Freeway Capacity Drop and the Definition of Capacity

FRED L. HALL AND KWAKU AGYEMANG-DUAH

Two aspects of the definition of freeway capacity are considered here. The first is whether there is a reduction of maximum flow rates when a queue forms. There appears to be roughly a 6 percent reduction in maximum flow rates after the onset of congestion, but not of the type discussed in the current *Highway Capacity Manual* (HCM). An indirect issue arises during this analysis pertaining to the question of where capacity can properly be measured. The answer, not surprisingly, is that it can only be measured in a bottleneck, and not in a queue. However, the HCM discussion identifies the potential capacity drop on the basis of operations in a queue. The analysis contained here explains why that is inappropriate. The second aspect considered is the distribution over time of maximum flows at a single location. The distribution approximates a normal one reasonably well, with a mean of 6,071 and a standard deviation of 262 vehicles per hour. The unanswered question from the analysis is what portion, or percentile, of the distribution is appropriate to use to meet the HCM definition of capacity, which calls for a value that can "reasonably to expected" to be achieved.

Despite the existence of an explicit definition of capacity in the *Highway Capacity Manual* (HCM) (*1*) and a value for it of 2,000 passenger cars per hour per lane (pcphpl) (under ideal conditions), there remains some contention about the concept, not to mention about the numerical value for it. This paper focuses on two of the critical issues underlying the ongoing debate: the possible existence of a capacity drop after a queue has formed and the distribution of maximum flow rates at a single location over time. Both are addressed with new data. During the investigation of the possibility of a capacity drop, the question is also considered of where it is appropriate to seek such a drop. The answers found in this paper will not resolve the troublesome problems relating to capacity on uninterrupted-flow facilities, but may help to focus the debate.

In the first section of the paper, the background is provided, drawing largely on the HCM and related discussion. In the second section, the data that form the focus for the analysis are discussed. There then follow two sections of analysis, the first dealing with the capacity drop issue, the second with the distribution over time of capacity flows. The conclusions of the paper are presented in the final section.

BACKGROUND

There are three important elements to raise from the background material for a discussion of capacity, all of them arising from the treatment of the concept of capacity in the HCM. The first is the definition itself and two of the key ideas within it. The second is the numerical value given for capacity. The third is the discussion within the HCM that leads to the suggestion of a capacity reduction under congested conditions.

The 1985 HCM defines capacity as "the maximum hourly rate at which persons or vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period under prevailing roadway, traffic, and control conditions" (*1*, p. 1–3). The 1965 HCM definition differed in only a few respects: "the maximum number of vehicles which has a reasonable expectation of passing over a given section of a lane or a roadway in one direction during a given time period under prevailing roadway and traffic conditions" (*2*, p. 5). It went on to say, "In the absence of a time modifier, capacity is an hourly volume." Both definitions refer to maximum rate, reasonable expectation, and prevailing conditions. The major differences are the removal of the presumption of hourly volume and the addition of control to the list of conditions. The presumption of hourly volume was in fact explicitly altered in a subsequent chapter of the 1985 HCM in which freeway capacity is defined as "the maximum sustained (15-min) rate of flow" (*1*, p. 3–3).

Perhaps the most critical idea within these definitions is that of "reasonable expectation." McShane and Roess offer a useful clarification of this idea:

> A rate of flow that can be repeatedly achieved during every peak period for which sufficient demand exists and that can be achieved on any facility with similar characteristics anywhere within North America. It is not the absolute maximum rate of flow ever observed on such a facility. . . . The defined capacity of a facility is that maximum rate of flow which the traffic engineer may be reasonably assured of being able to achieve day in and day out anywhere in North America. (*3*, pp. 192–193)

Two points come from this discussion, the first very useful, the second arguable.

The useful point is that capacity is not the absolute maximum flow observed. Rather, it is a rate of flow that can be repeatedly achieved, day in and day out. This brings to mind the notion of a distribution of such flows. Capacity is not the extreme value of this distribution, but it is not clear what portion of the distribution should represent capacity. Elsewhere, the HCM states that the manual utilizes "national average" traffic characteristics, and that "the recommended value of 2000 pcphpl represents a national average" (p. 2–2). But if the distribution is normal, the average is not reached roughly half of the time. It will be valuable to look at the distribution, which will be done later in this paper.

Departments of Geography and Civil Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4K1.

The arguable point is McShane and Roess's assertion that the rate of flow must be achievable on any similar facility in North America. This would suggest that it is not useful to talk about the capacity of a specific facility, that is, of any particular highway. In contrast, we suggest that it is useful to distinguish between the capacity and a *type* of facility (e.g., a six-lane level freeway) and the capacity of a *specific* facility. In this paper, we intend to analyze data from one specific facility and to draw conclusions about that facility. Whether these conclusions apply to other similar facilities is always an open question, but we will argue that in several important respects, they do. Whether the numerical value of capacity applies to any similar facility in North America is a different question. Nevertheless, we suggest that it is more practical to operate this particular facility in light of its capacity than to use some continent-wide capacity value.

The numerical value of capacity has already been cited as 2,000 pcphpl. The 1965 HCM stated that the "largest number of vehicles that can pass a point . . . averages between 1,900 and 2,200 passenger vehicles per hour" (*2*, p. 75). Having stated this, it went on to say "the capacity of a multi-lane highway under ideal conditions is considered to be 2,000 passenger vehicles per lane per hour" (p. 76). The 1985 Manual, as quoted earlier, justifies the same number as being in some sense a "national average." Maintaining the same numerical value implies that there has been no net effect of the downsizing of vehicles over the last generation or of increasing driver experience with freeway driving. The numerical value is of particular interest at this time because there is a proposal to change the HCM capacity for multilane rural roads to 2,300 pcphpl (*4*), which would suggest that such a facility can handle more traffic than a freeway, an idea that is counterintuitive. Other recently published work (*5*) suggests sustainable freeway flows similar to the 2,300 value, albeit in only one location.

The final important element in the HCM regarding freeway capacity is the suggestion of a capacity reduction for congested operations. This is phrased as follows:

> Some researchers have fit continuous curves through density-flow data, yielding a single maximum flow rate. Others have projected discontinuous curves through data, with one curve treating stable flow points, and another unstable or forced flow points. In these cases two maxima are achieved, one for each curve. All such models indicate that the maximum flow rate for the stable flow curve is considerably higher than that for the unstable flow curve, perhaps as much as 200 vph higher. (*1*, p. 2–23)

This discussion is clearly based on density-flow behaviour *within* a queue, because that is the only place where "forced flow points" can be observed. It is our contention that this discussion, although based on a number of published studies [for example, those of Drake et al. (*6*) and Ceder and May (*1*)], is based on a mistaken premise about where the data are collected. Reduced capacity within a queue says nothing about capacity in the bottleneck downstream of the queue. The hypothesis to be tested in this paper is that there is a capacity drop in the bottleneck flow at the time a queue forms upstream.

DATA

Three issues about the data will be discussed: the location at which to collect data, the time intervals to use for analysis, and the estimation of truck percentages. The location selected is a part of the Freeway Traffic Management System (FTMS) on the Queen Elizabeth Way (QEW) in Mississauga, west of Toronto, Ontario, just downstream of the Cawthra Road interchange (Figure 1). The rationale for selecting this location was explained by Hall and Hall (*5*) following ideas by Hurdle and Datta (*8*). Cawthra Road is the final entrance ramp in a series and experiences daily congestion immediately upstream of it. Downstream of the ramp, there is a straight, level section of roadway for 2 km, after which there is a downhill section. There is an additional entrance ramp 40 m downstream of Station 25, but the road widens to four lanes with the addition of this ramp, so no queue occurs here. The presence of a queue upstream of Cawthra Road for an extended period confirms that there is sufficient demand for capacity operations in the downstream section. Indeed, the queue is present because there is more demand than the system can accommodate in this downstream section. This location in the bottleneck is clearly the only place to look for capacity operations.

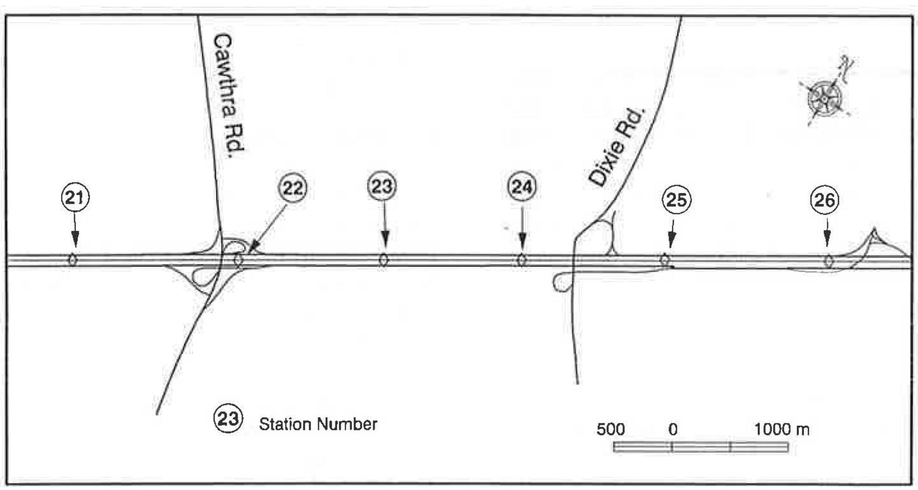

FIGURE 1 Study section: QEW Mississauga.

The analysis by Hall and Hall used the only station downstream of Cawthra Road at that time, which was located less than 200 m from the end of the on ramp. Since that time, the FTMS has been extended eastward so that there are now four additional stations, two of them within the section of interest. (After the second, the freeway widens to four lanes.) We have concentrated our bottleneck analysis on Station 25 (Figure 1), roughly 1.5 km downstream of the end of the entrance ramp. (For comparison with the earlier analysis, note that the system has since been renumbered, so that what was Station 22 in the paper by Hall and Hall is now Station 23.) Station 23 is very close to the end of the ramp, and Station 24 is only a single-loop station, so it does not provide speeds.

The data are available for 30-sec intervals 24 hr a day. Data for the morning peak period were used. The 1985 HCM refers to 15-min intervals in its definition of freeway capacity, but that is not a useful interval for the present analysis. Instead, three levels of detail are used: 30-sec values, 5-min values, and the full peak period. The 30-sec data were necessary for determining the beginning and end of a queue, and for a closer examination of the flow process, following Banks's approach (*9*). For the discussion of the distribution of capacity flows over different days, the unit of analysis is the peak period, which extends for 2 to 3 hr.

The focus of this study is freeway operations during good weather conditions. The study has been limited to normal weekdays, during which regular commuters use the facility, so the driver population is one that is quite familiar with the road and the usual traffic patterns. Data were available for the period April 25, 1990, through May 30, 1990. On May 7, peak-period flows were exceptionally low (5,400 vehicles/hr), suggesting that something was happening upstream that was not detected despite the use of closed-circuit television. This day was therefore left out of the analysis. May 21 was a holiday, so was omitted. There was rain during the peak period on May 16, 17, and 29, so these days were also omitted. For the remaining days, there were no incidents recorded by the operators that would affect this section. In addition, a cursory inspection of the Station 26 data indicated that it was consistently operating at high speeds, meaning that there was no downstream congestion that might not have been mentioned in the operators' log. In total, then, 20 days of data are used in these analyses, representing nearly ideal conditions.

The Mississauga FTMS collects flow data, but does not separately count truck volumes. To obtain these numbers, manual counts are needed. One such count was obtained on May 29, 1990. Although this was a rainy day, and not included in the analysis, the truck percentages can be taken to be representative of the other days, because the rain does not likely change the total volumes, but merely affects operating conditions. Truck percentages for the period 6:20 to 8:45 a.m. averaged 6 percent. There was no noticeable trend in the truck percentage over time.

CAPACITY DROP ISSUE

Two versions of the capacity drop issue are discussed. The first looks at the possibility of a capacity drop as described in the HCM and discussed above. The second looks at the idea of a capacity drop within the bottleneck flow.

Capacity Drop per HCM

The HCM quotation about the capacity drop arises from data like that shown in Figure 2, which is from Station 22 (see Figure 1). There are both a congested branch and an uncongested branch of the curve. There is even a visible gap between the two. According to the HCM discussion, the "stable" (i.e., uncongested) flow has a somewhat higher maximum than the "unstable" (i.e., congested) flow. But this station is not operating at capacity during "unstable" flow. The station is immediately upstream of a major entrance ramp.

Figure 3 presents data from Station 25 in the bottleneck for the same time period as that shown in Figure 2, namely, 5:30 a.m. to 10:00 a.m., covering the brief buildup to capacity (pre-queue data in the figure) over 2 hr of queue discharge flow and the flows after the morning rush has dissipated (post-queue). Because Station 25 is never in queue, there are no "unstable" data (i.e., high values of occupancy), and hence what would be the right-hand side of the flow-occupancy curve simply does not occur. Consequently there can be no drop in capacity described by a discontinuity of the curve. Yet this is the section of the freeway that is operating at capacity, and therefore the only place where one should be able to see if there is a drop in capacity.

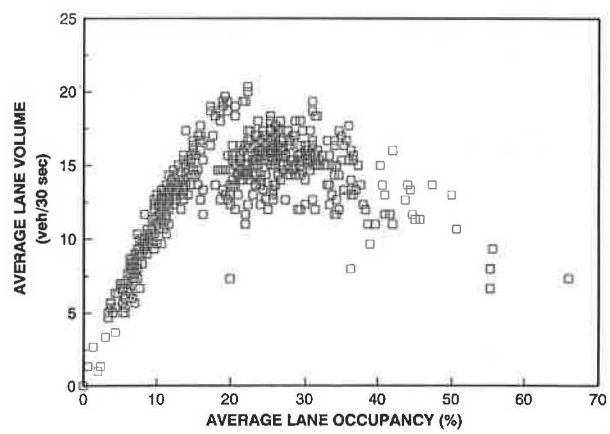

FIGURE 2 Flow versus occupancy: Station 22, May 11, 1990, 5:30–10:00 a.m., 30-sec data.

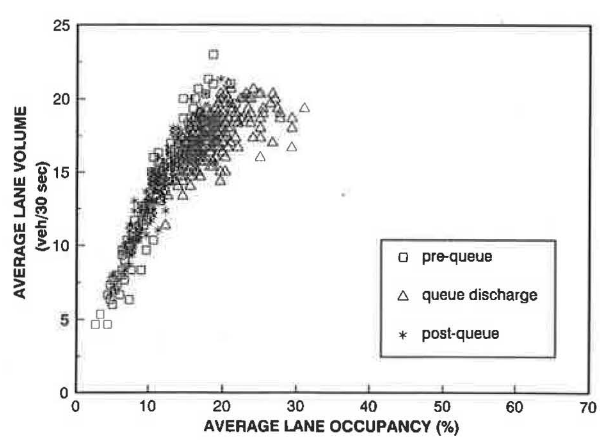

FIGURE 3 Flow versus occupancy: Station 25, May 11, 1990, 5:30–10:00 a.m., 30-sec data.

Closer inspection of the data for the three time periods within Figure 3 suggests that the pre-queue flows certainly occur at lower occupancies than do comparable queue discharge flows, and might reach a slightly higher maximum. The post-queue flows line up more closely with the pre-queue flows, and also occur at lower occupancies than comparable queue discharge flows. This occupancy difference is of course sensible if the queue discharge flows occur at lower speeds, which they do, for reasons explained by Persaud and Hurdle (*10*).

The data behind the HCM passage quoted earlier were collected in the queue upstream of a bottleneck, not in the bottleneck itself. The observed discontinuous curve simply reflects, for example, the volume of traffic entering an entrance ramp, which has caused the queue in the first place. The only place to look for a capacity drop (as opposed to simply a reduction in flow) is in the bottleneck itself. Banks did this (*9*), and found a minimal drop, but a statistically significant one, given that it occurred in eight out of nine cases. His method, comparing mean flows before and after queue formation, has been adopted here.

Capacity Drop Within the Bottleneck

Banks averaged the 30-sec flow rates before the onset of congestion with those for a similar interval immediately after congestion began. He used a "speed drop associated with the formation of the upstream queue" (*9*, p. 12) to identify the beginning of congestion. For the present analysis, queue presence was identified from the detector data at Station 22, where the queue first forms (see Figure 1). This station provides only volume and occupancy, not speed. Because it is known that the estimate of speed from volume and occupancy is not an unbiased one (*11*), and videotapes to calculate speeds were not available, the presence of the queue had to be determined directly from the volume and occupancy data.

Averages of flow and occupancy across the three lanes were used. The two tended to vary together in the period before congestion and to diverge during the congested period. Determining the exact beginning and end of congestion was, however, difficult from these numbers, so the ratio of occupancy to flow was used. Three values of the ratio were tested for the threshold level: 1.0, 1.1, and 1.2. A ratio of 1.0 gives a longer duration of bottleneck flows, some of which were very low, suggesting that demand was below capacity. A ratio of 1.2 excludes sustained periods (10 min) of high flows (5,800 vehicles/hr or more). A ratio of 1.1 or above persisting for 3 min was selected as the criterion for the identification of the start of a queue.

To consider capacity drop due to queue formation, it is necessary to restrict inspection of pre-queue flows to that period when demand is equal to capacity. If earlier, lower flows were included, the mean pre-queue flow rate would be reduced, biasing the results. The criterion for identifying the start of the queue is intentionally conservative. There is a considerable period before the start of the queue when the occupancy-flow ratio approaches congested conditions but does not maintain it. This is the period when the highest flow rates are being achieved, and therefore the ones to be kept distinct from queue discharge flow. In the following discussion, this portion of the pre-queue period is referred to as the transition period, that is, the period before the formation of a queue when demand is very high, resulting in high flows.

The ending time for this transition period of pre-queue flow is easy to identify in that it coincides with the start time for the queue. Identification of the start time for the transition period is more difficult. The occupancy-flow ratio for Station 22 could not be used because there was very little variation in the ratio below 1.1. Hence Station 25 data were used. In order to coordinate times when two stations were being used, it was necessary to take into account the travel time between them. The distance between the two stations is 2250 m. Calculation of the mean speeds during the period of high flows at Station 25 before the queue start time indicates that 1.5 min should be added to the start time of the queue at Station 22 to give the time of arrival of queue discharge flow at Station 25. The start time for the transition period was determined as follows, using the flow data at Station 25 and working backward in time from the interval when the queue discharge flow reached Station 25. When relatively low flows were observed in four consecutive 30-sec intervals, the latest time of such flows was taken as the tentative beginning of the transition period. Where two periods of comparatively high flows were separated by a sharp drop in flows for several consecutive 30-sec intervals, the t-test, or an approximation of it (depending on whether the variances in the two time periods were equal or not), was computed. When the difference in the means of the flow rates for the two periods was found to be significant, the tentative start time was accepted as the start of the transition. When the difference was found not to be significant, the two time periods were combined and were taken as one period.

The time at which the queue ended was identified in a similar fashion to the start of the transition period. A tentative ending time was identified on the basis of the first time that the occupancy-flow ratio at Station 22 fell below 1.0. Then attention was turned to Station 25. If low flows occurred there before the corresponding time, the same t-test approach was used to test average flows before and after the tentative end time. If the flows were equal, the queue end time was moved earlier accordingly.

With the pre-queue and queue discharge time periods identified, the task was then to determine whether the mean flow rates differed between the two periods. The two test statistics applicable in this case are the t-test for samples with equal variance and the approximation of t, which tests for the significance of difference between two means for samples with unequal variances. F-test results showed that at the 1 percent level, the variances for only 7 days were significantly different. At the 5 percent level, 10 of the 20 days have significantly different variances.

One-tailed tests were run for both t and its approximation, with emphasis on the "correct" t for each day, given its F-test result at the 5 percent level. There was a significant difference in the mean flow rates, pre-queue to queue discharge, for 12 days at the 5 percent level using the estimate for t and for 16 days using the t-test (Table 1). The "correct" test for each day (identified by an asterisk in Table 1) shows 12 days with a significant difference at the 5 percent level. At the 1 percent level, the difference in mean flows is significant for 10 days for the approximation of t, for 13 days using the t-test, and for 10 days using the "correct" test. There was only one day where queue discharge flow was higher than

TABLE 1 ONE-TAILED *t*-ESTIMATE AND *t*-TEST TO COMPARE MEAN VALUES OF 30-SEC RATES FOR PRE-QUEUE AND QUEUE DISCHARGE FLOWS

PRE-QUEUE DURATION			QDF DURATION		t	SIGNIFICANCE		t	SIGNIFICANCE	
Date	flowrate	minutes	flowrate	minutes	estimate	1%	5%	test	1%	5%
900425	6423	9.5	6173	156.0	1.48	no	no*	1.82	no	yes
900426	6498	13.5	5980	188.0	3.56	yes	yes	4.20	yes	yes*
900427	6400	9.0	6127	132.5	1.29	no	no*	2.14	no	yes
900430	6526	10.5	5919	180.0	4.40	yes	yes	4.78	yes	yes*
900501	6631	13.5	5995	183.0	3.59	yes	yes*	5.12	yes	yes
900502	6202	25.0	5920	179.5	2.77	yes	yes*	3.09	yes	yes
900503	6388	19.0	6097	161.0	2.42	no	yes*	2.95	yes	yes
900504	5750	6.0	6196	164.5	-1.42	no	no*	-2.54	no	no
900508	6669	16.5	5978	187.0	6.35	yes	yes	6.20	yes	yes*
900509	6471	13.0	6137	164.0	2.01	no	no*	2.74	yes	yes
900511	6505	17.0	6195	142.0	2.14	no	yes*	2.74	yes	yes
900514	6415	17.5	5959	176.5	3.75	yes	yes	3.77	yes	yes*
900515	6669	31.5	6199	118.0	4.98	yes	yes	5.20	yes	yes*
900518	6298	15.5	6218	132.5	0.62	no	no	0.69	no	no*
900522	6538	14.5	5902	170.0	5.31	yes	yes	5.10	yes	yes*
900523	6380	9.0	6214	148.0	0.94	no	no	1.11	no	no*
900524	6413	18.0	6313	128.5	0.64	no	no*	0.83	no	yes
900525	6424	25.5	6173	153.0	2.71	yes	yes	2.82	yes	yes*
900528	6484	14.5	5991	171.5	3.55	yes	yes*	4.36	yes	yes
900530	6123	17.5	6066	164.0	0.54	no	no	0.60	no	no*
Weighted mean	6432		6075							

* correct t calculation given the F-test results at the 5% level.

that in the transition period. Because this is an even split, it was deemed appropriate to consider the trend in the results, as Banks did. If the two means are equal, the probability of obtaining only one higher value for queue discharge flow in 20 observations is only 0.00001. Hence these results appear to support Banks's conclusion that there is a drop in achievable maximum flow rates when the queue forms. A comparison of the overall weighted mean flows in the two time periods, taken over the 20 days, shows a difference of 357 vehicles/hr, or 5.8 percent.

Two aspects of this result warrant discussion. First, there is the question of how these results relate to those reported by Hall and Hall (*5*). Second, the drop in flows when the queue forms highlights a problem associated with the concept of sustainable flows. Hall and Hall selected the maximum 40-min flow rates as representing capacity, which gave a value of about 6,500 vehicles/hr as the capacity for the same bottleneck. This value is roughly consistent with the 6,400 vehicles/hr found here as the average for pre-queue flows, but the duration of the pre-queue flows is not as long as the 40 min found in their paper. There are two possible reasons why the high flows did not last so long. The first is simply that Hall and Hall did not have as good information with which to ascertain the time of queue formation. The second is based on the idea that flows were not quite so heavy in the data they used (late 1987 and one day in July 1988), with the result that these high maximum flow rates were sustainable for a longer period before queue formation. This explanation is consistent with the HCM statement that Level of Service E (i.e., capacity operation) is "unstable." It will last longer on some days than on others.

However, given the information now available, it seems likely that the 40-min values that Hall and Hall identified as capacity represent the pre-queue flows. First, the current data show that it can take up to 30 min for a queue to stabilize. Second, a comparison of the observed speeds and reconsideration of the distances involved support the interpretation that the speed drop in their data represented the onset of queue discharge flow, not of operations in a queue as they suggested. For the current Station 25 data, the observed mean speed was 85 km/hr in the transition period and 73 km/hr for the queue discharge flows. Hall and Hall stated that their station was 800 m downstream of the entrance ramp, but in fact it is 650 m downstream from the beginning of the merging lane and less than 200 m downstream of the end of that lane. Hence their lower speeds are consistent with queue discharge flow, as defined by Persaud and Hurdle's expectations for speed recovery downstream of a queue (*10*).

But this correction of the Hall and Hall result serves only to highlight the difficulties with the HCM definition of capacity as "sustainable" flows. In Chapter 3, "Basic Freeway Segments," the HCM states that "freeway capacity is the maximum sustained (15-min) rate of flow" (*1*, p. 3–3). Clearly, the pre-queue flows meet this definition. In the Hall and Hall paper, such flows lasted for 40 min; in this paper they last from 6 min to over 30 min and 10 of the 20 exceed the 15-min requirement. (Two more fall short by only 30 sec). These 10 cases had average flows of between 6,123 and 6,660 vehicles/hr. According to the HCM definition of a 15-min sustained flow, these values represent capacity, and the 6,500 value put forward by Hall and Hall is valid. Hurdle and Datta (*8*), however, have suggested that capacity should be defined as queue discharge flow. On the basis of the start and end times, there was only one instance in which the queue discharge flow lasted less than 2 hr (on May 15, when it fell 2 min short of 2 hr). In four instances, it extended for 3 hr or more. There is no question that queue discharge flow is sustainable. If both types of flow are "sustainable," which one represents capacity? We return to this question later.

DISTRIBUTION OF QUEUE DISCHARGE FLOWS OVER TIME

For analysis of the behavior of queue discharge flows over time, data on the 5-min flow rates from all 20 available days were taken together. For convenience and comparability, the 5-min intervals for each day were begun on the hour. The

5-min interval that contained the start of queue discharge flow was not included, nor was the interval containing the end of queue discharge flow. All intervals between these two were included in the analysis. The 5-min counts for each of the three lanes were summed and multiplied by 12 to get the hourly flow rate across all three lanes. As stated earlier, no truck correction was made in these values. Table 2 contains examples of such data from 6 days.

The mean, standard deviation, and duration of queue discharge flow for the 20 days based on these 5-min data appear in Table 3. The values differ slightly from those in Table 1 because of the omission of a small number of 30-sec intervals. The daily mean queue discharge flows ranged from 5,891 to 6,307 vehicles/hr across the section. There were eight cases of mean flows under 6,000 vehicles/hr. Three of these had relatively high standard deviations, over 300 vehicles/hr. The high standard deviations seem to be a consequence of very low flow rates: the two with the highest standard deviations recorded flow rates below 5,000 vehicles/hr for two consecutive 5-min intervals.

Given this distribution of mean queue discharge flow over 20 days, what should one say is capacity? For eight days there were average flow rates below the HCM values; for 12 days there were averages above the HCM values. Twenty observations do not make for a very useful histogram, so all 620 of the 5-min volume counts have been used to create one (Figure 4) in which the data have been grouped in cells of 100-vehicle/hr width and frequencies plotted at the midpoint of the cells. The modal value (the peak of the diagram) is just above 6,000, as are the median (6,072) and mean (6,071). This frequency distribution represents a near-normal distribution, with a Pearson coefficient of skewness of only −0.011. It is reasonable to assume that a distribution of capacities for a larger number of days would show somewhat the same shape, although it might be a bit narrower. This frequency distribution is for queue discharge flows, but a similar diagram could be constructed for pre-queue flows. What is the proper segment of either distribution to select as "capacity," given the HCM wording of "reasonable expectation"? The expected value is, of course, the mean. In that case, these data suggest a discharge flow rate of 6,071 vehicles/hr. However, selecting the mean, with a normal distribution, implies that capacity will not be achieved half of the time. Does that suggest that a lower percentile of the distribution should be selected? The answer is not obvious, but certainly it will be valuable to identify explicitly the fact that there is a distribution of achievable maximum sustained flows at any given location, whether capacity is queue discharge flow or pre-queue flow. The same kind of analysis holds over space as well as over time.

TABLE 2 EXAMPLES OF QUEUE DISCHARGE FLOWS AT STATION 25: 5-MIN VALUES

End time	900427 flow/h	900501 flow/h	900508 flow/h	900509 flow/h	900515 flow/h	900524 flow/h
6:30	6636					
6:35	6372	6228		5664		
6:40	6324	6204	6000	6060		
6:45	6012	6132	6060	6084		7092
6:50	6300	5760	6060	5832		6252
6:55	5964	6228	6168	5988		6204
7:00	6120	5868	5868	6276	6156	6216
7:05	6216	5976	5880	6048	5820	6084
7:10	6228	5856	5892	6144	6084	6168
7:15	6180	6240	6156	6240	6468	6468
7:20	6072	5868	5676	6132	6264	6528
7:25	5880	5988	5928	6048	6264	6408
7:30	6024	5868	6084	6288	6252	6420
7:35	6276	6348	5640	6204	6384	6300
7:40	6156	5676	5772	6264	6708	6216
7:45	5964	5748	6012	6336	6168	6252
7:50	6432	6132	5988	6444	5712	6252
7:55	6192	6036	6012	6480	6468	6168
8:00	6084	5628	6336	6120	6264	6420
8:05	6084	5760	6056	6024	6072	5772
8:10	5844	5844	5892	6036	6288	6324
8:15	6072	5988	6180	6064	6156	6600
8:20	6144	5748	6012	6108	6024	6432
8:25	6276	5976	6384	6216	5952	6144
8:30	5952	5904	6240	5976	6132	6480
8:35	5988	6384	5952	5976	6300	6360
8:40	6000	5976	5868	6132	6036	6048
8:45		6012	6348	6336	6324	6060
8:50		6276	5796	6060	6048	
8:55		6156	6288	6192		
9:00		6288	5880	6420		
9:05		6264	5892	6048		
9:10		5940	6084	6060		
9:15		5580	6360			
9:20		5904	5700			
9:25		5844	5040			
9:30						
9:35						
9:40						
Mean	6121	5989	5985	6134	6189	6307
Std. Dev.	153	215	258	172	218	244
Duration(min)	130	175	170	160	115	125

TABLE 3 MEANS, STANDARD DEVIATION, AND DURATION OF QUEUE DISCHARGE FLOWS AT STATION 25: 5-MIN VALUES

Date	Mean flowrate	Standard Dev.	Duration minutes
900425	6170	181	150
900426	5976	301	185
900427	6121	153	130
900430	5913	173	175
900501	5989	215	175
900502	5921	235	180
900503	6094	184	155
900504	6205	182	160
900508	5985	258	170
900509	6134	172	160
900511	6198	232	140
900514	5959	397	175
900515	6189	218	115
900518	6227	248	125
900522	5891	365	165
900523	6208	191	145
900524	6307	244	125
900525	6169	182	150
900528	5970	160	165
900530	6062	188	160

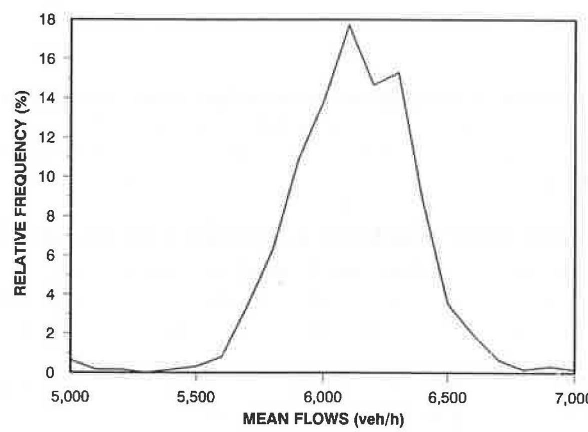

FIGURE 4 Frequency histogram for 5-min queue discharge flows, Station 25.

This discussion of queue discharge flows also raises the concept of stable flows, as used in the HCM: operations in Level-of-Service E are said to be "extremely unstable," in the sense that any disruption "establishes a disruption wave which propagates through the upstream traffic flow" (*1*, p. 3–10). It is indeed upstream that the effects are felt, where the queue forms. Downstream, in the bottleneck (such as Station 25 in this paper), flow rates remain high (although not quite as high as before), and speeds drop somewhat, with the actual value depending on how far downstream of the queue the measurement is taken. Downstream, where capacity flows are actually occurring, operations are in fact extremely stable, not unstable, as shown by their 2- to 3-hr duration. It therefore seems inappropriate for the Level-of-Service E description to assert that operations at or near capacity are unstable.

Note that the HCM also states that "average travel speeds at capacity are approximately 30 mph" (about 50 km/hr) (p. 3–10). These data contradict that assertion also. At capacity during pre-queue operations, speeds are reduced only slightly from free-flow values. In the bottleneck during queue discharge flow, speeds depend on the distance downstream. In the queue, speeds may well be roughly 50 km/hr (or less), but the queue is not operating at capacity. There is a restriction, such as an entrance ramp with high volume, that is forcing it to operate at lower flow rates.

CONCLUSIONS

There are four conclusions from this analysis, one relating to the notion of a two-regime flow-occupancy diagram, the other three relating to the definition of capacity. This section closes with some comments about future research.

For a flow-occupancy diagram to be able to show two regimes, the data must have been obtained within a queue (to get the right-hand portion of the curve). Figure 3, showing the flow-occupancy data for station 25, shows only the left-hand side of such a curve, reaching an occupancy of perhaps 28 percent. Similar figures from stations in a queue, such as Figure 2, show a lower maximum for the congested data than for the uncongested precisely because of the bottleneck effect: a queue forms at the location because something else (usually an entrance ramp) has taken up a portion of the available capacity, leaving a reduced flow possible within the queue. The only legitimate place to look for a capacity drop is in the bottleneck. The notion of two capacities may be right, but it has been raised for the wrong reasons.

The first conclusion about capacity is that there is a capacity drop in the bottleneck. Once a queue has formed upstream, the bottleneck location does not handle as many vehicles as it did before the queue formation. The demand is clearly there, as shown by the presence of the queue, so the reduced flow is a consequence of the way drivers accelerate away from the queue. From these data, the reduction in flow appears to be about 5 to 6 percent, from just over 6,400 vehicles/hr to just under 6,100, on average.

The second conclusion is about the duality of capacity. One capacity is the pre-queue flow, which can last for 30 min and more. Queue discharge flows can last for 2 to 3 hr. Both then meet the HCM requirement for sustainable flow. The important issue is the practical one: once congestion occurs on a facility, it is the queue discharge flow rate that will govern the time to recovery. If queues can be delayed during high flows, there is a 5 to 6 percent "bonus" flow available.

The final conclusion about capacity should be its numerical value, given the previous conclusion. However, the analysis of the distribution of queue discharge flows over time shows that there is indeed a distribution, not simply a single value achieved on all similar days. If the mean of that distribution is taken, then the value for queue discharge flow for this location is 6,071 vehicles/hr. Applying the truck percentage found by manual counting on the one day (6 percent) and a truck equivalent of 1.7 (*1*, p. 3–13) for a general segment of level terrain, capacity is just under 6,700 passenger cars/hr. Using the mean value from Table 1 for the pre-queue flows gives 6,432 vehicles/hr, or 7,088 passenger cars/hr. This corresponds favorably with the value of 2,300 pcplph for rural multilane highways (*4*). However, it is not obvious that the mean is the proper value to take from the distribution. More consideration is needed.

More research is needed on several of the topics addressed in this paper. The most important is probably the validation at other locations of the conclusion found with these data that there is a drop in capacity when a queue forms. As well, it would be useful to have equally well-specified information on the numerical value of capacity (both pre-queue and during queue discharge) at other locations. On a more theoretical note, Athol and Bullen (*12*) suggested 18 years ago that the probability of transition from uncongested to congested flow depends on the value of the flow. The discussion in this paper offers a different interpretation, namely, that operations at quite high flows fed by a queue can be sustained almost indefinitely (up to 3 hr in these data) without breaking down further. Instead, breakdown occurs at the entrance ramp where total flow exceeds capacity. It would be worthwhile to investigate in more detail which of these models is more reliable for freeway operations.

ACKNOWLEDGMENTS

The work reported in this paper represents the early stages of the analysis for Agyemang-Duah's master's thesis. Since presentation of this paper, but before its publication, he has completed his thesis (*13*), and a paper based on the full 52 days of data has been presented at the Karlsruhe conference (*14*). This work was supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada. The data were provided by the Freeway Traffic Management Section, Ontario Ministry of Transportation.

REFERENCES

1. *Special Report 209: Highway Capacity Manual 1985*. TRB, National Research Council, Washington, D.C., 1985.
2. *Special Report 87: Highway Capacity Manual 1965*. HRB, National Research Council, Washington, D.C., 1965.
3. W. R. McShane and R. P. Roess. *Traffic Engineering*. Prentice Hall, Englewood Cliffs, N.J., 1990.
4. W. R. Reilly, D. W. Harwood, J. M. Schoen, R. O. Kuehl, K. Bauer, and A. D. St. John. *Capacity and Level of Service Pro-*

cedures for Multilane Rural and Suburban Highways. Preliminary draft final report. NCHRP, TRB, National Research Council, Washington, D.C., 1988.
5. F. L. Hall and L. M. Hall. Capacity and Speed Flow Analysis of the QEW in Ontario. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, D.C., 1990, pp. 108–118.
6. J. S. Drake, J. L. Schofer, and A. D. May. A Statistical Analysis of Speed Density Hypotheses. In *Highway Research Record 154*, TRB, National Research Council, Washington, D.C., 1967, pp. 53–87.
7. A. Ceder and A. D. May. Further Evaluation of Single- and Two-Regime Traffic Flow Models. In *Transportation Research Record 567*, TRB, National Research Council, Washington, D.C., 1976, pp. 1–15.
8. 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.
9. J. H. Banks. Flow Processes at a Freeway Bottleneck. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, D.C., 1990, pp. 20–28.
10. B. N. Persaud and V. F. Hurdle. Some New Data That Challenge Some Old Ideas About Speed-Flow Relationships. In *Transportation Research Record 1194*, TRB, National Research Council, Washington, D.C., 1988, pp. 191–198.
11. F. L. Hall and B. N. Persaud. An Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems. In *Transportation Research Record 1232*, TRB, National Research Council, Washington, D.C., 1989, pp. 9–16.
12. P. J. Athol and A. G. R. Bullen. Multiple Ramp Control for a Freeway Bottleneck. In *Highway Research Record 456*, TRB, National Research Council, Washington, D.C., 1973, pp. 50–54.
13. K. Agyemang-Duah. Investigation of Some Unresolved Issues in Freeway Capacity. M.A. thesis. Department of Geography, McMaster University, Hamilton, Ontario, Canada, 1991.
14. K. Agyemang-Duah and F. L. Hall. Some Issues Regarding the Numerical Value of Capacity. In *Highway Capacity and Level of Service: Proceedings of the International Symposium on Highway Capacity, Karlsruhe* (U. Brennolte, ed.), Balkema, Rotterdam, 1991, pp. 1–15.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.