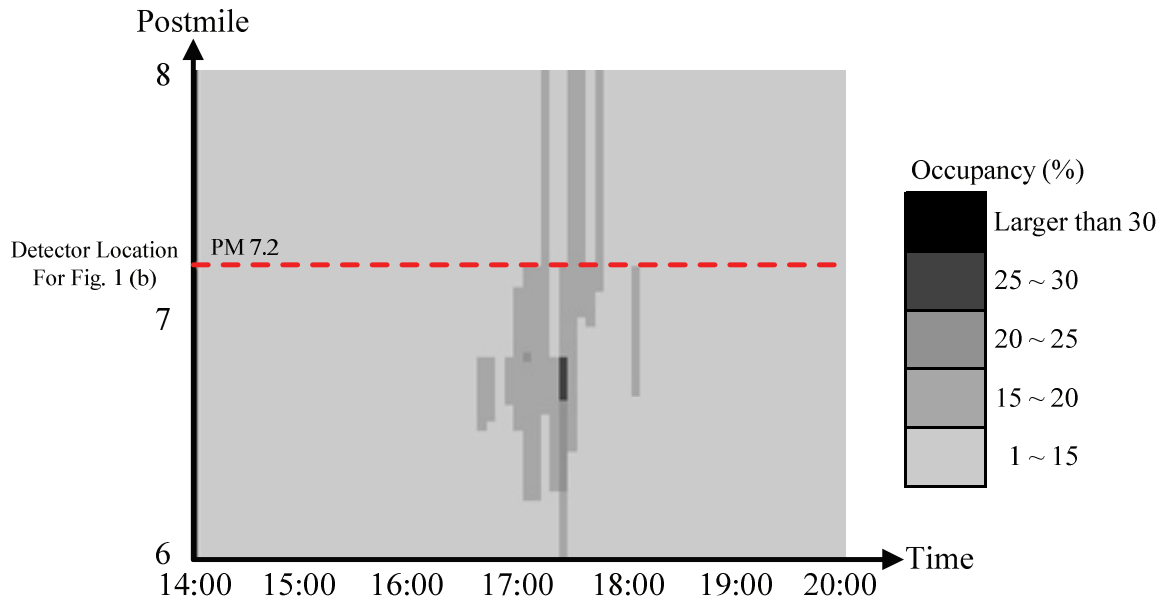


Fig. 5. Time-Space-Occupancy Plot, SR-237E (Aug 20, 2004)

3.5 The case of I-80E: Fig. 1 (a)

Finally, we have attempted to reassess the traffic conditions of Fig. 1(a). These attempts were unsuccessful, in part, because we are unsure of the day from which these speeds actually came. The observation day was reported in EATO to be August 8, 2004. Yet, the detectors on this day (a Sunday) were not functioning.⁸ As an alternative, we present occupancies for this freeway stretch, but for another day (August 18, 2004) whose data resemble those in Fig. 1(a).

Figure 6 shows the resulting time-space-occupancy plot.⁹ A queue first formed locally (at a diverge) near PM 11 and soon thereafter a short distance downstream (at a merge) near PM 11.5. The queue became denser as it propagated past a busy on-ramp near PM 10. These events took place well before 15:00 hrs, suggesting that queues were not formed by the HOV lane. The queue's dissipation, in this case, appears to have coincided with the HOV lane's expiration time.

Summary: Perhaps, in this case there could be some connection between the HOV lane and increased congestion. Although the HOV lane was not triggering the queues, perhaps it was exacerbating them by reducing bottleneck discharge flows. Unfortunately, we cannot verify this conjecture, reasonable as it may seem, because the detectors were unreliable. For example, during the 15 minutes immediately prior to queue formation, flow at PM 9.9 was measured to be well below 4,000 vph; i.e., less than 800 vph per lane. Such low flow is unlikely during the build-up to a rush.

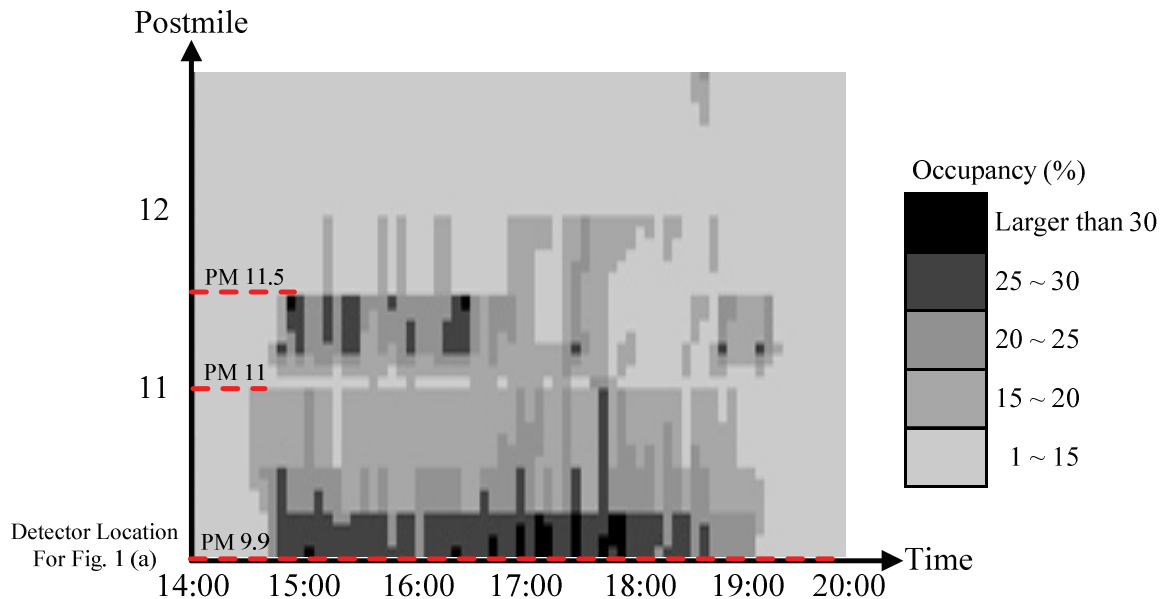


Fig. 6. Time-Space-Occupancy Plot, I-80E (Aug 18, 2004)

4. DETAILED OBSERVATIONS AT A BOTTLENECK

This section examines the bottleneck of Sec. 3.3 in more detail. Readers interested only in its conclusions can skip to its last paragraph.

Figure 7(a) displays the freeway geometry in the bottleneck's vicinity. Video cameras were erected on the pedestrian over-crossing and these recorded traffic at locations X_1 , X_2 and X_3 during part of an afternoon rush (on July 19, 2006). Vehicle arrival times were manually extracted from the videos and, as is customary, cumulative curves of vehicle count were plotted on an oblique coordinate system (O-curves).¹⁰ The curves are shown in Fig. 7(b). Note that curves 2 and 3 are superimposed, and below curve 1. Thus, traffic was in free-flow between X_2 and X_3 , but delays existed between X_1 and X_2 . These two curves diverged for good at about 14:43 hrs when a disruption reduced the flow at X_2 . Less than 3 minutes later (at approximately 14:45:30) flow dropped further to about 6,960 vph. These events, including the "capacity drop" that occurred around 14:45:30, are typical of merge bottlenecks and do not necessarily imply any involvement of the HOV lane.

In fact, the video data unambiguously establish that the HOV lane did not trigger this bottleneck by pushing LOVs into the GP lanes. Figure 7(c), which displays the net inflow of vehicles into the median lane between locations X_1 and X_3 , shows that the inflow was positive when the queue was forming. Thus, the HOV lane was relieving pressure from the GP lanes, rather than "pushing" traffic into them. Conceivably, this HOV activity could also create a bottleneck, but further investigation reveals that this was not the case.

Lane-specific data reveal that the queue started in the shoulder lane, triggered by merging activity. Figure 7(d) displays two O-curves for location X_2 : one for the shoulder lane (in boldface) and another for the remaining lanes. Note how the flow in the shoulder lane suddenly diminishes (from 1,960 vph to 1,760 vph) at 14:43 hrs (the time when queuing began in Fig. 7(b)) without any effect on the remaining lanes.¹¹

Note too that the capacity drop that occurred around 14:45:30 coincided with a reduction in the flow of the adjacent lanes (light curve in Fig. 7(d)) signifying that the queue had by then spread across the whole width of the freeway. This, again, is typical of merge bottlenecks without HOV lanes. (For a detailed account of the onset of congestion at merge bottlenecks see *Cassidy and Rudjanakanoknad, 2005*.)

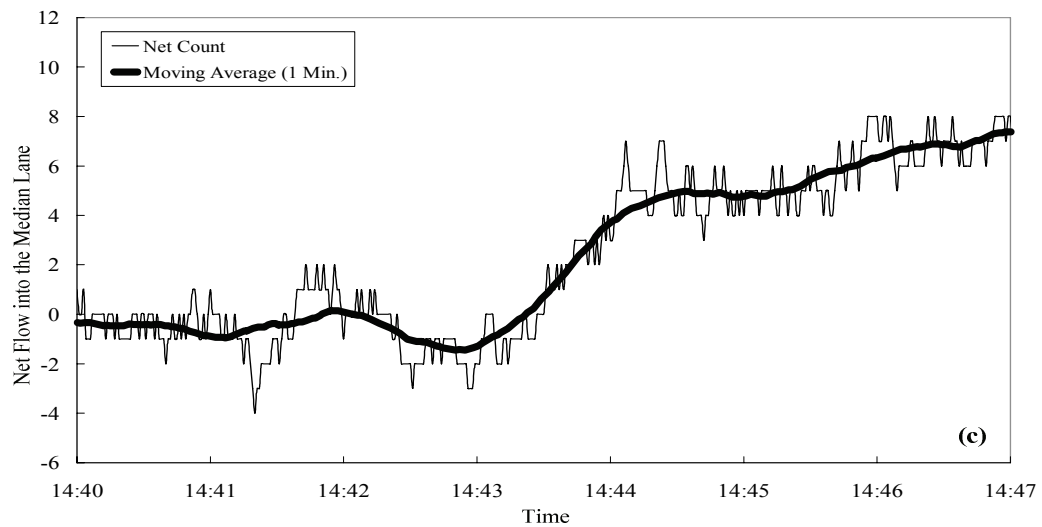
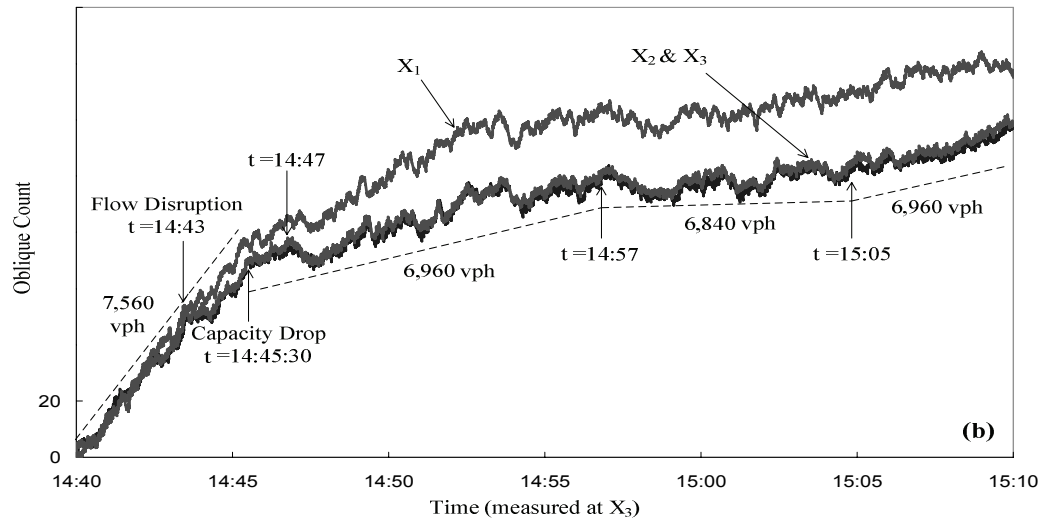
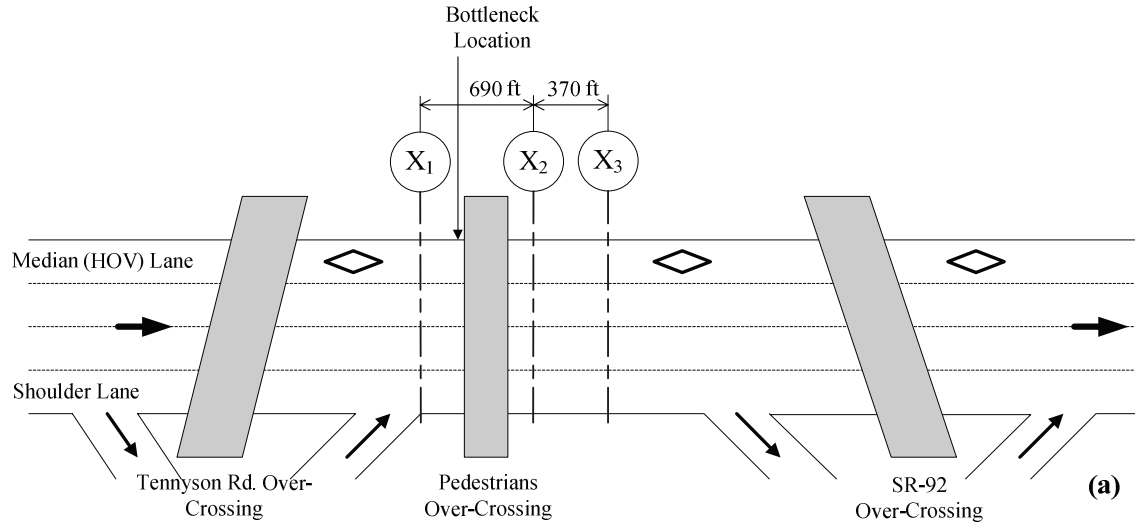


Fig. 7. July 19, 2006: (a) Study Site, Northbound Freeway 880, Alameda, California, (b) *O*-curves at X₁ through X₃, (c) Net Flow into the Median Lane between X₁ and X₃

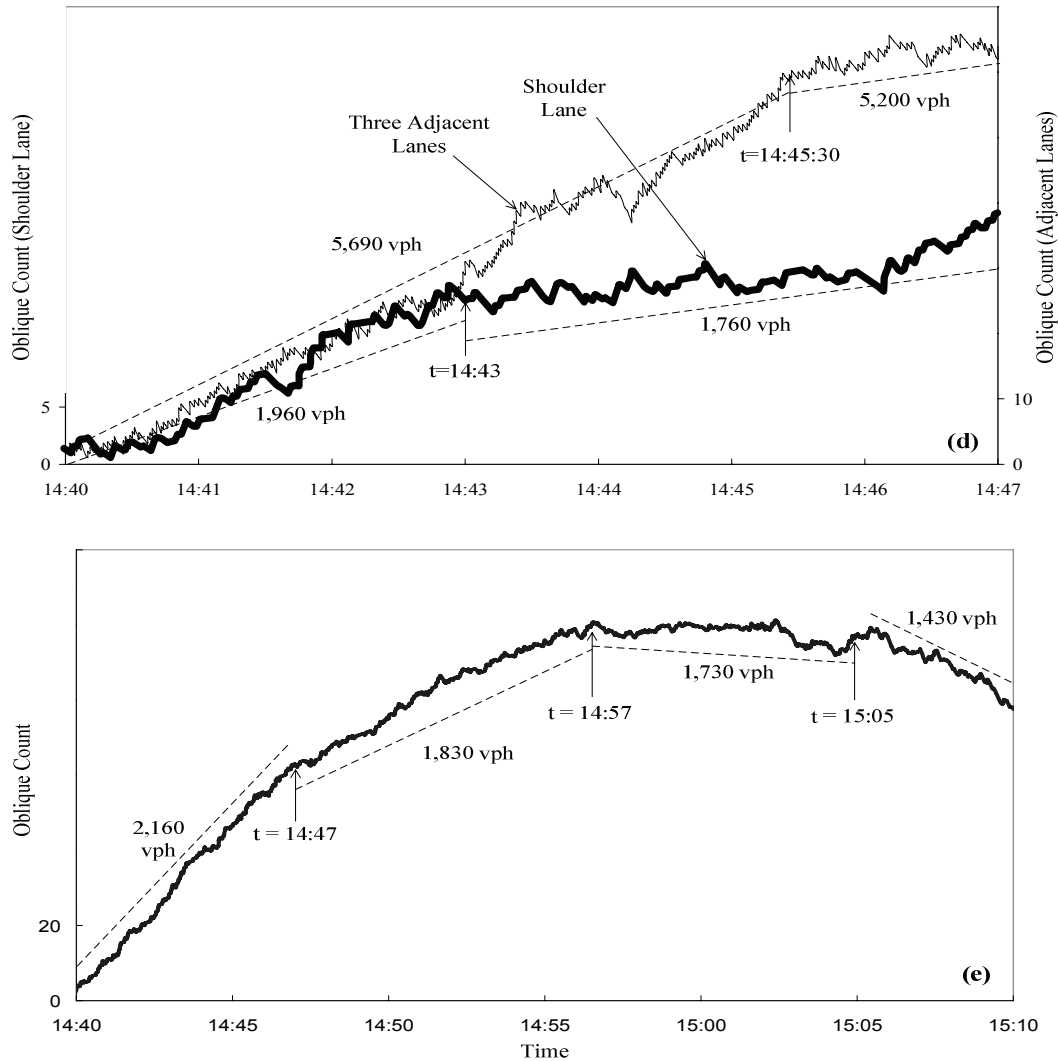


Fig. 7.(continued) July 19, 2006: (d) Oblique Count Curves of Shoulder and Adjacent Lanes at X_2 , (e) Oblique Count Curve of HOV lane at X_3

Having established that the HOV lane did not trigger the queue formation, we next explore whether the lane caused the capacity drop. Recall that capacity-drops are typically in the range of 5 to 10%; see endnote 6. We note from Fig. 7(b) that if ours was a typical bottleneck, its flow would drop (by 7.5%) from 7,560 vph to 6,990 vph, which compares well with the observed discharge flows (6,960, 6,840 and 6,960 vph) from time 14:45:30. So, it seems that the existence of the HOV lane is not causing a flow drop any greater than drops typical of freeways without HOV lanes.

To shed further light on this issue, consider Fig. 7(e), which displays an O-curve at X_3 for the HOV lane. Note how flows on this lane remained high until 14:47, well after the capacity had dropped at around 14:45:30 hrs, and how they gradually diminished thereafter. The time of the initial flow reduction indicates that the migration of LOVs out of the HOV lane did not start before 14:47 and had nothing to do with the capacity drop.

Furthermore, a comparison of Figs. 7(e) and (b) from 14:45:30 onward reveals that the migration, while strong, did not reduce the bottleneck discharge rate at all. This was a surprise,¹² but the fact is incontrovertible. Note how the flow in the HOV lane (Fig. 7(e)) drops from an initial rate of 2,160 vph to about 1,430 vph just after 15:05 (indicating that the HOV lane became underutilized by about 33%). Yet, the total flow across all lanes (including the HOV lane) shown in Fig. 7(b) remained quite steady at about 6,900 vph. Interestingly, the queue discharge flow returned to its highest rate (6,960 vph) during the last 5 minutes when the HOV-lane flow was lowest (1,430 vph), demonstrating that the smoothing effect arose.

Thus, despite its underutilization, the HOV lane at this site has no adverse effects on the bottleneck. Queues were triggered close to 15:00 hrs only because this is when demand exceeded capacity. Coincidences of this type, both at 15:00 and 19:00 hrs have recurred in Sec. 3. One can argue that speeds in the charts of Fig. 1 are generally lower in the interval between 15:00 and 19:00 not because HOV lanes are causing the reductions, as claimed in EATO, but because this is when people want to travel in large numbers. To Caltrans' credit, HOV-lane operating times seem to have been scheduled to coincide with the rush, albeit not perfectly since the rush varies from site to site.

5. CONCLUSIONS AND RECOMMENDATIONS

The EATO study asserts that HOV lane installations at (five) sites in the San Francisco Bay Area contribute to congestion; and concludes (with some caveats) that HOV lanes should be eliminated. Our analysis could not duplicate the empirical findings in EATO. We found a more positive picture. Specifically, we found: one site (Sec. 3.1) where the HOV lane could be creating a small bottleneck on some days (though the evidence was inconclusive and the bottleneck sporadic); two sites (Secs 3.2 and 3.3) where bottlenecks would have existed even without the HOV lanes, and where these lanes seemed to be having no negative effect on the bottlenecks; one site (Sec. 3.4) where the congestion was caused by downstream construction and where no delays recurred on days without construction; and one last site (Sec. 3.5) where one could surmise through the noise

of ill-functioning detectors that the HOV lane did not trigger the queue, although it *could* have reduced the bottleneck discharge rate.

In short, we were unable to place unambiguous blame for congestion on the HOV lanes at any of the EATO sites. Two (out of the five) sites could be adversely affected by their HOV lanes, but the effects (if any) were too small to be confirmed. We believe that unfavorable effects can and do arise, but see no basis for the large damages attributed to HOV lanes in the EATO study; and no justification for their blanket removal. Current understanding of HOV-lane physics suggests instead that modest changes in policy at problematic sites would do a better job for society than would removal, especially since the latter would be expensive and could be counterproductive. Recommendations are given below. We begin by discussing what ought to be an objective when deploying HOV lanes.

5.1. A Goal

For a specified set of commuter trips (input), an important measure of a freeway's performance is the number of total hours that vehicles spend traveling on it, on its on-ramps and on the surface streets that access these ramps (the VHT). This metric is readily converted into total vehicle delay; and it correlates well with commuters' monetary costs and externalities such as emissions, fossil fuel depletion and noise.

A freeway is like any input-output system in that, for a given set of input demands, the total time spent in it and accessing it (including the time spent on on-ramps and access streets) depends only on the output flows (at the downstream end of the freeway and from all its off-ramps); see for example *Newell (1982)*, *Lin and Cao (1997)* and *Cassidy (2003)*. This means that an HOV lane will increase VHT only if its deployment causes freeway output flow to diminish. Since typically a freeway's output flows are largely dictated by the discharge flow(s) through its bottleneck(s), the goal should be to install HOV lanes without significantly reducing bottleneck flows.¹³

If this is achieved, and the HOV lane allows HOVs to reach the bottleneck with less delay, then VHT remains nearly invariant and most of it is allocated to LOVs. As a result, the HOV lane decreases people-hours of travel (PHT) without significantly changing VHT and its concomitant externalities. Note that the relevant metric to assess discrimination favorable to HOVs is not whether HOVs get to travel close to their free-flow speed (a metric used in EATO) but whether HOVs travel appreciably faster than LOVs. By this criterion, the HOV lanes of Fig. 1 are discriminating splendidly in favor of HOVs.¹⁴

5.2 Policy Recommendations and Caveats

In light of the above, an HOV lane should not be installed at a location where it would trigger a bottleneck. Nor should an HOV lane run through a bottleneck if the lane is so severely underutilized that it would reduce bottleneck discharge flow. But an HOV lane that violates these conditions does not have to be removed entirely. Instead, one should lift the lane-use restriction near the bottleneck (or near the location that would otherwise become a bottleneck) and preserve the HOV lane upstream, all along the queue. In this way, HOVs reach the bottleneck sooner if/when a bottleneck does materialize.

Simulations confirm that terminating HOV lanes upstream of bottlenecks does not diminish the bottlenecks' discharge flows (see again *Menendez and Daganzo, 2006*). These simulations (and the results of Sec. 4) also suggest that if an HOV lane is only slightly underutilized and is allowed to run through a bottleneck, the lane can *increase* the total bottleneck flow by smoothing it. A run-through HOV lane with this desirable property would reduce PHT, and possibly even VHT and the externalities. It could be a win-win proposition for society.

We also stress that if an HOV lane is terminated upstream of a bottleneck, so that bottleneck flow does not change, then the underutilization of the HOV lane does not affect the VHT of vehicles that travel through the bottleneck. Curiously, underutilization actually benefits the LOVs traversing the bottleneck (whose drivers are often vocal critics) because it implies that fewer HOVs are availing themselves of the opportunity to bypass the LOV queue and cut in ahead of those waiting.

The only downside of underutilization is that it slightly wastes freeway storage space: our preliminary analyses (to be completed in a forthcoming paper) show that severely underutilized HOV lanes on 4- or 5-lane freeways extend queue lengths by about 10-20%; and well utilized HOV lanes by 0%. This lengthening phenomenon has consequences only when the short extra tail of the queue retards another major traffic stream; e.g., by spilling back to another congested freeway through an interchange, or by blocking a busy off-ramp. But this can be easily checked in the field. Barring this, slightly longer queues contribute little to vehicle delay.¹⁵

There are situations where, after treating all bottlenecks along an HOV lane as described above, congestion would disappear. In these cases (only) one could eliminate the complete HOV lane, although no additional benefits would come of this.

5.3 Action Items and Future Research

Although not clearly observed at the EATO sites, HOV lanes do, in some cases, trigger problems. *Menendez and Daganzo (2006)* identify a bottleneck on a Bay Area freeway where this seemed to

occur. Therefore, efforts should go toward diagnosing problematic installations and implementing suitable corrections. Field experiments are also needed to solidify our understanding of how HOV lanes impact bottlenecks. If institutional hurdles can be overcome, we should also field test less conventional dynamic operating policies, such as lifting the HOV restrictions near bottlenecks at certain times only, as proposed in *Daganzo et al. (2002)*.

The simulations in *Menendez and Daganzo* suggest that the above-cited policy can *increase* a bottleneck's capacity above what it would otherwise have been without an HOV lane; and that this can be achieved *independent of the HOV lane utilization level*.¹⁶ Although the field observations in Sec. 4 lend support to this theory, experiments are needed to improve our understanding of dynamic control strategies, as well as conventional ones. These experiments would entail removing HOV lanes at bottlenecks (permanently or dynamically) and measuring the resulting changes in discharge flow.

Ultimately, control strategies found to be effective could be implemented using technology to measure when conditions are ripe for enacting a strategy; and for informing commuters whenever the strategy is put into effect. Strategies could, of course, be implemented first in relatively simple static fashion, based upon historic knowledge of a freeway's operating conditions.

We believe there is also a need to change the general approach used to deploy HOV lanes. It appears to us that these lanes have been installed in a one-size-fits-all fashion (e.g., operated from 15:00 to 19:00 hrs at all the sites of Fig. 1) without due consideration to each freeway's bottlenecks. This seems no more appropriate than a policy that would have all of a city's traffic signals operate with the same signal settings. Clearly, just as signal settings should vary both across intersections and by time of day, HOV lanes should be deployed at different times in different places (perhaps dynamically in the near future), and with due consideration to the character of existing bottlenecks. The authors of this report would be pleased to work with Caltrans to help make this happen.

References

- Banks, J., 1991. Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering? *Transpn Res. Rec* 1320, 83-90.
- Caltrans. 2004. Report on Bay Area HOV lanes. California Dept of Transportation, District 4, Oakland, Office of Highway Operations.
- Cassidy, M.J. 2003. Freeway on-ramp metering, delay savings and the diverge bottleneck. *Transpn Res. Rec.* 1856, 1-5.

Cassidy, M.J. and Windover, J.R. 1995. Methodology for assessing dynamics of freeway traffic flow. *Transpn Res. Rec. 1484*, 73-79.

Cassidy, M.J. and Rudjanakanoknad, J. 2005. Increasing the capacity of a merge by metering its on-ramp. *Transpn Res Part B 39*, 896-913.

Chen, C, Varaiya, P. and Kwon, J. 2005. An empirical assessment of traffic operations. In: Mahmassani, H.S. editor, *Transpn and Traffic Theory*, Procs of 16th International Symp on Transpn and Traffic Theory.

Chung, K., Rudjanakanoknad, J. and Cassidy, M.J. 2007. Relation between traffic density and capacity drop at three freeway bottlenecks. *Transpn Res Part B 41*, 82-95.

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

Hall, F., Agyemang-Duah, K., 1991. Freeway capacity drop and the definition of capacity. *Transpn Res. Rec. 1320*, 91-98.

Lin, W.H. and Cao, J.M. 1997. A system optimal dynamic traffic assignment model with limited system control. *Presented at the 76th Annual Meeting of the Transpn Res. Board*, Washington, D.C.

Menendez, M. and Daganzo, C.F. 2006. Effects of HOV lanes on freeway bottleneck. *Submitted for publication*.

Munoz, J.C. and Daganzo, C.F. 2002. Fingerprinting traffic from static freeway sensors. *Cooperative Transpn Dynamics 1*, 1-11.

Newell, G.F 1982. Applications of Queueing Theory. Chapman and Hall, New York.

Varaiya, P. 2004. California's performance management system: Improving freeway efficiency through transportation intelligence. *TR News 218*, Jan-Feb.

APPENDIX: ADDITONAL COMMENTS

The body of this report has reassessed the empirical evidence in EATO. In this appendix, we comment on some statements in the EATO report which, in our opinion, unduly castigate HOV lanes. It will be useful to keep in mind the following principle, which follows from deterministic queueing theory:

VHT Principle: *Given a specified pattern of vehicle trips, a freeway's total VHT (including VHT on its on-ramps and surrounding access streets) cannot decrease unless freeway output flow increases. Therefore, anything done to increase freeway speeds without*

increasing freeway output flow can only add to total VHT: reductions in freeway VHT would cause equal or larger transfers of VHT to on-ramps and city streets.

The VHT principle explains why the focus of urban freeway evaluation should be on output flows and not exclusively on speeds.

A1. HOV-lane speeds: rewards versus penalties

The EATO report correctly notes that the HOV lanes exhibited speed reductions during the rush (see the bold, dark lines in Figs 1(a) – (f)). EATO gives possible reasons for the reductions and calls each reduction a penalty. It then states that increasing the utilization of an HOV lane would increase the penalty and that this is counter-productive.

We agree that there is probably a connection between increased HOV-lane utilization and diminished speeds in those lanes. Yet, even if speeds drop in HOV lanes, HOVs should still enjoy an advantage over LOVs – since HOV drivers have a choice of traveling in either set of lanes. The data of Fig. 1 corroborate this fact: despite slowing in the HOV lanes, vehicles in these lanes still move faster than those in the adjacent GP lanes. At worst, HOVs will travel at speeds equal to those of LOVs (see endnote 14).

Further, we disagree with EATO’s metric (diminished speeds) for assessing HOV-lane effectiveness (see VHT principle above). Since a properly functioning HOV lane should be reallocating delay without significantly changing freeway VHT, the relevant comparison should be between HOV and LOV travel times per unit distance traveled. Rather than characterizing diminished HOV-lane speeds as a penalty, one should focus on the difference between HOV and LOV trip times – a *reward*. This reward can be readily estimated from field data, such as those in Figs. 1(a) – (f). In nearly all these cases, the reward is quite significant; e.g., the HOV reward is 4.5 min for a 3-mile-long trip traveled at 40 mph, as compared with an LOV travel speed of 20 mph (as in the data of Figs. 1(e) and (f)).

A2. Impracticality of using P-metering in lieu of HOV lanes

The recommendation in EATO to remove HOV lanes was accompanied by a qualification concerning freeway management. It was claimed that HOV lanes should be removed, assuming that they could be replaced with “proper ramp metering” (henceforth called P-metering). This is said to be the kind of metering that guarantees GP lanes will carry flows that are close to the maximum observed vehicle flows (at maximum speeds). We agree that if ramp meters could guarantee near-maximal flows everywhere along a freeway, total VHT could be reduced (see

VHT principle). It would then be counter-productive to squander a lane to serve only (low) HOV demands, because output flows would then diminish. But as we now show, P-metering is generally infeasible.

Visualize the morning commute along a busy freeway with a fixed number of lanes, connecting the suburbs of an urban area to its downtown center or some other major destination. Suppose that during the morning rush, very few commuters exit prior to arriving at this destination. On such a freeway, the maximum observed flow at each location (before queues form) would be its capacity. Therefore, the P-metering strategy must allow enough input flows on the suburban on-ramps to guarantee that all freeway lanes flow close to capacity in suburbia. Since most of the suburbanites travel downtown, they would fill the downstream freeway segments. Under P-metering then, downstream on-ramps would have to be (nearly) closed. Otherwise, queues would form on the freeway, and flows within these queues would fall below the maximum observed level. Therefore, downstream on-ramps could only be opened after all upstream commuters had passed by.

As a real-world example, consider the stretch of westbound freeway I-80 shown in Fig. A1. It connects some of San Francisco's East Bay suburbs to the 580/880/80 freeway interchange. Since a majority of this freeway's morning commuters travel to the interchange, we infer that P-metering would require multi-hour closures of on-ramps in the cities of Berkeley and Emeryville. This unfavorable set up would be typical for radial freeways in many cities. In fact, we cannot think of any busy freeways where freeway congestion could be eliminated without creating long on-ramp queues that would spill-over to the city streets. This is a well-known drawback of metering with the goal of improving freeway traffic conditions only. And this is why we think P-metering is an unattainable idealization in most real-world situations.

A3. Accounting for people delay and commuter costs

In a discussion of mode choice, EATO includes calculations of person-flow for a range of vehicle occupancies (i.e. people per car) that might be induced by HOV lanes; and on calculations of commuter costs that arise when HOV restrictions are, and are not, in effect. The outcomes in EATO are not favorable to HOV lanes, but they are biased and exaggerated for three reasons: they assume that existing HOV lanes are causing problems when they are not (as was demonstrated in Secs. 3 and 4); they extrapolate conditions seen at a single point in space to an entire freeway link; and they ignore the VHT principle (and the potential impacts on on-ramps and city streets) by focusing on the freeway.

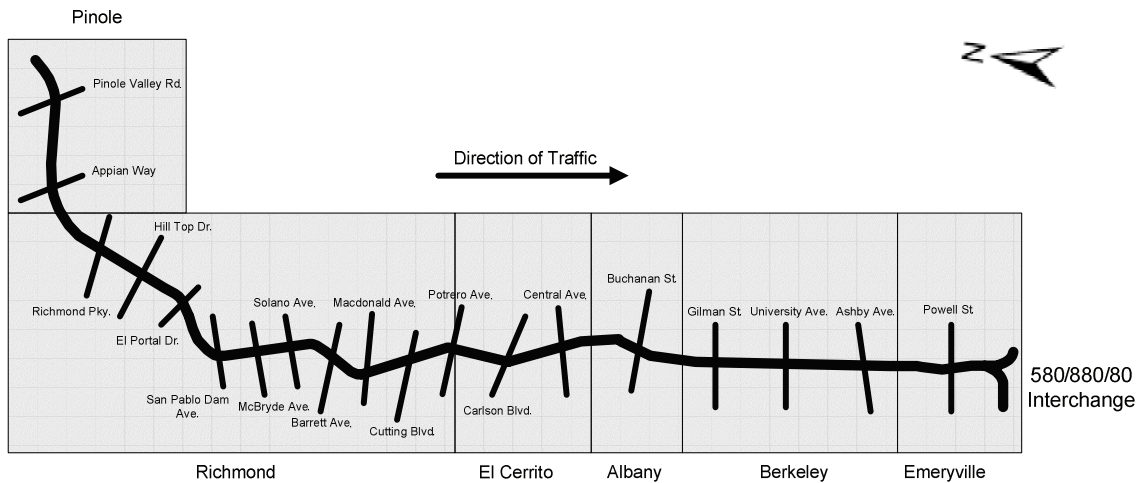


Fig. A1. Map of I-80 Westbound

ENDNOTES

¹ EATO qualifies this conclusion, and this is discussed in the appendix.

² Calculations of “person flow” and of commuter “costs” were also included in the EATO study. These too were based on measurements from isolated detector locations that were then extrapolated to describe a complete freeway link; for more detail, see the appendix.

³ The speed recovery in Fig 1(a) *might* be an exception but we cannot tell for sure; see Sec. 3.5.

⁴ Occupancies of about 20 percent or more denote the presence of queues. The exact occupancy threshold between queued and unqueued states seems to vary somewhat from one detector to the next, however, and the threshold likely depends on factors that include the detector’s general condition, how it was tuned, etc.

⁵ The precise location (upstream-downstream) is not noted in the PeMS database, but the customary location for Caltrans’ detectors is upstream of an on-ramp. Furthermore, if the detectors were some distance downstream, they would not have recorded the large speed reductions that were observed.

⁶ Capacity drops are reductions in bottleneck discharge flows when queues first form upstream. They are observed frequently at merges, and are typically in the range of 5%-10%; see e.g.: *Banks, 1991; Hall and Agyemang-Duah, 1991; Cassidy and Rudjanakanonand, 2005; Chung et al., 2007.*

⁷ The lightly shaded rectangle pinned at the bottleneck was the result of a collision that occurred at 15:49 at PM 25; see again <http://pems.eecs.berkeley.edu>.

⁸ On this day, the PeMS system (*Varaiya, 2002*) did not record data for the detectors at PM 9.9.

⁹ On this and other days, detectors upstream of PM 9.9 were reported (in PeMS) to be malfunctioning.

¹⁰ The important thing to know about O-curves is that superimposed curves indicate free-flow traffic and separated curves indicate delays; the wider the separation the longer the delays. The slopes of the O-curves

are the excess flow over a background flow, which is 6,800 vph in the present case. For additional information, see *Cassidy and Windover (1995)* and *Munoz and Daganzo (2002)*.

¹¹ One can see on the video that disruptions began in the shoulder lane when vehicles decelerated to make room for merging vehicles.

¹² It was not a total surprise because the fact had been predicted by the simulations of merge bottlenecks in *Menendez and Daganzo (2006)*.

¹³ There could be situations in which deployment of an HOV lane would significantly reduce off-ramp flows and would therefore be inappropriate. This is unlikely for queues that span only a few off-ramps, but could be true for heavily congested circular beltways. Work on these matters continues.

¹⁴ The only exception is Fig. 1(c), which shows that speeds in some GP lanes exceeded those in the HOV lane sporadically. This phenomenon, however, was not observed elsewhere in the queue. It was therefore a localized phenomenon and might have been due to the traffic collision on that day (see endnote 7).

¹⁵ Our upcoming paper also shows that unfavorable effects of an even smaller order can also arise if HOV lanes take effect on freeway stretches where queues had already formed on un-metered on-ramps. This kind of problem can be eliminated by slightly metering the ramps.

¹⁶ The simulation model was developed based on facts concerning driver behavior that were known as of 2005. The model successfully reproduced known features of bottleneck operation, including capacity drops. When HOV lanes were introduced, the simulation model also produced previously unobserved (and unexpected) predictions; the smoothing effect in particular. That the smoothing effect has now been observed in real traffic (see Sect. 4) gives us reason to expect that the benefits of dynamic control policies are also real.