Flow Processes at a Freeway Bottleneck

JAMES H. BANKS

Flow processes in the vicinity of a high-volume fixed bottleneck on a metered freeway were studied by comparing detector data with videotapes. Evidence is presented to support the hypothesis that maximum flow rates decrease when queues form. In addition, despite volumes significantly in excess of accepted values of capacity downstream of the on-ramp, the normal point of queue formation was about 1,500 ft upstream of the merge area. Queues appear to start in the leftmost lane, which carries extremely high volumes at this point. Shock waves resulting from obvious conflicts in the merge area are rare and normally occur only after the establishment of the queue upstream, despite 6-min merge rates of about 2,500 to 2,800 veh/hr.

Precise understanding of traffic flow processes in the vicinity of freeway bottlenecks is of obvious importance to the study of freeway capacity and traffic flow. Understanding the circumstances surrounding the initiation and growth of queues is particularly important. Potentially, freeway bottlenecks are of many different types, including upgrades, lane drops, merge areas, and weaving sections. How well these areas function depend on the details of vehicle spacing, lane use, and other features of driver behavior; these, in turn, may be affected by environmental conditions, traffic control measures (such as ramp metering), trip purpose, geographical area, and the like.

Despite this diversity and the importance of the subject, literature carefully documenting the flow processes leading to queue formation at freeway bottlenecks is sparse. A more common approach—that followed in the development of the weaving and merging chapters of the 1965 and 1985 *Highway Capacity Manuals* (HCMs) (*1,2*), for example—has been to attempt to generalize about traffic flow phenomena (*3–9*). Consequently, research has been based on rather small samples of data taken at many locations for which characteristics were not known (or at least not reported) in detail.

A few case studies of queue formation at bottleneck locations do exist. Persaud (*10*) described the initiation of queueing at a bottleneck in Toronto involving a lane drop. Newman (*11*) described the initiation of queueing at a merge location in Los Angeles. He found that, although the merge itself functioned at rates exceeding 2,000 veh/hr, queueing developed downstream as drivers increased their gaps following the merge. Edie and Foote (*12,13*) described flow processes and initiation of queueing at a tunnel in New York. Although a few other such studies may also exist, the full range of conditions leading to queue formation at freeway bottlenecks has not been documented.

A detailed case study of a high-volume bottleneck on a metered section of Interstate 8 in San Diego is described. The overall objective was to determine whether ramp metering can increase the capacity of freeway bottlenecks. The most direct way to answer this question would be to experiment with metering rates, including turning the meters off. As a practical matter, however, the operators of the system are reluctant to do this unless a clear case can be made that a particular action might increase capacity. Another approach is to document flow processes carefully in the vicinity of bottlenecks to determine the possible effects of metering. Such studies identify actions that could be taken on an experimental basis.

Past case studies of ramp metering have indicated that even crude metering systems can sometimes significantly reduce total travel time with no great decrease in total vehicle miles of travel (*14,15*). One possible explanation that is widely believed is that maximum flow rates at bottlenecks diminish when queues form, so metering increases the capacity of the bottlenecks by preventing queueing on the freeway.

This two-capacity hypothesis was discussed in the 1985 HCM (*2*) and the *Traffic Control Systems Handbook* (*16*). Athol and Bullen (*17*) discussed the possible implications of the hypothesis for metering strategy. The hypothesis is often related to two-regime or dual-mode traffic flow theories, which hold that relationships among speed, flow, and density are discontinuous at near-capacity flows (*17–20*). However, the idea that such discontinuities occur at bottlenecks has been challenged. Critics argue that most of the literature advocating dual-mode flow theories is vague as to where the data are to be collected, and that kinematic wave theory predicts flow drops upstream of the bottleneck when the shock wave at the upstream edge of a queue passes, even though flow at the bottleneck has not decreased (*21,22*). Edie (*18*) stated that the phenomenon had been observed at known bottlenecks but he did not give details.

Recent attempts to investigate the two-capacity hypothesis have been carried out by Hurdle and Datta (*23*) and Persaud (*10*). These studies observed flow at or just downstream from known bottlenecks. The proof of the hypothesis was taken to be a flow pattern in which the mean flow rate at the bottleneck drops abruptly just as the queue forms. Persaud (*10*) discussed the question of whether capacities should be stated in terms of maximum short-term counts or mean flows and investigated the effects of prequeue flow rates that increased with time. However, because of insufficient data, both of these studies proved inconclusive.

Although the two-capacity hypothesis is controversial, it could form a plausible basis for the hypothesis that ramp metering can increase bottleneck capacity. Consequently, the first stage of this research project is an attempt to test the two-capacity hypothesis. The results of the first of a series of case studies undertaken for that purpose are reported.

Civil Engineering Department, San Diego State University, San Diego, Calif. 92182.

STUDY SITE

The study site is located on westbound Interstate 8 on the eastern edge of San Diego. It includes the three westernmost metered interchanges in the morning-peak ramp-metering system on this freeway. Figure 1 shows the lane configuration and approximate distances between key features. Proceeding downstream from the Lake Murray Boulevard–70th Street interchange, Interstate 8 is relatively flat. Between the loop on-ramp from 70th Street and the College Avenue off-ramp, the westbound lanes consist of four main lanes and an auxiliary lane, which is trapped off at the College Avenue off-ramp. Just past the College Avenue off-ramp junction and about 1,500 ft upstream of the on-ramp, a horizontal curve of 2,000-ft radius overlaps an 800-ft vertical curve between a −0.7 and a −3.0 percent grade. Just beyond the College Avenue on-ramp is another vertical curve to a grade of approximately −3.5 percent. This curve extends to just upstream of the Waring Road off-ramp, 5,500 ft downstream of the College Avenue on-ramp. At the Waring Road on-ramp, an auxiliary lane is added, and the freeway transitions to a slight upgrade.

A previous study (*24*) had indicated a bottleneck somewhere in the vicinity of the College Avenue on-ramp. Average flow per lane just downstream of the on-ramp is maximum compared to other nearby locations. This flow is among the highest ever recorded. Sustained average flows of more than 2,400 vehicles per lane per hour across all lanes are common and 6-min flows exceeding 2,600 vehicles per lane per hour occur frequently. Furthermore, detector data produced by the ramp metering system showed that the lower speeds associated with queueing most often begin at College Avenue and move upstream from there.

Given these facts, a natural assumption is that the bottleneck results from flow processes at the ramp junction. Chapter 5 of the 1985 HCM (*2*) identifies two potentially critical checkpoints: the right lane at the merge point, and the entire freeway just downstream of the ramp. The logic for assuming that these points are critical is obvious. Volumes and densities in the right lane increase instantaneously at the merge point. Assuming that queues occur because flow exceeds some critical value (as both one-capacity and two-capacity hypotheses assume), the most likely location for the queue to form is in the right lane at the merge point. Possibly, however, drivers tolerate overcritical flows or densities momentarily, so the queue could form downstream of the merge point as drivers move out of the right lane or increase their gaps. This is the behavior that Newman reported (*11*).

Despite the plausibility of this scenario, previous study (*22*) indicated that it might be false. The downstream end of the queue at this location was observed to stabilize in the vicinity of the curve of 2,000-ft radius, which is about 1,500 ft upstream of the nose of the on-ramp. At the same time, occasional dense queues had been observed in the right lane, but without a permanent visual record, determining their cause or whether they initiated the more general queueing was impossible.

DEVELOPMENT OF HYPOTHESES

To test the two-capacity hypothesis, first contrast it with the hypothesis that capacity does not change when the queue forms. Under this hypothesis, each roadway segment has a capacity that is independent of the presence of queues, although that capacity may be assumed to be a random variable and to depend on other factors such as time of day or driver population. Whenever arrivals exceed the service rate (over any time period, however short) a queue forms, so the distribution of service rates functions as an upper bound for the observed distribution of bottleneck discharge rates. In contrast, the distribution of maximum nonqueue bottleneck discharge rates may not be the same as the service rate distribution—specifically, the discharge rate distribution may have a higher mean. Deciding between these hypotheses requires

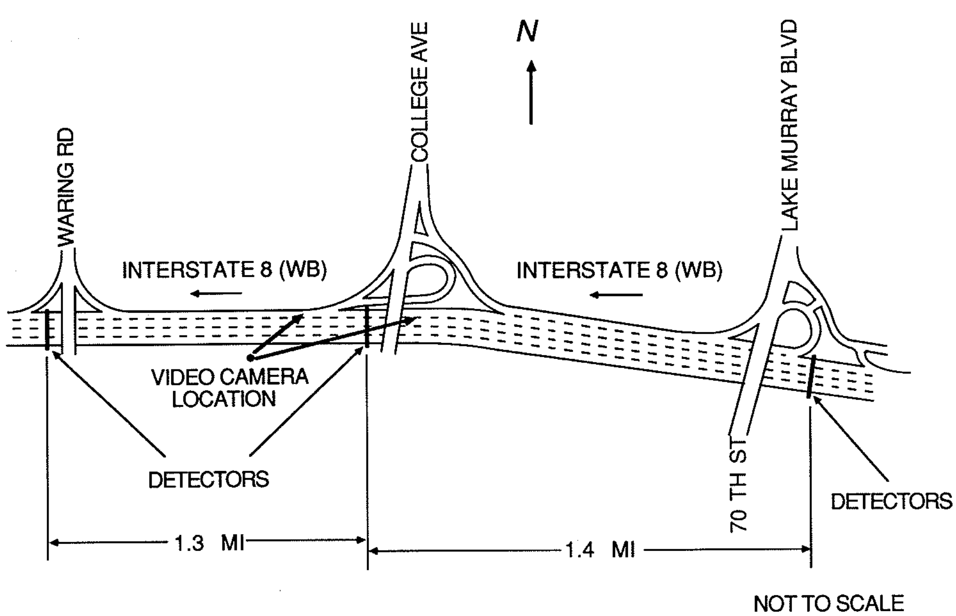

FIGURE 1 Schematic diagram of study site.

determining how their contrasting assumptions affect expectations about the distributions of bottleneck flow rates before and after queue formation.

Consider the following experiment. An observer is stationed at or just downstream from the bottleneck and is able to identify queues whenever they form. This observer measures flow over very short time intervals (say, 30 sec) until the first queue forms during a particular peak period. From these data, a distribution of prequeue flows through the bottleneck is computed. Subsequently, the observer measures queue discharge rates and computes the distribution of service rates. How do the two distributions compare?

First, consider the case of a metered freeway for which the meters are capable of holding the flow steady before the formation of the queue. If the alternative hypothesis is correct, this situation is typical stochastic queueing. If the mean flow rate produced by the meters is sufficiently close to the service rate, the arrival rate for some short time period eventually should exceed the service rate and a queue will form. Assume also that vehicle counts during successive time intervals are independent random variables and that the distribution of arrivals is independent of the distribution of service rates. Although the counts are actually discrete, consider continuous approximations of their distributions, which will simplify the derivation and should not make any practical difference in the conclusions.

Let $p_a(x)$ be the probability density function of arrivals and $p_s(x)$ be that of the service rate. Let $C_a(x)$ and $C_s(x)$ represent the corresponding cumulative distribution functions. Now, consider the probability $p(x > X)$ that a prequeue count exceeds X. For this to happen, both the arrival and service rates must exceed X. The probability that the arrival rate exceeds X is $1 - C_a(x)$, and the probability that the service rate exceeds X is $1 - C_s(x)$. If arrival and service rates are statistically independent,

$$\begin{aligned} p(x > X) &= [1 - C_a(x)][1 - C_s(x)] \\ &= 1 - C_a(x) - C_s(x) + C_a(x)C_s(x) \qquad (1) \end{aligned}$$

Further, the cumulative distribution function of prequeue flows $C_b(x)$ is $1 - p(x > X)$, so that

$$C_b(x) = C_a(x) + C_s(x) - C_a(x)C_s(x), \qquad (2)$$

and the probability density function of prequeue flows $p_b(x)$ is $dC_b(x)/dx$, or

$$p_b(x) = p_a(x) + p_s(x) - C_a(x)p_s(x) - C_s(x)p_a(x) \qquad (3)$$

Because both $C_a(x)$ and $C_s(x)$ are always less than one, $C_a(x)C_s(x)$ must always be less than both $C_a(x)$ and $C_s(x)$, and from Equation 2, $C_b(x)$ must always be greater than either $C_a(x)$ or $C_s(x)$. The consequence is that the mean of the prequeue flows should always be less than the mean of the service rate, and the probability of counts greater than the mean of the service rate should always be less before queue formation than afterward. Figure 2 shows the relationship among the arrival, service rate, and prequeue flow distributions.

Now, consider the case in which the prequeue flow is increasing with time, as would be characteristic of unmetered or inadequately metered facilities. In this case, the mean of counts taken just before queue formation should underestimate the mean flow rate at the instant of queue formation. Thus, mean prequeue flows are even less likely to exceed mean queue discharge flows, although, as Persaud (*10*) pointed out, a mean prequeue flow rate occasionally may be greater than the service rate.

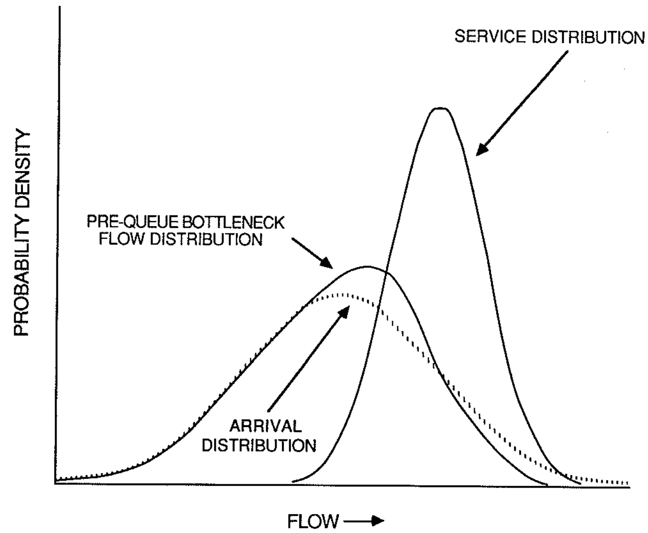

FIGURE 2 Expected prequeue bottleneck flow distribution.

Finally, consider what happens if queues form in individual lanes. Different lanes probably have different capacities; even if they do not, but arrivals vary by lane, queueing should begin in some particular lane, rather than in all lanes at once. Just after queue formation, the distribution of traffic across lanes should shift as drivers switch from the lane in which the queue has formed to other lanes. Thereafter, the queueing may also spread to the other lanes. Under the alternative hypothesis, the pattern of flows described would occur in the critical lane. In the other lanes, flows should exceed prequeue levels because of the change in lane use.

A finding that even one lane has a lower mean flow rate after the queue forms or fewer high counts after queue formation tends to support the two-capacity hypothesis. However, increases in flow in noncritical lanes or in total flow across all lanes do not negate the hypothesis because they may occur as a result of shifts in lane use. Indeed, increases in flow in noncritical lanes are to be expected in any case in which the mean prequeue arrival rate for the lane is less than its mean queue discharge rate.

METHODOLOGY

To test these hypotheses, observations of flow should be made at or just downstream from the bottleneck. Identification of the time when queueing first begins should be possible, and the specific external circumstances and traffic flow events associated with the formation of queues should be identified, if possible. For instance, being able to distinguish between the normal queueing associated with the site and that resulting from minor incidents is important. Being able to identify and exclude queues resulting from downstream conditions, whether they are routine or incident related, is also important.

Study procedures were designed to achieve these objectives. These procedures involved combining detailed analysis of detector data taken just upstream of the nose of the on-ramp with videotapes of the traffic flow. The detector data were used to calculate flow characteristics before and after queue formation; because the detector data include occupancies as well as flows, they could be used to calculate approximate average speeds and thus provide evidence of queue formation. The videotapes were used to identify conditions to be screened out in the data analysis, such as incidents and queues growing into the section from downstream, and to confirm the time of formation of queues. The videotapes were also used to provide supplementary information such as patterns of vehicle spacing before and after queue formation, counts of heavy vehicles, and verification of detector counts.

A high-quality, consumer-grade, 8-mm camera-recorder was used for videotaping. This unit featured a minimum light sensitivity of 4 lux and backlight features, which proved especially important because viewing conditions were less than ideal. Most of the time, the camera was set up on a canyon rim south of the freeway about 2,000 ft downstream from the ramp nose, facing east. This site gave a good view of the merge area but not of the curve upstream, which proved to be the critical area for queue formation. A number of other locations were tried, but none proved superior to this one. Lighting conditions varied from nearly dark, in which most vehicles had headlights on, to cases in which the rising sun was actually in the field of vision. Videotapes were of adequate quality in all cases except those in which the sun was actually in the field of vision.

Videotaping was conducted on weekday mornings beginning at the time the College Avenue meters came on (usually around 6:10 or 6:15 a.m.) and extending until just after 8:00 a.m. on most days. Videotaping began in June 1989 and was carried out in 23 days, resulting in about 37 hr of tape.

Detector data consisted of 30-sec volume and occupancy counts for each freeway lane and 30-sec on-ramp passage counts for the period 6:00 to 9:00 a.m. These data were available both for the College Avenue detectors and for those immediately upstream and downstream.

The primary data analysis technique was a generalization of the event-based averaging technique described by Allen et al. (*25*) and previously applied to the two-capacity hypothesis by Persaud (*10*). As in the previous section, the significant evidence is the mean flow rates before and after queue formation and the shapes of the corresponding bottleneck discharge flow distributions. The overall procedure began with identification of the approximate time that speeds dropped by plotting time series of calculated speeds aggregated across all lanes and over 6-min intervals. Next, time series of 30-sec speeds (both for individual lanes and aggregated across lanes) were plotted to determine the time of the speed drop more exactly.

From these plots, two 12-min periods were identified. The first was selected to be entirely before the formation of the queue and the second to include only queue discharge. In most cases, a short time interval occurred in between, a period in which speed was dropping rapidly but not instantaneously. Also, in some cases, apparent periods of intermittent queueing were avoided; also, isolated gaps in the detector data (presumably because of transmission failures) had to be avoided.

For these periods, linear regressions of flow versus time were performed to determine whether mean flow rates were varying with time, histograms of the distributions of 30-sec volume counts (both for individual freeway lanes and for the average count across all lanes) were prepared, and means and standard deviations of the counts were calculated and compared. Later, before-queue and after-queue data for all days included in the study were aggregated; histograms and frequency polygons were prepared for these composite distributions, and their means and standard deviations were calculated.

RESULTS

Of the 23 days for which taping was carried out, only 9 days proved suitable for analysis. Some days were excluded because no queues were observed; others were excluded because queues grew into the section from downstream before local queueing began, presumably because of incidents.

When queues formed in the immediate vicinity of College Avenue, they appeared to begin upstream of the on-ramp and the detectors, around the horizontal curve of 2,000-ft radius. Available camera locations did not permit a clear view of this area, but shock waves were sometimes observed in the vicinity, especially in the left lane. More conclusive evidence of queue formation in this area was a sudden drop in calculated speeds at the detectors just downstream, from more than 50 mph to about 35 to 45 mph. These speed drops were confirmed by timing selected vehicles on the tape. Also, a change in vehicle spacing and lane use that occurred at the same time as the speed drop was apparent both on the videotapes and in the detector data.

Flows at this location are somewhat less than those downstream of the ramp; average flow across all lanes just before queue formation is about 2,250 veh/hr, with approximately 950 veh/hr coming on at the ramp downstream. Percentages of heavy vehicles (trucks, buses, and larger recreational vehicles such as motorhomes) for selected 6-min intervals just before and after queue formation ranged from 2.6 to 5.4 percent; about 90 percent of the heavy vehicles are in the two rightmost lanes, with the highest percentage being found in the second lane from the right.

Merge conflicts were rarely observed. When conflicts did occur, it was always after the formation of the upstream queue or else in dense queues that had grown into the section from downstream; conflicts almost never produced shock waves that reached upstream even as far as the detectors. This absence was in spite of 6-min merge rates (as counted from the videotapes) of up to 2,500 veh/hr in free flow and up to 2,800 veh/hr in the discharge from upstream queues.

Dense platoons of vehicles were observed under all flow conditions. Those in the left lane before queue formation were especially large and dense, consisting in some cases of more than 30 vehicles in 30 sec at occupancies of 30 to 35 percent. Before queue formation, these platoons were separated by gaps of varying sizes. The gaps tended to be smallest in the left lane and largest in the right lane, reflecting considerable imbalance in the distribution of flow across the lanes. Once the queue formed, the gaps tended to disappear, and the flow took on a smoother appearance.

Regressions of time versus 30-sec counts averaged over four lanes were carried out for the 12-min periods identified as before and after queue formation to determine whether flow was changing with time. Table 1 presents the results. As can be seen, positive and negative regression slopes occur with about equal frequency before and after queue formation. Individual slopes were statistically significant at the 5 percent level (as determined by *t*-tests for significance of slope) in only two cases, both of which occurred after queue formation. This finding indicates that the meters are effective in holding the mean prequeue flow rate steady.

Comparisons of the before-queue and after-queue detector data showed that the distribution of flow across lanes became more uniform after queue formation, as expected. Table 2 presents these changes. In particular, the fraction of the traffic in the left lane decreased, as did the mean 30-sec count for the left lane, whereas the corresponding measures for the right lane increased. Those for the two middle lanes were almost unchanged.

Meanwhile, the variances of the time series of 30-sec counts tended to drop for each lane considered separately, as well as for the average count across all lanes. When compared by means of *F*-tests, variances decreased significantly at the 5 percent level (one-sided) in only about 35 percent of the cases observed, but overall, variances decreased about 80 percent of the time. To a large extent, the reduction in gaps between platoons, observed on the tape, may account for the decrease in variance. However, in the case of the left lane the maximum 30-sec counts also decreased, reflecting an increase in the average headway in the densest platoons.

Overall, comparisons of data from periods before and after queue formation tend to confirm the two-capacity hypothesis. According to the theoretical arguments, a drop in mean flow in even one lane would be evidence in favor of the hypothesis, as would distributional shapes in which the highest 30-sec counts were more frequent before queue formation than afterward. Both these conditions were met.

Table 3 presents mean 30-sec counts before and atter queue formation for the left lane, which appears to be the critical lane, as well as the average across all lanes. Because the variances were significantly different in a number of cases, means before and after queue formation were compared using a test for significance of difference between two means with different variances, rather than the *t*-test. The mean flow in the left lane was less after the queue formed in all nine cases, and the difference was significant at the 5 percent level in eight out of nine cases.

In eight out of nine cases, the average flow across all lanes also decreased when the queue formed. As was to be expected, the differences in before-queue and after-queue flow were less for all lanes than for the left lane; only three of the nine cases were significant at the 5 percent level. However, under the hypothesis that no change in flow is associated with formation of the queue, the probability that such a decrease will be observed in any given case is 0.50, and the number of cases in which flow drops should follow the binomial distribution for $p = 0.50$. Considered this way, the probability of observing flow drops in eight out of nine cases, given no real difference, is only about 2 percent. On the average, the decrease in mean flows across all lanes is about 0.60 vehicles per lane per 30 sec (70 to 75 vehicles per lane per hour) or about 3 percent of the mean prequeue flow.

Figure 3 shows frequency polygons of the distributions of left-lane counts before and after queue formation, as determined by combining data for all 9 days; Figure 4 shows similar frequency polygons for counts averaged across all lanes. Both

TABLE 1 REGRESSION SLOPES, 30-sec COUNT VERSUS TIME IN 30-sec INTERVALS, BEFORE AND AFTER QUEUE FORMATION

Date	Before		After	
	Slope	Significant?[a]	Slope	Significant?[a]
6-20-89	-.040	no	+.051	no
6-22-89	+.010	no	-.073	yes
6-26-89	+.015	no	+.004	no
6-30-89	+.045	no	-.089	no
7-13-89	-.048	no	+.014	no
7-14-89	-.021	no	-.026	no
7-18-89	+.006	no	-.093	yes
7-19-89	-.092	no	+.092	no
7-20-89	-.034	no	+.040	no

[a] Five per cent level of significance

TABLE 2 RATIOS OF INDIVIDUAL LANE VOLUMES TO AVERAGE VOLUMES ACROSS ALL LANES, BEFORE AND AFTER QUEUE FORMATION

Date	Condition	Lane (from Left)			
		1	2	3	4
6-20-89	Before	1.27	1.11	0.93	0.69
	After	1.13	1.02	0.98	0.88
6-22-89	Before	1.30	1.09	0.92	0.69
	After	1.20	1.07	0.98	0.75
6-26-89	Before	1.26	1.06	0.95	0.72
	After	1.16	1.06	0.90	0.89
6-30-89	Before	1.23	1.11	0.94	0.73
	After	1.17	1.08	0.94	0.81
7-13-89	Before	1.31	1.07	0.93	0.69
	After	1.21	1.09	0.92	0.78
7-14-89	Before	1.23	1.06	0.98	0.73
	After	1.18	1.03	0.96	0.82
7-18-89	Before	1.30	1.08	0.93	0.68
	After	1.18	1.10	0.91	0.80
7-19-89	Before	1.25	1.07	0.95	0.72
	After	1.16	1.07	0.93	0.84
7-20-89	Before	1.31	1.06	0.94	0.69
	After	1.21	1.07	0.94	0.77

figures indicate that the frequency of the highest counts is considerably more before queue formation than afterward, which is consistent with the two-capacity hypothesis.

QUESTIONS AND FUTURE DIRECTIONS

Bottleneck capacities decrease when queues form and queues near on-ramps sometimes form at locations not identified as critical by Chapter 5 of the HCM (*1,2*). These findings raise further questions.

On the theoretical side, an obvious question is why the formation of the queue should reduce the maximum flow rate. Previous attempts have been made to answer this question. A clear qualitative discussion that addresses most of the facts observable at this location was given by Edie and Foote (*13*). The answer appears to be related to certain time lags observable in traffic flow (a phenomenon sometimes known as hysteresis); car-following models (the literature for which is extensive) usually incorporate such lags. Treiterer and Myers (*26*) empirically studied the propagation of disturbances in vehicle platoons; however, in their study extensive queueing does not appear to have resulted. Construction and verification of a complete quantitative model explaining the flow decrease would obviously be an ambitious project.

More practical questions are whether metering can take advantage of the two-capacity phenomenon, and if so, how to identify optimal metering strategies. These questions hinge on the relationship (if any) between the prequeue flow and time of queue formation, and the magnitude of the flow decrease across all lanes. The magnitude of the flow decrease, in turn, depends in part on the degree to which prequeue flow is concentrated in particular lanes.

Athol and Bullen (*17*) have previously discussed the relationship between the two-capacity phenomenon and metering strategy. They assume that the probability of flow breakdown during some short time interval (in their case, 1 min) is a function of flow; then they consider the probability of breakdown over a number of successive time intervals. The result is that the expected time of flow breakdown (after the beginning of the peak) is a declining function of prequeue flow. They then select an optimum flow (to be produced by meter-

TABLE 3 COMPARISON OF MEANS OF 30-sec COUNTS BEFORE AND AFTER SPEED DROP

Date	Flow	Mean Before	Mean After	Change	Significant?[a]
6-20-89	Left Lane	24.75	20.71	-4.04	yes
	All Lanes	19.52	18.54	-0.98	yes
6-22-89	Left Lane	24.54	22.54	-2.00	yes
	All Lanes	18.83	18.74	-0.09	no
6-26-89	Left Lane	23.42	20.58	-2.84	yes
	All Lanes	18.54	17.80	-0.74	no
6-30-89	Left Lane	21.71	21.21	-0.50	no
	All Lanes	17.72	18.14	+0.42	no
7-13-89	Left Lane	25.13	21.92	-3.21	yes
	All Lanes	19.21	18.10	-1.11	yes
7-14-89	Left Lane	23.29	21.29	-2.00	yes
	All Lanes	18.96	18.05	-0.91	yes
7-18-89	Left Lane	24.71	21.63	-3.08	yes
	All Lanes	18.95	18.26	-0.69	no
7-19-89	Left Lane	23.38	21.17	-2.21	yes
	All Lanes	18.72	18.29	-0.43	no
7-20-89	Left Lane	25.63	22.71	-2.92	yes
	All Lanes	19.63	18.82	-0.81	no

[a] Five per cent level of significance

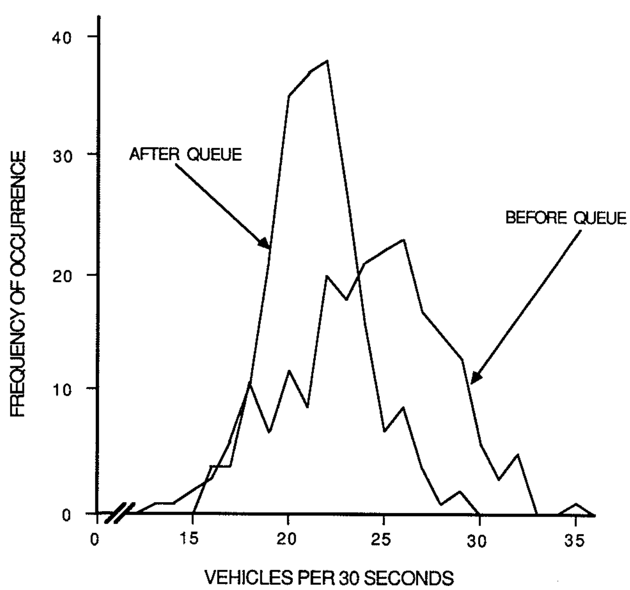

FIGURE 3 Frequency polygons, left-lane counts.

ing) to maximize the expected value of uncongested flow for the peak period.

An alternative objective would be to minimize total delay. Figure 5 shows delay for metered and unmetered conditions and could be extended to compare different metering rates. The cumulative demand function represents the cumulative flow that would arrive at the bottleneck in the absence of either queues or metering. Assuming that no traffic is diverted, this flow is the same regardless of the metering rate. Any shift in cumulative discharge to the right of this line represents delay, either in a main-line queue or in the queues at the ramp meters. Under the assumption that the discharge rate decreases when the queue is formed and that the time the queue takes to form is a function of the prequeue flow rate, limitation of the prequeue flow to the metering rate delays queue formation from t_1 to t_2. Area *A* in the figure represents delay experienced with metering, because the metering rate is initially less than the unmetered prequeue flow; Area *B* represents delay experienced without metering, because the queue discharge rate is less than the metering rate, and metering delays queue formation. Overall, metering decreases delay by no more than the difference between Areas *A* and *B*; if

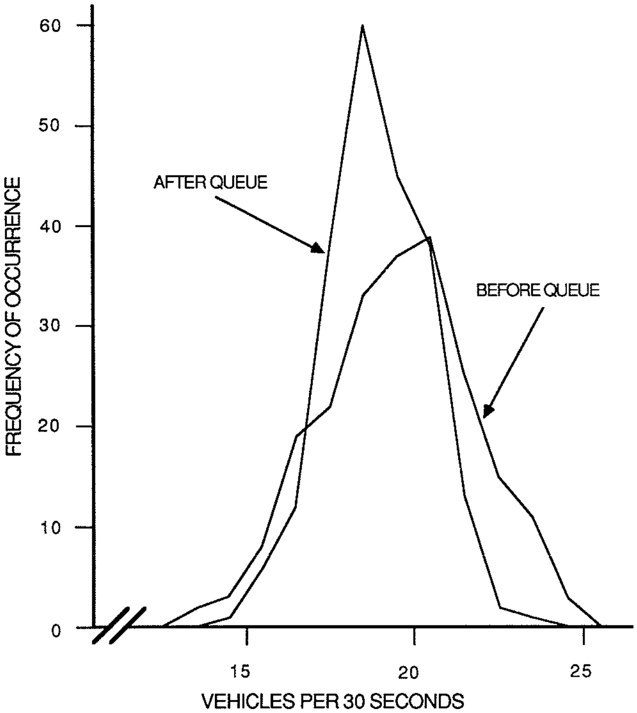

FIGURE 4 Frequency polygons, four-lane average counts.

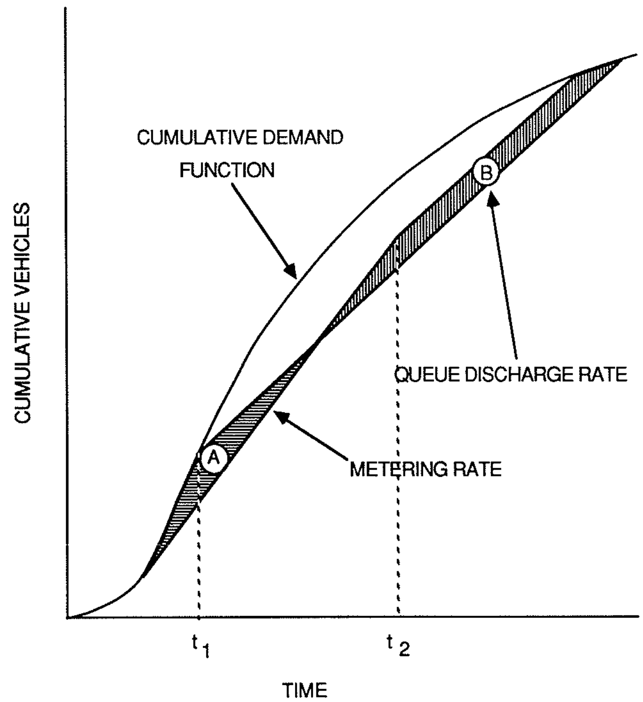

FIGURE 5 Comparison of delay with and without metering.

metering extends for several ramps upstream, it will also delay some vehicles not passing through the bottleneck, so the true difference in delay is actually less than the difference between Areas *A* and *B*. Clearly, the change in delay depends both on the decrease in flow when the queue forms and on the relationship between the prequeue flow and the time of queue formation.

Mean prequeue flow rates were plotted against time of queue formation to see if a pattern would emerge, but none did; furthermore, mean prequeue flow rates for days with queueing fall in the same range as those for which queues did not form. Clearly, a great deal more data are required to quantify the relationship. The necessary detector data are available, but verifying the causes of queueing would not be practical for the size of sample required. Nevertheless, an attempt at quantification might be worthwhile, even if some inappropriate data had to be included.

The question of lane use is of obvious importance, not only because lane use is related to the magnitude of the flow drop across all lanes when the queue forms but also because it determines where the queue forms. In this case, the queue formed where the prequeue flow was maximum—flows in the left lane upstream of the ramp exceeded those in the right lane at the merge point. The question is why drivers distribute themselves across the lanes in this manner. A number of possibilities exist. For example, drivers may avoid the right lane in anticipation of the high-volume merge, or they may perceive that speeds are higher in the left lane as long as no queue is present. Whatever the reason, the imbalance in lane use does not appear to be unique to this location; evidence of similar imbalances and similar shifts in lane use once queues form were found by Persaud (*10*) and Hall and Gunter (*27*).

A model relating use of each lane [not just the right lane, as studied by Hess (*4,5*)] to factors such as volume, ramp type and location, and congested or noncongested flow would be useful. Considerable empirical work will be required to produce such a model. The data base needed for at least a part of this work is available in San Diego, but these data would require extensive manipulation and analysis.

CONCLUSION

This study of flow processes in the vicinity of a high-volume freeway bottleneck, in which detailed detector data were analyzed and compared with videotapes of traffic flow, reveals that bottleneck capacities decrease when queues form and that queues near on-ramps sometimes form at locations not identified as critical by Chapter 5 of the HCM (*1,2*). These findings raise several questions, which include the theoretical question of why the flow drops when the queue forms and several empirical questions related to the issue of whether ramp metering can exploit the two-capacity phenomenon. Among the latter questions are how the time to queue formation varies with the prequeue flow rate and what factors influence the overall pattern of lane use.

ACKNOWLEDGMENTS

This research was funded by the State of California Department of Transportation. Special thanks are owed to Karel Shaffer.

REFERENCES

1. *Special Report 87: Highway Capacity Manual.* HRB, National Research Council, Washington, D.C., 1965.
2. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
3. K. Moskowitz and L. Newman. Notes on Freeway Capacity. In *Highway Research Record 27*, HRB, National Research Council, Washington, D.C., 1963, pp. 44–68.
4. J. W. Hess. Capacities and Characteristics of Ramp-Freeway Connections. In *Highway Research Record 27*, HRB, National Research Council, Washington, D.C., 1963, pp. 69–115.
5. J. W. Hess. Ramp-Freeway Terminal Operation as Related to Freeway Lane Volume Distribution and Adjacent Ramp Influence. In *Highway Research Record 99*, HRB, National Research Council, Washington, D.C., 1965, pp. 81–116.
6. R. P. Roess. Development of Modified Procedures for Analysis of Ramp Capacity. In *Transportation Research Record 772*, TRB, National Research Council, Washington, D.C., 1980, pp. 66–72.
7. O. K. Normann. Operation of Weaving Areas. *Bulletin 167*, HRB, National Research Council, Washington, D.C., 1957, pp. 38–41.
8. R. P. Roess, W. R. McShane, and L. J. Pignataro. Revision of NCHRP Methodology for Analysis of Weaving-Area Capacity. In *Transportation Research Record 772*, TRB, National Research Council, Washington, D.C., 1980, pp. 58–65.
9. R. P. Roess. Development of Weaving Area Analysis Procedures for the 1985 *Highway Capacity Manual.* In *Transportation Research Record 1112*, TRB, National Research Council, Washington, D.C., 1987, pp. 17–22.
10. B. N. Persaud. *Study of a Freeway Bottleneck to Explore Some Unresolved Traffic Flow Issues.* Ph.D. dissertation. Department of Civil Engineering, University of Toronto, Toronto, Canada, 1986.
11. L. Newman. Traffic Operation at Two Interchanges in California. In *Highway Research Record 27*, HRB, National Research Council, Washington, D.C., 1963, pp. 14–43.
12. L. C. Edie and R. S. Foote. Traffic Flow in Tunnels. *HRB Proc.*, Vol. 37, 1958, pp. 334–344.
13. L. C. Edie and R. S. Foote. Effect of Shock Waves on Tunnel Traffic Flow. *HRB Proc.*, Vol. 39, 1960, pp. 492–505.
14. C. Pinnell, D. R. Drew, W. R. McCasland, and J. A. Wattleworth. Evaluation of Entrance Ramp Control on a Six-Mile Freeway Section. In *Highway Research Record 157*, HRB, National Research Council, Washington, D.C., 1967, pp. 22–70.
15. L. Newman, A. M. Dunnet, and G. J. Meis. An Evaluation of Ramp Control on the Harbor Freeway in Los Angeles. In *Highway Research Record 303*, HRB, National Research Council, Washington, D.C., 1970, pp. 44–55.
16. *Traffic Control Systems Handbook.* FHWA, U.S. Department of Transportation, 1976.
17. P. J. Athol and A. G. R. Bullen. Multiple Ramp Control for a Freeway Bottleneck. In *Highway Research Record 456*, HRB, National Research Council, Washington, D.C., 1973, pp. 50–54.
18. L. C. Edie. Car Following and Steady-State Theory for Non-Congested Traffic. *Operations Research*, Vol. 9, 1961, pp. 61–76.
19. J. Drake, J. Schofer, and A. May, Jr. A Statistical Analysis of Speed Density Hypotheses. In *Highway Research Record 154*, HRB, National Research Council, Washington, D.C., 1967, pp. 53–87.
20. H. J. Payne. Discontinuity in Equilibrium Freeway Traffic Flow. In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 140–146.
21. F. L. Hall, B. L. Allen, and M. A. Gunter. Empirical Analysis of Freeway Flow-Density Relationships. *Transportation Research*, Vol. 20A, 1986, pp. 197–210.
22. J. H. Banks. Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations. Presented at 68th Annual Meeting of the Transportation Research Board, Washington, D.C., 1989.
23. V. F. Hurdle and P. K. Datta. Speeds and Flows on an Urban Freeway: Some Measurements and a Hypothesis. In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 127–137.
24. J. H. Banks. Performance Measurements for a Metered Freeway System. In *Transportation Research Record 1173*, TRB, National Research Council, Washington, D.C., 1988, pp. 30–32.
25. B. L. Allen, F. L. Hall, and M. A. Gunter. Another Look at Identifying Speed-Flow Relationships on Freeways. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C., 1985, pp. 54–64.
26. J. Treiterer and J. A. Myers. The Hysteresis Phenomena in Traffic Flow. In *Proc., 6th International Symposium on Transportation and Traffic Theory.* D. J. Buckley, ed., American Elsevier, New York, 1974, pp. 13–38.
27. F. L. Hall and M. A. Gunter. Further Analysis of the Flow-Concentration Relationship. In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986, pp. 1–9.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.