Gap Acceptance Characteristics for Ramp-Freeway Surveillance and Control

DONALD R. DREW, Associate Research Engineer, Texas Transportation Institute, Texas A & M University

The purpose of this investigation was to determine gap acceptance characteristics and merging delay characteristics for 6 inbound entrance ramps on the Gulf Freeway Surveillance and Control Project. A 20-pen graphic recorder operated by 4 men stationed adjacent to the merging area between the freeway and the frontage road was used in collecting the data.

Merging vehicles were divided into 2 groups: those in which the driver rejected gaps before finally accepting a gap and those in which the driver of a ramp vehicle accepted the first gap. The former was referred to as "stopped" vehicles and the latter as "moving" vehicles. The critical gap for stopped vehicles was found to be about 20 percent higher than for moving vehicles. In addition, it was concluded that the critical gap (median) for the merging maneuver from an entrance ramp is independent of the freeway volume but is apparently affected by ramp geometrics and ramp controls.

A distribution of critical gaps was formed and fitted to a gamma distribution. Merging delay values calculated using this distribution were shown to be higher than those calculated assuming that all drivers have

If the distribution of time spent by the merging vehicle at the head of the queue is approximated by a gamma distribution, the entrance ramp merging operation may be considered within the context of classical queuing theory. Based on this queuing model, a ramp metering technique was developed which takes into account the individuality of entrance ramps. Finally, it is shown how the need exists for an automatic ramp control technique combining the microscopic approach developed in this paper with the systems or macroscopic approach which has been used as the basis for the past manual ramp metering experiments on the Gulf Freeway.

•THE Gulf Freeway Surveillance and Control Project sponsored by the Texas Highway Department and U.S. Bureau of Public Roads has the basic objective of developing criteria for the design and operation of automatic surveillance and control systems which would permit the attainment of acceptable levels of service on urban freeways during periods of peak traffic demand.

A traffic surveillance system should involve the continuous sampling of basic traffic characteristics for interpretation by established control parameters, to provide a quantitative knowledge of operating conditions necessary for immediate rational control and future design. Figure 1 shows some relationships between many of these control parameters, the basic traffic characteristic, and their ultimate applications. Most of

Paper sponsored by Committee on Characteristics of Traffic Flow and presented at the 45th Annual Meeting.

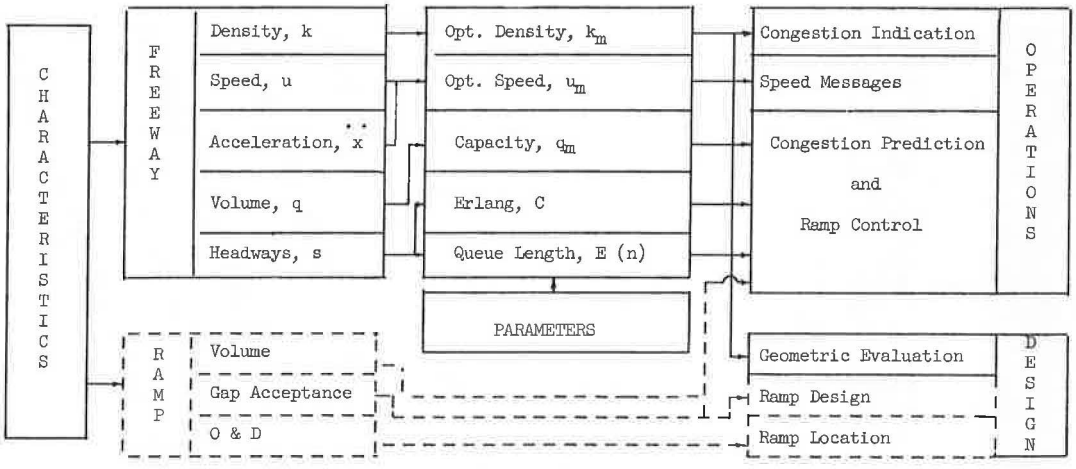


Figure 1. Analysis of freeway traffic congestion.

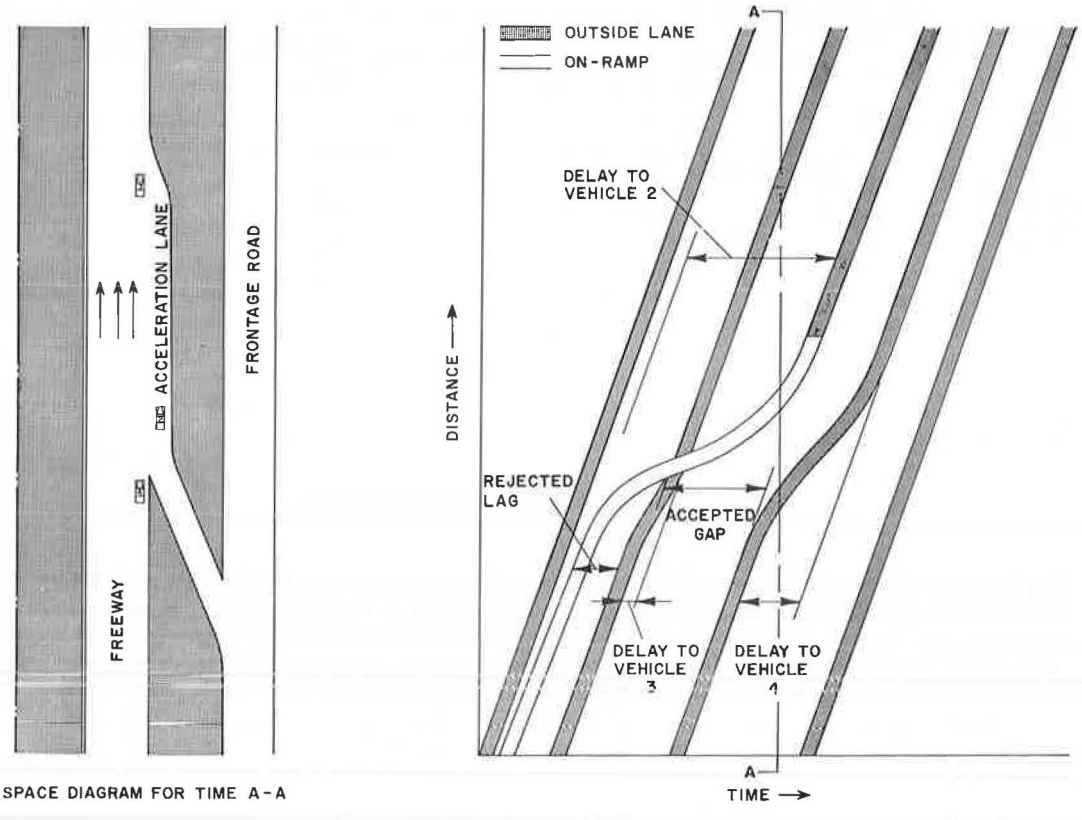
the freeway characteristics and freeway control parameters shown in Figure 1 were evaluated in previous studies and have been discussed in various project research reports (1, 2). These operational studies also pointed to the need for some form of freeway ramp control. Utilizing a comprehensive systems plan for metering and controlling the inbound entrance ramps, a significant improvement in the inbound level of service was obtained (3). As control techniques become more sophisticated, it is felt that greater reliance will be placed on ramp metering and less on complete ramp closure, making it imperative that the characteristics associated with both normal and controlled freeway ramp operation be appreciated.

FREEWAY MERGING MANEUVER

One of the most important elements of freeway operation is the merging maneuver from an entrance ramp onto the shoulder lane of a freeway. Merging, in this case, is simply the absorption of the stream of ramp traffic. A vehicle is said to merge when it moves from the acceleration lane to the shoulder lane. The driver in the merging vehicle must, of course, select the proper moment to execute the merging maneuver based on his judgment as to whether or not a gap which he can enter is large enough for a safe merge.

A merging gap is the interval of time between the arrival of two successive shoulder lane vehicles at a point in the merging area. The concept of gap acceptance is important in describing the interaction of the freeway and ramp streams. It is assumed that the ramp driver measures each gap t in the shoulder lane and either merges (accepts the gap if $t>T$) or waits (rejects the gap if $t<T$) where T may be assumed to be the driver's critical gap for that decision (Fig. 2).

Most of the work on gap acceptance applies to the crossing maneuver at intersections. Greenshields (4) defined the "acceptable average-minimum time gap" as a gap accepted by half of the drivers. Raff (5) used a slightly different parameter, the "critical lag." A lag is the interval from the arrival of a minor stream vehicle at the point of conflict to the arrival of the next major stream arrival (Fig. 2). The critical lag is the size lag which has the property that the number of accepted lags shorter than the critical lag is equal to the number of rejected lags longer than the critical lag. For the data presented in Raff's report, the Greenshields gap quantity averages 0.2 sec longer than Raff's critical lag. The principal use of both parameters, or any such parameter, is to simplify the computation of the number of cars delayed by permitting the assumption that all intervals shorter than the critical value (lag or gap) are rejected whereas all longer intervals are accepted.



The speed of the merging vehicle is important in considering the distribution of gaps acceptable to the merging driver at a freeway entrance ramp. By changing the speed of his own car, a ramp driver also changes his time criterion (critical gap) for a safe merge. In simulating entrance ramp operation, the Midwest Research Institute (6) used a distribution of gaps for nonstop merges with a range of values roughly equal to half those values used for vehicles merging from a stop. Actually, because of the scarcity of published characteristics on entrance ramp operation, the exact relationship between relative speed and gap acceptance is still one of conjecture.

Although numerous data are available regarding merging at an intersection, this information would probably not be applicable as a model for describing ramp merging. The availability of an acceleration lane not only increases the ability of the merging vehicle to minimize its relative speed, but provides the driver with a limited opportunity to inspect different parts of the freeway stream for places to merge. Such considerations as a driver's desire to merge as safely, as quickly, or as directly as possible make the problem a question of policy and complicate the exact formulation of a mathematical model (7).

PURPOSE AND OBJECTIVES

The purpose of this investigation is to determine the gap acceptance characteristics for entrance ramps on the Gulf Freeway Surveillance and Control Project. This information will be used to calibrate some of the ramp metering models slated for future use on this project (1, 2). It is also believed that this research will help to fulfill the very definite need for gap acceptance data for ramp simulation inputs (6), as well as the formulation and testing of ramp-freeway merging models (8).

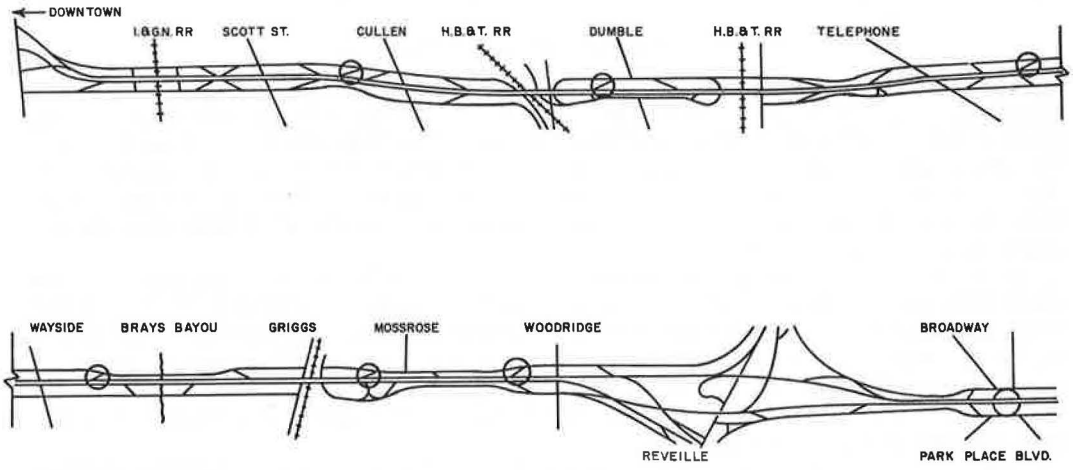


Figure 3. Location of study ramps on Gulf Freeway Surveillance Project.

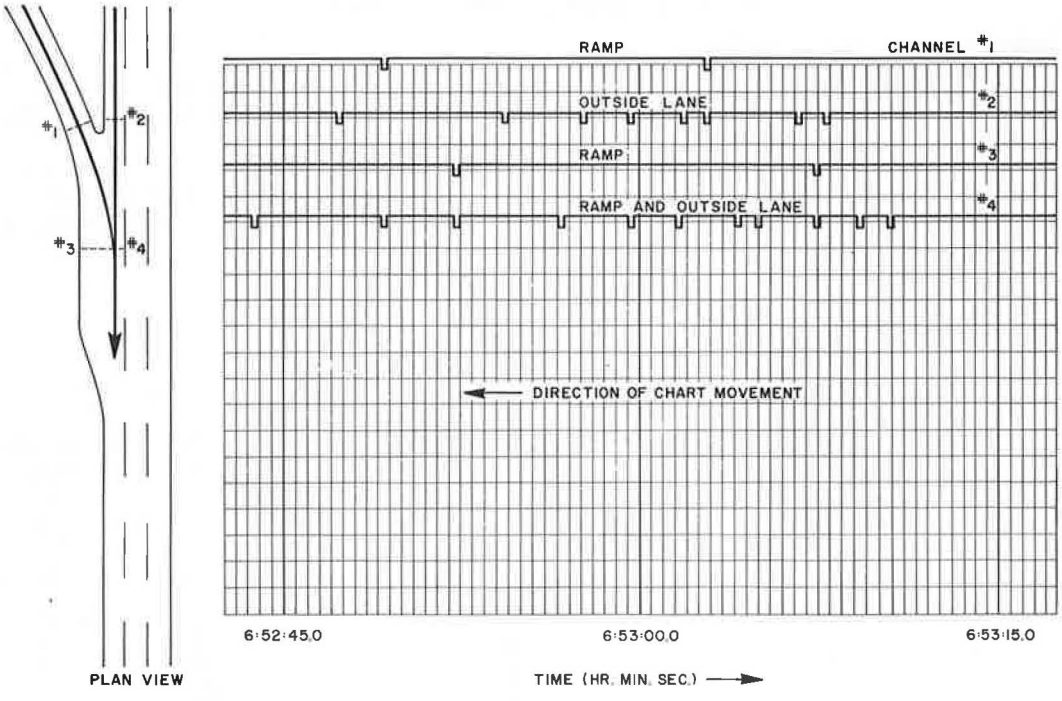


Figure 4. Section of recording chart showing data collection procedure.

Specific objectives of this investigation are (a) to measure gap acceptance characteristics; (b) to determine delay characteristics at the merging area; and (c) to relate these characteristics to freeway operations and design.

STUDY PROCEDURE

All study sites were located on the inbound entrance ramps of the Gulf Freeway in Houston, Texas (Fig. 3). This facility has three 12-ft lanes in each direction. The through lanes of the Gulf Freeway overpass the intersecting roadways producing a "roller-coaster" effect. This design seems to compromise the merging maneuver to the extent that where ramps enter before the overpass some of the acceleration lanes terminate at the overpass structures, and where ramps enter after the overpass the sight distance may be reduced.

A 20-pen graphic recorder operated by four men stationed adjacent to the merging area between the freeway and the frontage road was used in collecting the data. The code for recording traffic behavior is summarized in Figure 4. One pen was assigned to the decision point on the entrance ramp and one pen to a similar point opposite the ramp nose on the outside lane of the freeway. The other two pens were assigned to a point in the merging area, one pen recording both outside lane and merging vehicles and one pen merging vehicles only.

The studies were planned so as to be conducted during periods of congestion-free operation, yet during periods when the ramp demand was high enough to reduce study duration.

The clock times for the four channels shown in Figure 4 were placed on punch cards. A program was written for the IBM 7094 which would output the ramp vehicle arrival time, outside lane arrival time, and time all vehicles passed the designated point in the merging area. The output also included all the gaps rejected by a ramp vehicle plus the gap finally accepted. The study included over 12,000 gaps for 6 entrance ramps.

GAP ACCEPTANCE CHARACTERISTICS

Most studies of gap acceptance have been concerned with the determination of the critical gap, a time gap just as likely to be accepted as it is to be rejected. The probability that it will be accepted is thus equal to one-half.

In evaluating the critical gap, a given gap must be either accepted or rejected by a given driver and as such is binomial in nature. Yet each driver can accept only one gap while he can reject several of them. This means that if all rejected gaps are given

TABLE 1

ACCEPTED AND REJECTED GAPS AT DUMBLE RAMP[a]

Length of Gap t (sec)	Stopped Vehicles		Moving Vehicles		All Vehicles	
	No. Accepted Gaps<t	No. Rejected Gaps>t	No. Accepted Gaps<t	No. Rejected Gaps>t	No. Accepted Gaps<t	No. Rejected Gaps>t
0.0	0	100	0	89	0	189
Δt = 0.5	0	100	0	89	0	189
1.0	0	95	0	80	0	175
1.5	0	71	1	52	1	123
2.0	2	49	7	27	9	76
2.5	11	34	a = 13	c = 16	a = 24	c = 50
3.0	a = 15	c = 20	b = 26	d = 7	b = 41	d = 27
3.5	b = 23	d = 10	38	4	61	14
4.0	32	5	46	3	78	8
4.5	41	4	55	3	96	7
5.0	48	2	63	2	111	4
5.5	57	0	70	1	127	1
10.0	100	0	106	0	206	0

[a]Critical Gap, $T = t + \frac{(c-a)\,\Delta t}{(b+c)-(a+d)}$; T (stopped) = 3.1; T (moving) = 2.5; and T (all) = 2.8

the same weight as accepted gaps, then the percentage of intervals accepted for a particular size will not be a true measure of the percentage of drivers who find such an interval acceptable. If the percentage of intervals accepted is to be used to determine the percentage of drivers who are willing to accept them, then the same number of intervals must be counted for each driver. Raff (5) accomplished this by counting only lags and ignoring the gaps.

For this investigation, merging vehicles were divided into two groups, those in which the driver rejected gaps before finally accepting a gap and those in which the driver of a ramp vehicle accepted the first gap. The former is referred to as "stopped" vehicles and the latter "moving" vehicles, the implication being that if a ramp vehicle rejects the first gap he must stop. Theoretically, this would not have to be true where there is an acceleration lane present. Yet, on the Gulf Freeway, when traffic is heavy, ramp vehicles tend to shy away from using the acceleration lane, perhaps for fear of being "trapped" at the end while other ramp vehicles behind them accept the available gaps.

This arbitrary classification of stopped and moving merges afforded the means of eliminating any bias due to the inclusion of all rejected gaps. For example, in the case of stopped vehicles only two gaps were considered for each vehicle, the largest rejected gap and the gap finally accepted. To evaluate the gap acceptance characteristics for moving vehicles only the first gap available to those ramp vehicles not delayed by previous ramp vehicles was considered, much in the same manner that Raff treated lags.

The number of gaps accepted and rejected at the Dumble entrance ramp is given in cumulative form in Table 1 for stopped and moving vehicles, and then for all vehicles. The critical gap may be determined algebraically as indicated in Table 1 or graphically as shown in Figure 5, in which two cumulative distribution curves depict the number of accepted gaps shorter than t and the number of rejected gaps longer than t. The value of t for which these two curves intersect is the critical gap.

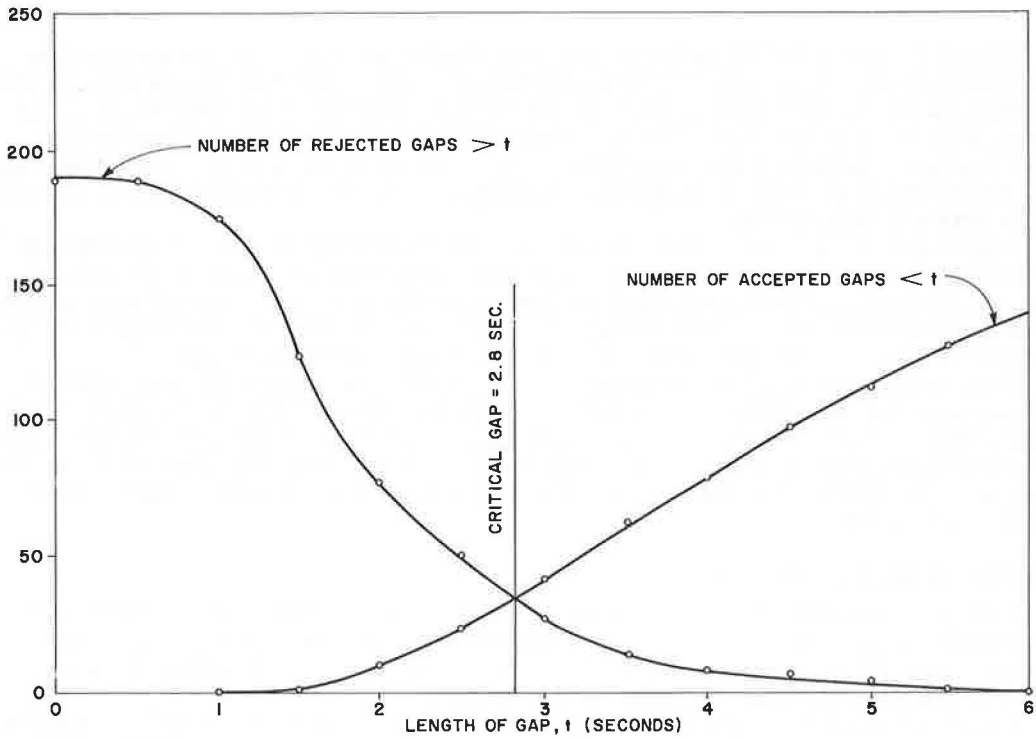


Figure 5. Accepted and rejected gaps for all vehicles at Dumble ramp.

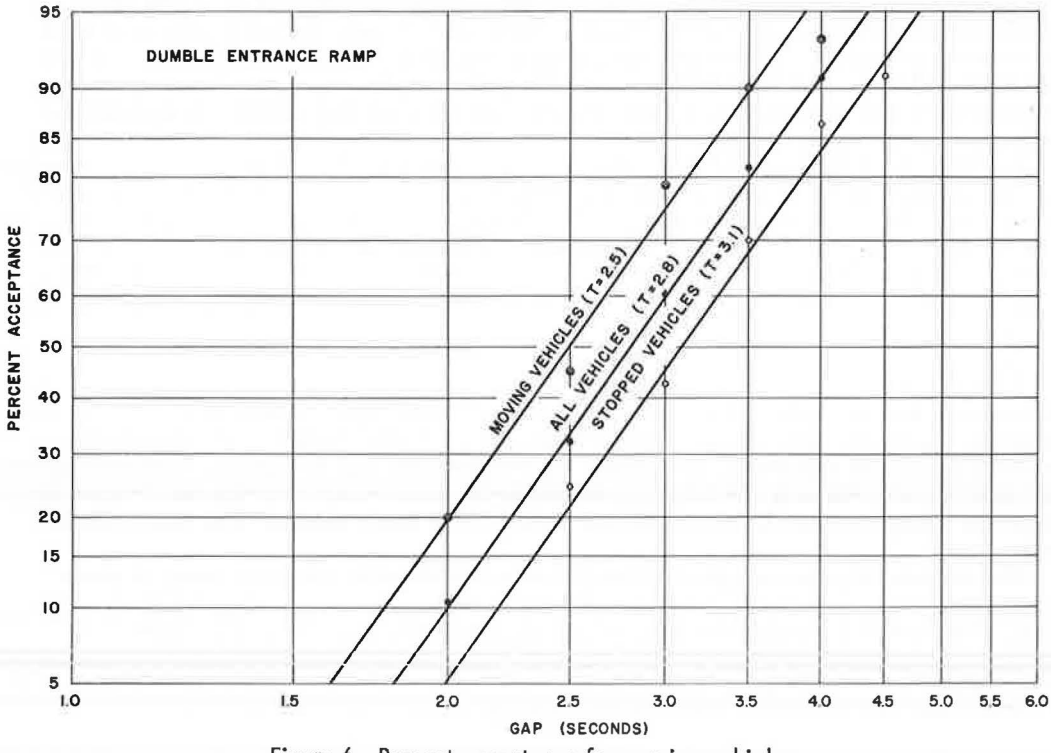


Figure 6. Percent acceptance for merging vehicles.

Bissell (9) calculated the percentage of acceptance observed at an intersection for each time group, plotted the data on log probability paper, and then drew a straight line through the point. Solberg and Oppenlander (10) also calculated the percentage acceptance, but converted the probability scale to probits and then fitted a regression line with confidence limits through the points. The critical gap was taken as the median gap, the 50 percentile in the case of Bissell, the probit of acceptance equal to 5.0 in the case of Solberg and Oppenlander.

Figure 6 shows a procedure combining some of the features of the Raff and Bissell methods. Using Table 1, the number of accepted gaps less than t was divided by the sum of the number of accepted gaps less than t plus the number of rejected gaps greater than t. A straight line is drawn through these points on the log-probability graph (Fig. 6) with the 50 percentile or median gap used to establish the critical gap.

The reason for using this new procedure for establishing the critical gap is that it serves to define the parameter in a way relating directly to the manner in which it is used. The probability of accepting a smaller gap than the cirtical gap is exactly equal to the probability of rejecting a longer gap, thus equating the critical gap to the median gap for a distribution of critical gaps. This concept will be discussed in more detail later in the paper.

Factors Affecting Critical Gap

Raff (5) investigated factors influencing the critical lag at urban intersections and concluded that the critical lag (a) varies directly with main street speeds and main street width; (b) is independent of the main street volume; and (c) is influenced by sight obstructions and the side-street traffic pattern.

The effect of relative speed on the critical gap for merging ramp vehicles is evident in Figure 6. At the Dumble entrance ramp, the critical gap for moving vehicles (based

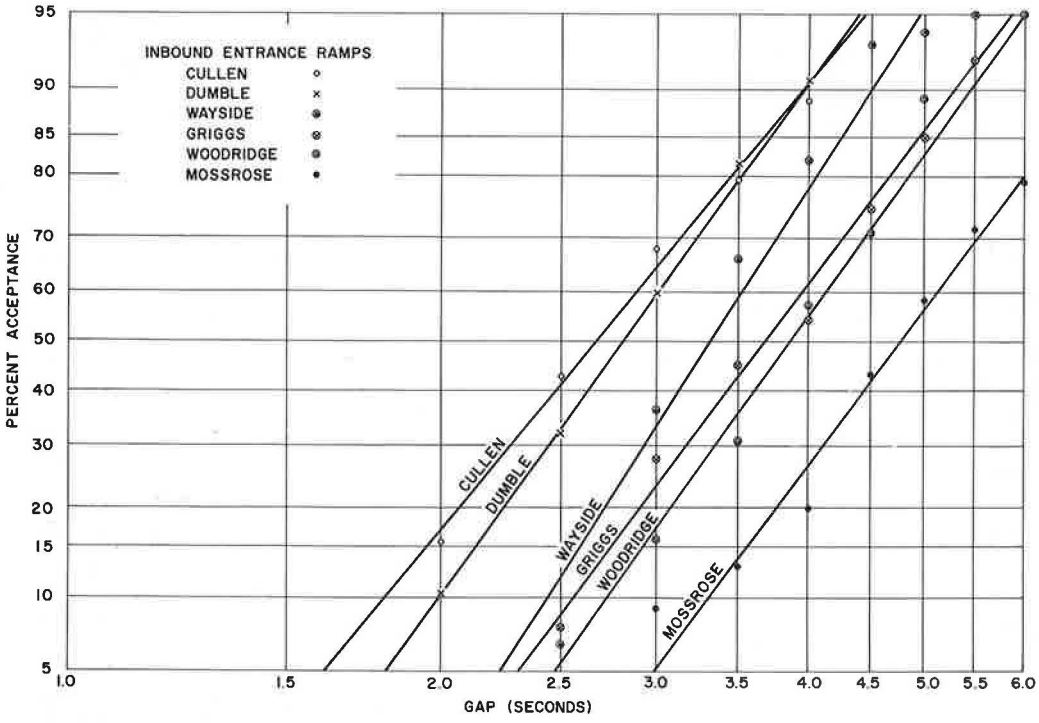


Figure 7. Variations in percent acceptance for merging vehicles at different ramps.

TABLE 2

RELATIONSHIP BETWEEN CRITICAL GAP AND OTHER ENTRANCE RAMP ELEMENTS

Element	Cullen	Dumble	Wayside	Griggs	Mossrose	Woodridge
Critical gap (sec)	2.7	2.8	3.3	3.7	4.7	3.8
Avg. freeway volume outside lane during period of observ. (veh/hr)	1,350	1,240	1,100	700	1.060	970
Percent grade on freeway	0.0	+0.5	-1.5	+1.5	+1.0	-3.0
Ramp length (ft)	320	320	330	230	190	300
Ramp width (ft)	20	20	20	20	15	20
Angle of entry (deg)	12	12	11	11	10	12
Acceleration lane length (ft)	–	330	–	310[a]	250[a]	370
Auxiliary lane length (ft)	350	–	510	–	–	–

[a]Taper-type acceleration lane terminating at overpass structure.

on the first gap evaluated) is seen to be 2.5 sec as against a critical gap of 3.1 sec for stopped vehicles (vehicles rejecting at least one gap before finally accepting a gap).

The graph of the percent acceptance for merging vehicles at different ramps (Fig. 7) indicates that the critical gaps were not the same, but varied from ramp to ramp as they do among intersections. To determine to what extent this variability was due to ramp geometrics, the relationship between critical gap and entrance ramp elements is given in Table 2. First, the table indicates that the assumption that the critical gap for a ramp is independent of the freeway outside lane volume is justifiable. Second, analyzing the table objectively, there is no significant correlation between any single

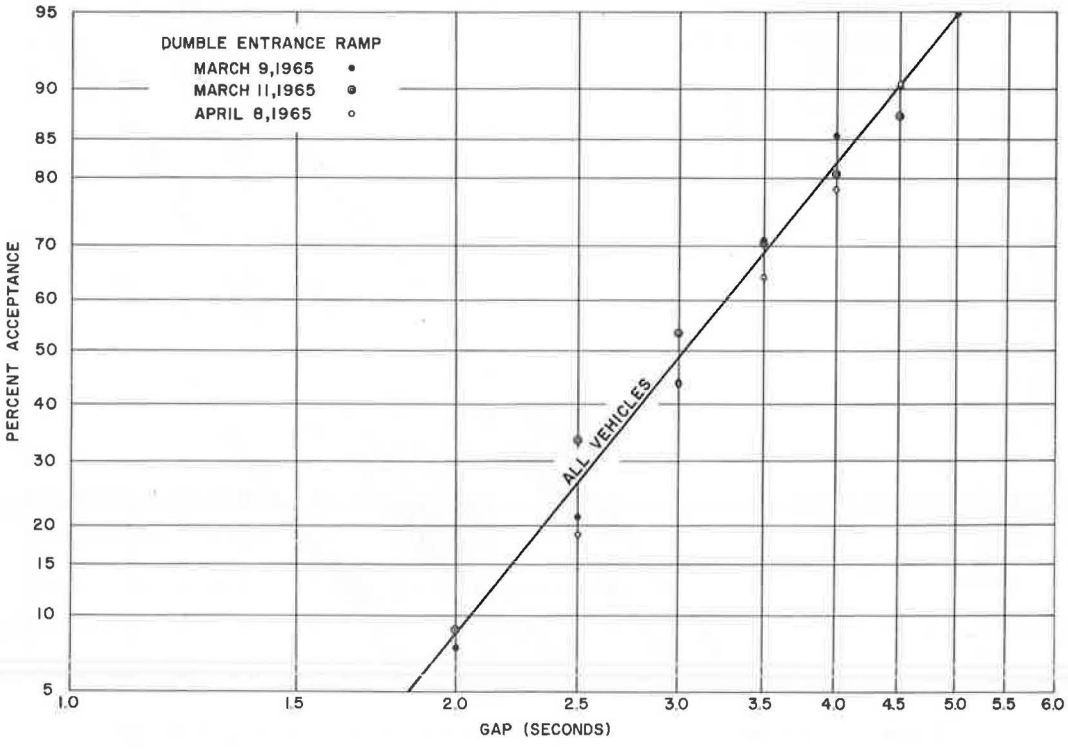


Figure 8. Variations in percent acceptance for merging vehicles at same ramp.

element and the critical gap. Subjectively, however, to the person familiar with the study ramps, it was not surprising that the Cullen, Dumble and Wayside ramps exhibited lower critical gaps than the Griggs, Mossrose and Woodridge ramps. The Griggs and Mossrose ramps have taper-type acceleration lanes which terminate abruptly at overpass structures, whereas operation at the Woodridge ramp is complicated by a high grade differential between the ramp (upgrade) and freeway (downgrade).

To be certain that the variability in the critical gaps between ramps was due to differences in the ramps rather than random fluctuations in the critical gap parameter, the gap acceptance studies were repeated at a single ramp at different times. Figure 8 shows that the critical gap is a stable measure of the operational performance of a given ramp. The parameter, like capacity, is a measure of the ability of a traffic stream to absorb merging vehicles from a given ramp.

In the case of intersections, the side street traffic pattern (i.e., whether the side streets carried one-way or two-way traffic) seemed to affect the critical gap appreciably. Of course, ramps are limited to one-way operation, but ramp metering is a form of operational control which could change the minor street traffic pattern. One objective of this study was to observe some of the characteristics associated with semiautomatic metering to determine equipment requirements for future automatic ramp control systems. This phase of the study was conducted at the Dumble on-ramp. A post-mounted traffic signal with red, amber, and green lenses was installed on the frontage road at the P.C. of the on-ramp. Overhead signals mounted over each lane of the frontage road were employed to separate the two movements (ramp usage and frontage road usage). The signal phasing was designed for bulk-service metering utilizing a three-dial pre-timed controller with a 30-sec cycle length. The three dials were set to give $10\frac{1}{2}$, 8 and $13\frac{1}{2}$ sec of green with a constant amber of $2\frac{1}{2}$ sec, capable of dispatching platoons of 4, 3, and 5 vehicles per cycle.

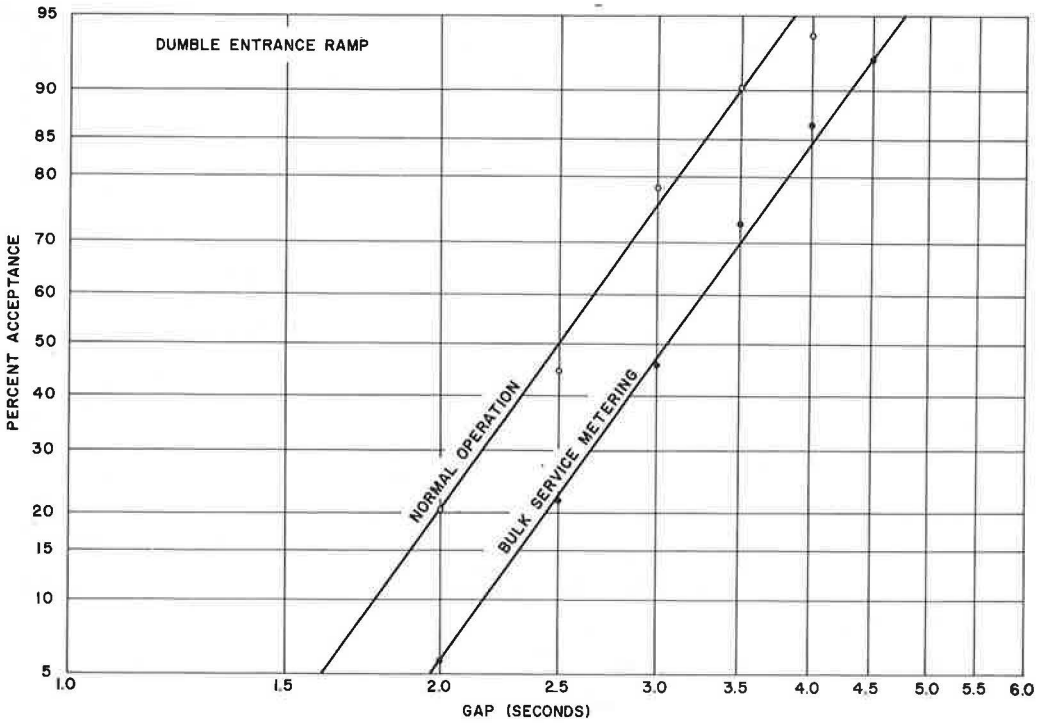


Figure 9. Comparison of percent acceptance for moving merging vehicles during normal and controlled operation.

Figure 9 shows the effect of this technique of ramp metering on the critical gap for moving vehicles. The critical gap for bulk service metering is seen to be 3.1 sec as against 2.5 sec for normal operation. The reason for this is not clear, but it is suspected that (a) metered vehicles have a greater relative speed, and (b) metered drivers are more conscious of the merging maneuver and are therefore more cautious.

Distribution of Critical Gaps

The critical gap for merging indicates how large a time interval is required for the typical ramp driver to enter a freeway. The use of a single typical figure rather than the whole range of observed human behavior has the advantage of simplicity in such advanced stages of the analysis of ramp behavior as delay calculations. In spite of the importance of the critical gap concept, very little consideration has been given to the properties of the parameter. Bissell (9) estimated the standard deviation of the gap acceptance distribution by taking the difference between the median acceptance value and the time corresponding to an acceptance of 15.9 percent. Yet, it is a well-known fact that few traffic phenomena can be described by a single measure of central tendency and dispersion, but can be described by a frequency distribution.

One technique explored in this investigation to obtain a frequency distribution for critical gaps is shown in Figure 10. Only drivers who rejected at least one gap before merging were considered. Assuming that a driver who accepts a given gap size at a certain ramp can be expected to accept any gaps of greater length and that a driver who rejects a given gap size can in a similar manner be expected to reject all gaps of shorter length, it is then evident that a driver's critical gap must lie somewhere in the range between the largest gap he rejected and the gap finally accepted. The number of times the critical gap range included a certain gap interval divided by the total number of gap intervals establishes the observed frequency.

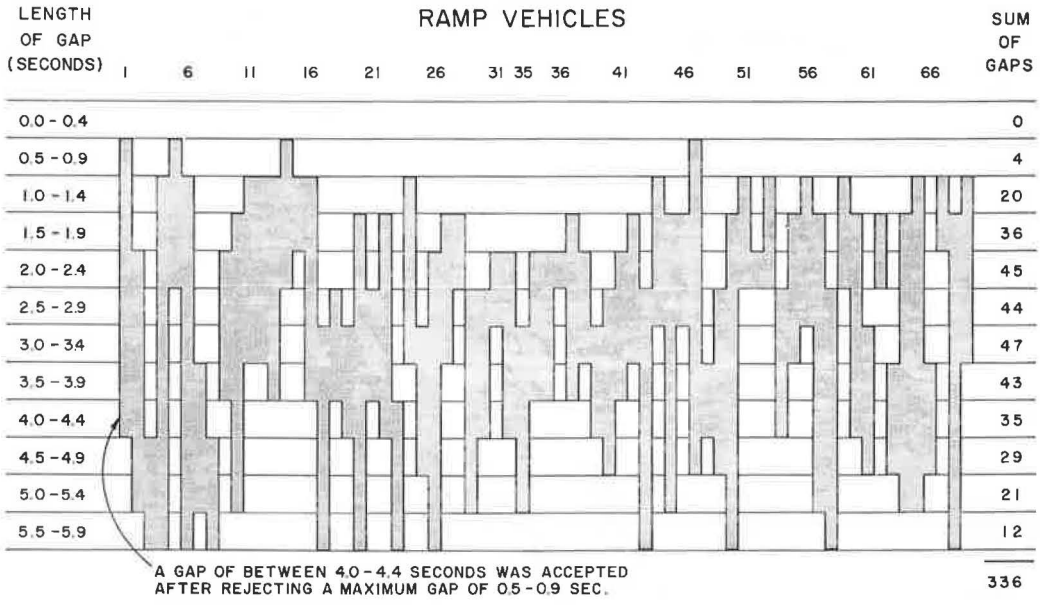


Figure 10. Limits of critical gaps for stopped vehicles at Dumble inbound ramp.

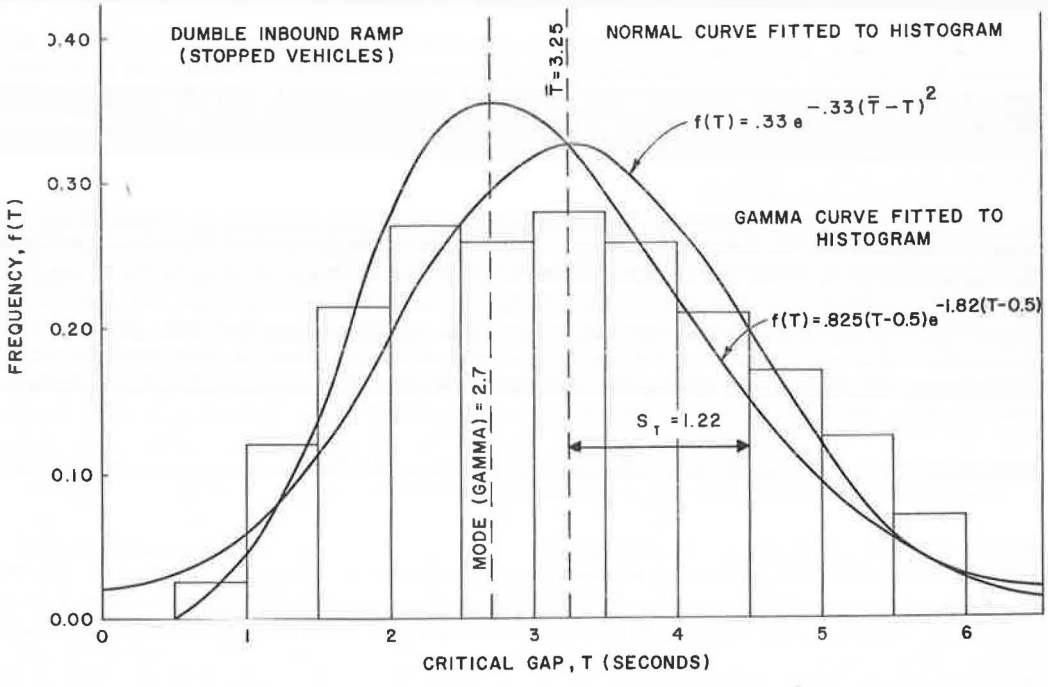


Figure 11. Comparison of observed and theoretical critical gap distributions.

The histogram of observed frequencies obtained from Figure 10 is shown in Figure 11. Since the task of gap selection is primarily a function of the human component of the composite driver-vehicle population, a logical choice in fitting a probability distribution to the observed data is the normal distribution. Since the normal distribution is completely determined by its mean and standard deviation and these quantities are rather accurately estimated from the histogram, a normal curve was easily fitted to the observed data (Fig. 11).

Although there is reasonably good agreement between the normal curve and the data, there are certain theoretical, not to mention practical, considerations which suggest a less restrictive distribution. The principal characteristics of the normal distribution are (a) it is symmetrical, (b) it assigns a finite probability to every finite deviation, and (c) the mode or most probable result is equal to the mean. Of course, possession of these three requisites is by no means general in considering vehicular traffic characteristics. It is unrealistic to think of a driver with a negative critical gap. Yet, if critical gaps were distributed normally, the second property of the distribution would assign a finite probability a negative critical gap.

Among the families of distributions formulated for the purpose of enabling the statistician to deal with a wide variety of data are those developed by Karl Pearson. Pearson (12) noted that the binomial and Poisson distributions satisfy the differential equation

$$\frac{dP}{dx} = \frac{(a + x)\,P}{b + cx + dx^2}$$

for some set of values of the constants a, b, c and d. Certain solutions of this equation have been sorted out, largely because of their algebraic simplicity, to form the Pearson frequency distributions.

Because of the non-negative restriction, we are looking for a curve limited at one end. The Pearson Type III distribution is of this type (see Appendix). Because this distribution uses the gamma function it is also known as the gamma distribution. The element of randomness inherent in this distribution, as exemplified by its formulation from the Poisson distribution, gives it some conceptional appeal for describing the critical gap phenomenon. In Figure 11, the Pearson Type III (Gamma) curve has been superimposed on the histogram along with the normal curve.

The cumulative type curves of Figures 6 to 9 suggest a second procedure for obtaining the desired frequency distribution. A distribution of minimum or critical gaps can be estimated by finding the proportion of drivers who accept a certain sized gap and the

TABLE 3

SUMMARY OF CRITICAL GAP MOMENTS AND PARAMETERS OF THE PEARSON TYPE III DISTRIBUTION APPROXIMATING OBSERVED CRITICAL GAP DISTRIBUTION

Ramp	Critical Gap Parameters (sec)			Pearson Type III Parameters (sec)		
	Median	Mean	Std. Dev.	a	b	c
Cullen	2.7	3.1	0.98	4.5	2.2	1.0
Dumble	2.8	3.2	0.85	6.6	3.0	1.0
Wayside	3.3	3.6	0.94	7.9	3.0	1.0
Griggs	3.7	4.1	1.29	4.1	1.6	1.5
Mossrose	4.7	5.2	1.57	5.5	1.5	1.5
Woodridge	3.8	4.2	1.18	9.7	2.6	0.5

proportion who accept a slightly longer gap. The difference in these proportions then gives an estimate of the proportion of drivers who will accept the slightly longer gap but not the shorter one, i.e., the proportion of drivers whose critical gap is the slightly longer gap. A summary of the critical gap distribution moments obtained using this second method is given in Table 3. The Pearson Type III distribution was fitted to these frequency distributions; the estimates of the parameters of the theoretical distribution are also summarized in Table 3.

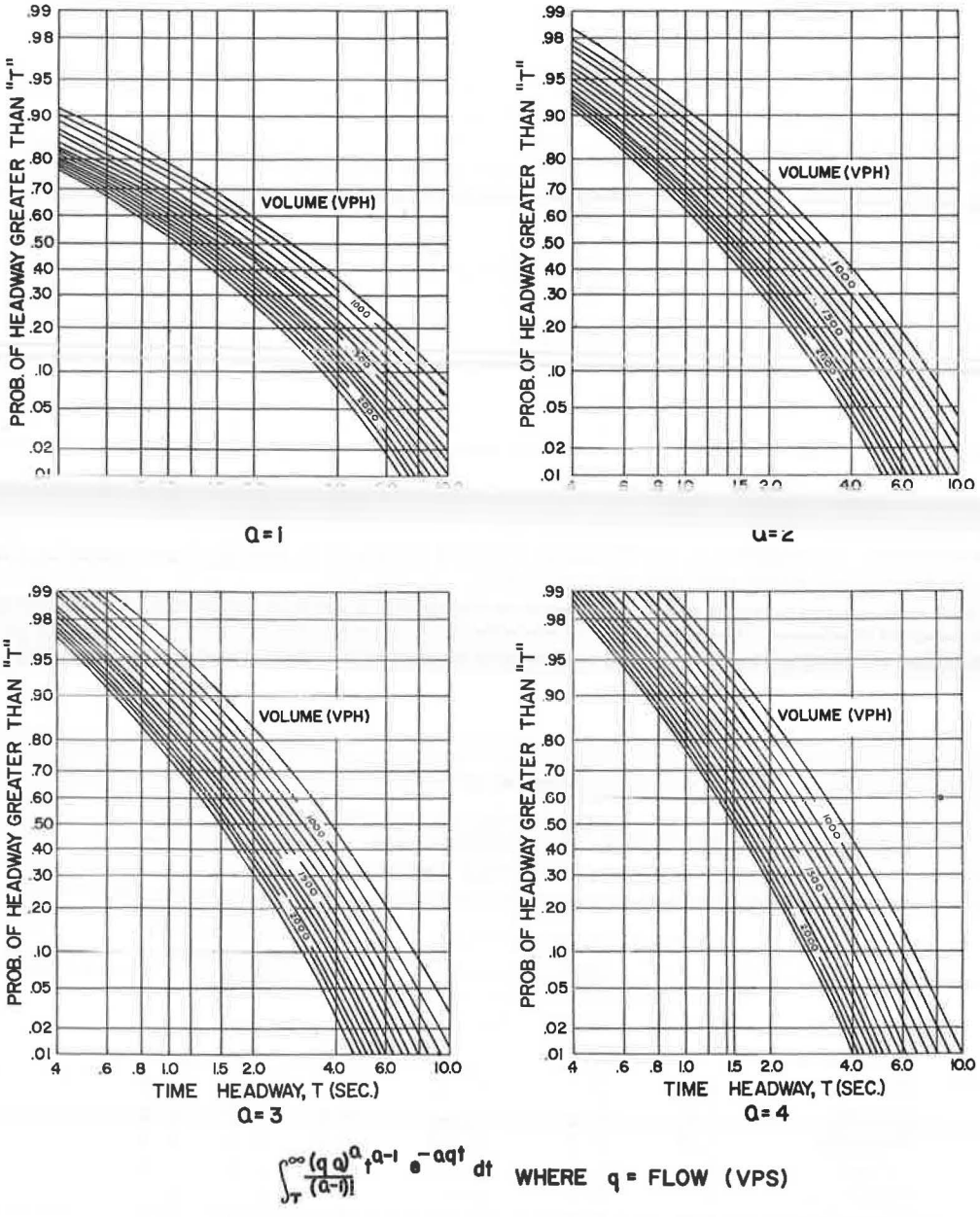


Figure 12. Cumulative probability curves for the Erlang distribution.

There is a subtle yet distinct difference between the frequency distributions obtained by the two methods. The frequency distribution obtained using the method of Figures 10 and 11 enables one to predict the probability of a driver having a certain critical gap. The area under the frequency distribution obtained from the cumulative distribution gives the probability of a driver accepting a certain size gap.

What is desired is a description of the critical gaps among a population of drivers. The mode, mean, and variance of the critical gap distribution add significantly to this description and therefore must be considered in the applications along with the median critical gap, as shall be shown.

DELAY CHARACTERISTICS

Gap Availability

In using mathematics to estimate delay due to merging, it is necessary to have a description of the distribution of time headways or gaps in the outside freeway lane as well as a description of the critical gap.

The Erlang frequency distribution, a special case of the Pearson Type III distribution in which the parameter, a, must be a positive integer, combines many of the advantages of the parent distribution plus simplicity in application. The cumulative form of the Pearson Type III distribution must be evaluated by either numerical methods or tables of the incomplete gamma function unless a is a positive integer, making it difficult to determine the probability of a ramp driver finding a gap larger than the critical gap, for example.

The cumulative frequency curves (Fig. 12) show the theoretical relationship between gap availability and the critical gap, assuming that the distribution of headways on the outside lane of the freeway may be described by an Erlang distribution. Although use of the Erlang distribution affords the opportunity of considering the distribution of freeway vehicles for all cases from randomness (a = 1) to complete uniformity (a = ∞), experience (2) has shown that the first four curves (a = 1, 2, 3, 4) are adequate for the description of vehicular headways in a traffic stream.

Determination of Merging Delays

In the theoretical approach to describing a system composed of two interacting streams, it is necessary to determine the average time for vehicles in the minor stream to merge. It may be assumed that the waiting driver measures each time gap, t, in the traffic on the major highway until he finds an acceptable gap, T, which he believes to be of sufficient length to permit his safe entry. If he accepts the first gap (t >T), his waiting time is zero. If he rejects the first gap (t < T), but accepts the second gap, his expected waiting time would be one interval. If we assume that the driver's gap acceptance policy does not change with time, then by induction, the individual waiting periods form a geometric distribution and the probability, P_n, of any driver having to wait for n intervals each less than T sec before merging is

$$P_n = p^n (1 - p), \quad n = 1, 2, \ldots, \tag{1}$$

where

$$p = P(t<T) = \int_0^T f(t)dt \tag{2}$$

and f(t) is the distribution of gaps in the major stream. The expected number of intervals for which a driver has to wait is given by

$$E(n) = \frac{p}{1-p} = \frac{\int_0^T f(t)dt}{\int_T^\infty f(t)dt} \quad (3)$$

The average time for a ramp vehicle which is in position to merge to find an acceptable gap in the freeway traffic stream will be the product of the expected number of intervals less than T, E(n), and the average length of interval. The average length of interval less than T is, in turn, equal to the total time less than T sec divided by the number of intervals less than T sec:

$$\text{Avg length of intervals } <T = \frac{q\int_0^T t\, f(t)dt}{q\int_0^T f(t)dt} \quad (4)$$

where q is the rate of flow. Multiplying Eqs. 3 and 4 yields the average waiting time for a ramp vehicle in position to merge:

$$d = \frac{\int_0^T t\, f(t)dt}{\int^\infty} \quad (5)$$

Recalling that the proportion of ramp vehicles actually delayed is given by Eq. 2, it is apparent that the average waiting time of those who incur delay is

$$d' = d \Big/ \int_0^T f(t)dt \quad (6)$$

The delays expressed in Eqs. 5 and 6 are for single vehicles approaching the merging area.

Theoretical solutions (13, 14) to the determination of delays for vehicles attempting crossing or merging maneuvers have been based on the assumption that the main stream of traffic is exponentially distributed,

$$f(t) = qe^{-qt} \quad (7)$$

Substituting Eq. 7 in Eq. 5, the mean delay d for a critical gap T and freeway lane flow q is

$$d = q^{-1}(e^{qT} - qT - 1) \quad (8)$$

If the more general Erlang distribution is used to describe the distribution of main stream gaps, then

$$f(t) = \frac{(aq)^a}{(a-1)!} t^{a-1} e^{-aqt}, \quad (a = 1, 2, \ldots) \tag{9}$$

and the mean delay for a = 2, 3 and 4 is

$$d_{a=2} = \frac{e^{2qT} - 2(qT)^2 - 2qT - 1}{q(2qT + 1)} \tag{10}$$

$$d_{a=3} = \frac{e^{3qT} - 4.5\,(qT)^3 - 4.5\,(qT)^2 - 3qT - 1}{q[4.5\,(qT)^2 + 3qT + 1]} \tag{11}$$

$$d_{a=4} = \frac{e^{4qT} - 10.67\,(qT)^4 - 10.67\,(qT)^3 - (8qT)^2 - 4qT - 1}{q[10.67\,(qT)^3 + 8(qT)^2 + 4qT + 1]} \tag{12}$$

Since the negative exponential is a special case of the Erlang distribution, $d_{a=1}$ is given by Eq. 8. Eqs. 8, 10, 11 and 12 are plotted in Figure 13.

It must be remembered that the theoretical delay values are based on a fixed critical gap for all drivers. Blunden (11) correctly suggests that a more complete description of delays to traffic would be given if the fixed critical gap were replaced with a distribution of critical gaps, g(T). If it is assumed that the critical gap distribution is of the gamma type

$$g(T) = \frac{b^a}{(a-1)!} T^{a-1} e^{-bT} \tag{13}$$

a reasonable assumption as seen by Figure 11, then the mean delay D is readily obtainable from

$$D = \int_0^\infty d(T)\ g(T)dT \tag{14}$$

Substituting the delay function given by Eq. 8 and the critical gap distribution given by Eq. 13 in Eq. 14 yields

$$D = q^{-1} \int_0^\infty e^{qT} g(T)dT - \int_0^\infty Tg(T)dT - q^{-1} \int_0^\infty g(T)dT \tag{15}$$

Realizing that the second term is the form for the mean of a distribution and the integral in the last term must equal unity, gives

$$D = q^{-1} \int_0^\infty e^{qT} g(T)dT - \bar{T} - q^{-1} \tag{16}$$

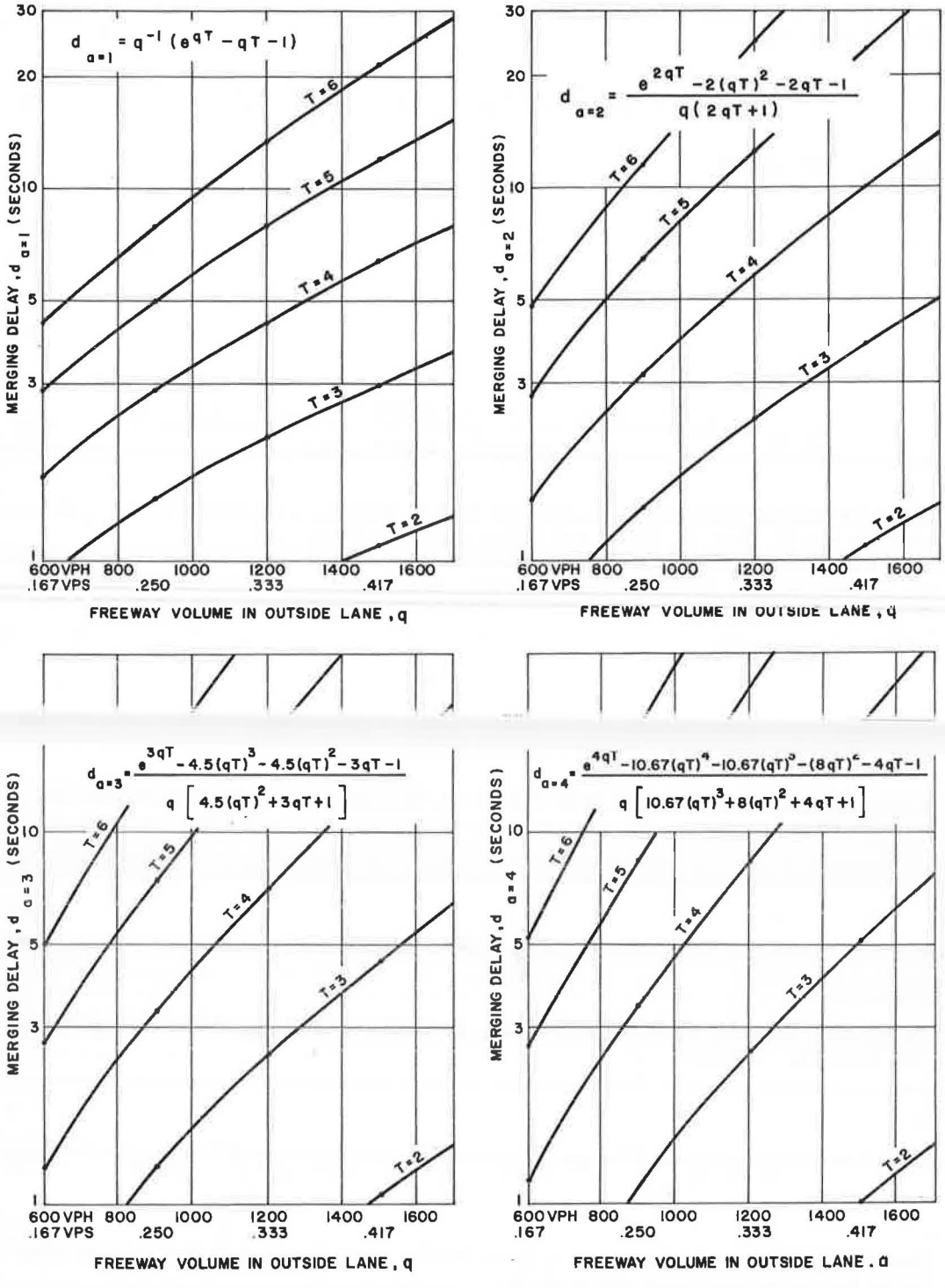


Figure 13. Merging delay in terms of the freeway flow q, critical gap T, and Erlang constant, a.

Letting b = a/T, the first term may be put in the form

$$\frac{(a/\bar{T})^a}{q\left(\frac{a-q\bar{T}}{\bar{T}}\right)^a} \int_0^{\infty} \left(\frac{a-q\bar{T}}{\bar{T}}\right)^a \frac{T^{a-1}}{(a-1)!}\, e^{-(a-q\bar{T})\, T/\bar{T}}\, dT \tag{17}$$

$$= \frac{1}{q}\left(\frac{a}{a-q\bar{T}}\right)^a \tag{18}$$

provided that $a > q\bar{T}$, since the integral represents the area under a gamma distribution with parameters a and $(a - q\bar{T})/\bar{T}$. Gathering terms, the general expression for delay to a merging vehicle is

$$D = \frac{1}{q}\left(\frac{a}{a-q\bar{T}}\right)^a - \bar{T} - \frac{1}{q} \tag{19}$$

Equation 8 can be verified as the special case of Eq. 19 in which the critical gap is taken as a fixed value equal to the mean critical gap. For $a = \infty$, we know that the variance of the gamma distribution is zero; so we may interpret $g(T; u, \infty)$ as the

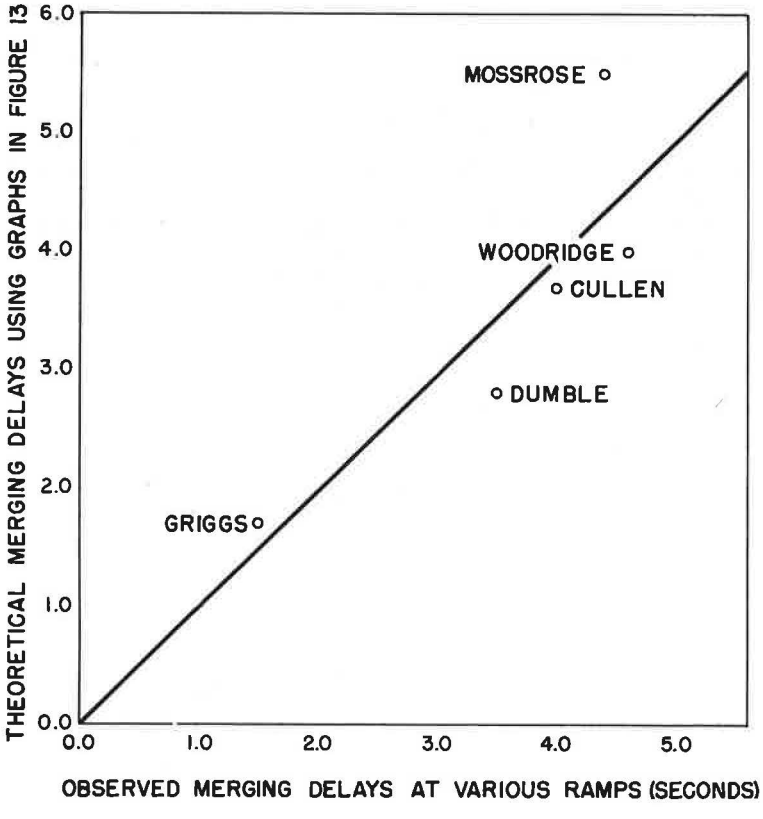


Figure 14. Comparison of observed and theoretical merging delays at study ramps.

family of situations for which the critical gap is constant and has the value $\bar{T}$. From the definition of e it is apparent that

$$\lim_{a\to\infty} \left(1 - \frac{q\bar{T}}{a}\right)^{-a} = e^{q\bar{T}} \tag{20}$$

Substitution of Eq. 20 in Eq. 19 gives Eq. 8.

Delay values calculated from Eq. 19 using finite values of a will always be higher than those obtained from Eq. 8 assuming a fixed critical gap. Moreover, the fact that the delay obtained using the fixed value is the special case of the delay obtained using the distribution for a variance of zero proves that the mean of the critical gap distribution, not the median, should be used in delay computations.

Observed average merging delays were compared to the average merging delays predicted by the equations, using the mean critical gap (Fig. 14). The value of a for the Erlang distribution used to enter the graphs in Figure 13 was calculated from the time headways in the outside lane using the method of moments in which a is the integer closest to the quotient obtained by dividing the mean headway squared by the variance in the headways. The merging delay concepts discussed in this section become important as inputs in a queuing model developed in the next section.

Queuing Considerations

A queuing system is composed of three elements: (a) a demand or flow of arrival requiring service, (b) some restriction on the availability of service, and (c) irregularity in either the demand or in the servicing operation or in both.

Consider vehicles on an entrance ramp arriving at the merging area at a rate described by f (q_r) where q_r is the average rate of arrivals. These vehicles are obliged to yield to the freeway traffic, thus forming a single lane waiting for successive vehicles at the head of the line to merge. If the distribution of time spent by vehicles at

dall's (15) approach for a queuing system with random inputs and arbitrary service times, formulas are developed for the mean queue length and mean waiting time for a queue of ramp vehicles waiting to merge.

Let n_0, n_1 be the ramp queue lengths immediately after two successive ramp vehicles C_0, C_1 have merged, t be the service time of C_1, r be the number of ramp vehicles arriving while C_1 is being served. If a random variable δ is introduced such that $\delta = 1$ if $n_0 = 0$ and $\delta = 0$ if $n_0 \neq 0$, then it follows that

$$n_1 = n_0 + r - 1 + \delta \tag{21}$$

The definition of δ shows that

$$\delta^2 = \delta$$

and

$$n_0\delta = 0,$$

and hence, from Eq. 21, on taking expected values,

$$E(n_1) = E(n_0) + E(r) - 1 + E(\delta). \tag{22}$$

If the system is assumed to be in a state of statistical equilibrium, $E(n_1) = E(n_0)$ and

$$E(r) = q_r E(t) = q_r/\mu = \rho$$

Thus substituting in Eq. 22

$$E(\delta) = 1 - \rho$$

Squaring both sides of Eq. 21 and taking expected values as before,

$$E\,(r-1)^2 + E(\delta^2) + 2E\,[n_0\,(r-1)] + 2E\,[\delta\,(r-1)] = 0$$

which reduces to

$$E(n_0) = \rho + \frac{E(r^2) - \rho}{2(1-\rho)} \tag{23}$$

It is now necessary to calculate $E\,(r^2)$, the second moment of the number of arrivals in the service time T, making use of its relationship to the mean and variance in arrivals. Assuming that ramp arrivals are Poisson and remembering that averaging here must be carried out with respect to both r and the service time t,

$$E(r^2) = q_r\,E(t) + q_r^2\,E\,(t^2)$$

Since $E(t) = \mu^{-1}$ and

$$E\,(t^2) = \sigma^2 + E\,(t)^2$$

then

$$E\,(r^2) = \rho + \rho^2 + q_r^2\,\sigma^2 \tag{24}$$

Substituting Eq. 24 in Eq. 23 gives the expected queue length on the ramp as

$$E\,(n)_r = \frac{q_r}{\mu} + \frac{\mu q_r^2\,(\mu^{-2} + \sigma^2)}{2(\mu - q_r)} \tag{25}$$

If w is the waiting time (before merging) of C_1, then n_1 ramp vehicles arrive in time t + w. Thus since the mean arrival rate is q_r,

$$E(n)_r = q_r\,E\,(t + w)$$

TABLE 4

SUMMARY OF SERVICE TIME MOMENTS AND PEARSON TYPE III PARAMETER "a"

Ramp	Service Time Moments		Type III "a" Value
	Mean	Std. Dev.	
Cullen	4.0	5.2	0.6
Dumble	3.5	5.5	0.4
Wayside	6.2	8.8	0.5
Griggs	1.5	2.8	0.3
Mossrose	4.4	6.2	0.5
Woodridge	4.6	6.0	0.6

It follows that the average waiting time for a ramp vehicle before merging is

$$E(w) = [E(n)_r/q_r] - \mu^{-1} \tag{26}$$

and the mean wait in the system for a ramp vehicle is

$$E(v) = E(n)_r/q_r \tag{27}$$

Major and Buckley (16) have interpreted the service time for the queue μ^{-1} as identical to the summation of the rejected gaps for a ramp vehicle in position to merge (what we have referred to as the merging delay d in this paper). The Pearson Type III distribution was fitted to the observed distribution of delays at each ramp and the results are given in Table 4 for all ramps in the study. Based on the assumptions that the average service time is equal to the average merging delay and the distribution of service times is of the form of the Pearson Type III distribution, we obtain from Eq. 25

$$E(n)_r = dq_r + \frac{(dq_r)^2 (1 + a^{-1})}{2(1 - dq_r)} \tag{28}$$

where d is given by Eq. 8, 10, 11 or 12 depending on the distribution of freeway headways and a is the parameter describing the distribution of service times.

If the a value in Eq. 28 calculated in Table 4 is rounded off to unity, Eqs. 25-28 take the form of the conventional, Poisson-negative exponential, queuing formulas.

APPLICATIONS

Ramp Control

Almost any engineering problem may be described as a systematic attempt to re-

build enough strength into the materials of a layered pavement system, for example, to withstand shear stresses due to anticipated loads. However, the mere fact that the strength (capacity) exceeds the load stresses (demand) does not guarantee an acceptable level of service. The deflection, smoothness, texture and color contrast also affect the driver's ride and, as such, are level of service factors that must be considered.

Traffic engineering is the science of measuring traffic characteristics and the application of this information to the design and operation of traffic systems. The traffic engineer's basic problem of resolving a capacity-demand relationship is similar to that of any other engineer: he must be able either to measure the parameters defining capacity and demand very accurately, or to control them. Returning to the pavement design analogy: although the strength of the materials in a pavement can probably not be estimated as accurately as the capacity of a freeway lane, the pavement designer knows that the loads (demand) on the facility are controlled, and in most states limited by law. If urban freeways are to operate at the levels of service for which they were designed, the demand on these facilities must also be regulated.

Because the control of vehicles entering the freeway, as against the control of vehicles already on the freeway, offers a more positive means of preventing congestion, considerable emphasis is being placed on the technique of ramp metering. Any scheme for controlling the rate of flow of vehicles from an entrance ramp onto the freeway should be based on, and be capable of, reacting to operation on the freeway lanes. Chicago (17) utilizes an occupancy (density) measurement in the middle freeway lane in advance of the merging area. Wattleworth (18) proposes a systems approach in which the capacity of a freeway system is limited to the section of smallest capacity (bottleneck). If the bottleneck capacity is known, all ramps in the system can be metered according to the differential between the upstream freeway demand and bottleneck capacity.

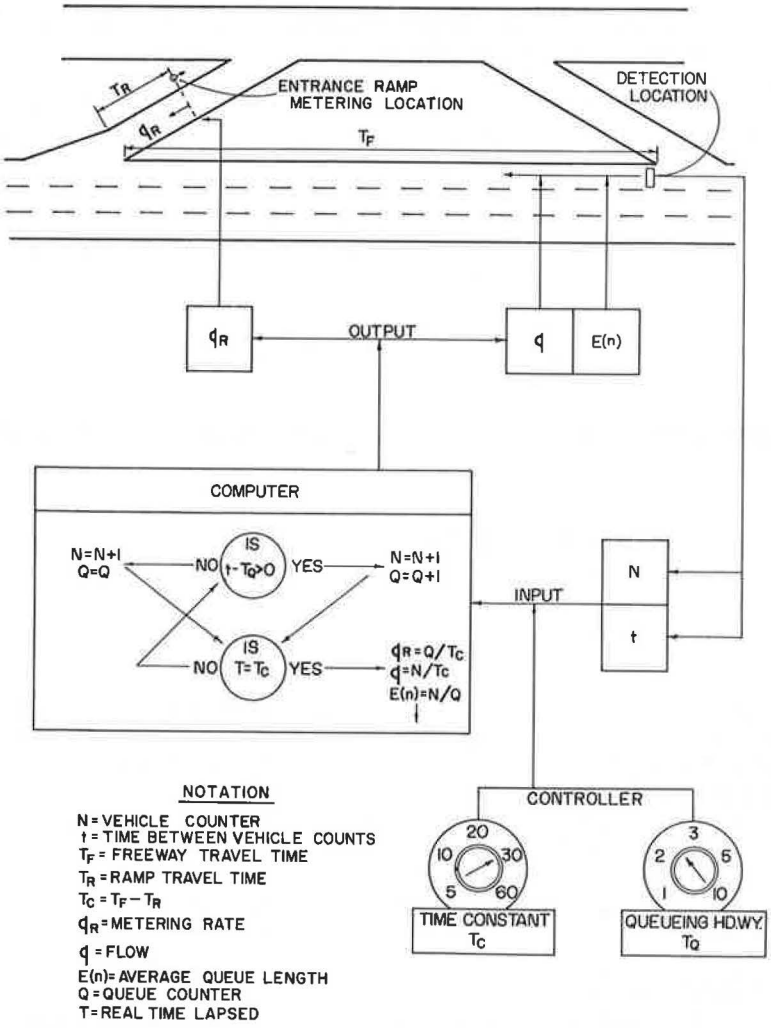


Figure 15. Application of queuing model to freeway control.

The key to entrance ramp operation is, however, the synthesis of two lanes of traffic into a single lane. For this reason, it is believed that a microscopic approach sensitive to the operating conditions on the outside freeway lane should be considered. Drew (2) describes a "moving queues" model based on coordinating ramp metering with the detection of acceptable gaps in the outside freeway lane. Figure 15 shows the application of this model to the control of a merging situation at an entrance ramp. Thus, a vehicle in the outside lane of the freeway with a time headway less than the arbitrary queuing headway T_Q is considered to be queued to the preceding vehicle. The control system illustrated consists of the flow of information from a detector located on the outside freeway lane to a computer and then to the metering signal on the ramp. For the closed loop system pictured, either a digital or analog computing device could be utilized. However, use of the former would necessitate a reduction in the "time constant" (time over which traffic conditions are averaged) by an interval equal to the time necessary for computation.

In considering an example, suppose the travel time during the peak period from detector to merging area is T_F = 35 sec and from the metering station to the merging

area is T_R = 5 sec. The critical gap (that headway in the outside freeway lane for which an equal percentage of ramp traffic will accept a smaller headway as will reject a larger one) is assumed to be 2.5 sec. If control adjustment is to be made during the same period as detection, the time constant T_c cannot be greater than $T_F - T_R$. Moreover, if the arbitrary queuing headway T_Q is equated to the critical gap, the number of platoons or moving queues Q will equal the number of critical gaps. The latter determines q_r, the number of ramp vehicles that can merge during T_c. Thus, in the example the dials on the controller would be set to T_c = 30 sec and T_Q = 2.5 sec. If during T_c, N = 10 vehicles were detected and Q = 5 of the headways were greater than the queuing headway on the dial ($t > T_Q$), the metering rate during T_c would be:

$$q_r = \frac{Q}{T_c} = \frac{5 \text{ veh}}{30 \text{ sec}} = 1 \text{ veh every } 6 \text{ sec}$$

The rate of flow and congestion index at the detection station during T_c would be:

$$q = \frac{N}{T_c} = \frac{10 \text{ veh}}{30 \text{ sec}} = 1 \text{ veh every } 3 \text{ sec}$$

$$E(n) = \frac{N}{Q} = \frac{10 \text{ veh}}{5 \text{ queues}} = 2 \text{ veh/queue}$$

The variability of the critical gap among entrance ramps (Fig. 7) shows that controller settings would vary from entrance ramp to entrance ramp. The geometrics of the freeway would also affect the detector location and hence the time constant.

The significance of the control parameter E(n) is twofold: (a) by definition, it is

approach infinity); and (b) the moving queue length, E(n), was formulated in such a way that it is the reciprocal of the probability of getting a gap larger than the queuing criteria, and therefore is actually a measure of gap availability.

Comparison Ramp Metering Criteria

An important contribution of this microscopic approach to ramp control is that the individuality of entrance ramps is taken into consideration. Figure 16 shows the relationship between the freeway outside lane volume, q, the headway distribution of the freeway volume, a, the critical gap, T, and the ramp volume, q_r. Consider an entrance ramp operating with a critical gap of 4.0 sec and with the distribution of freeway traffic conforming to an Erlang distribution with a = 2. The sum of the coordinates of any point on the line T = 4 in the graph a = 2 describes the merging service volume for that ramp. For example, the point described by q = 1,500 and q_r = 120 tells us that the merging service volume is 1,620, and that under these operating characteristics, a ramp arrival has a 67 percent chance of finding the ramp empty or a 33 percent chance of finding a vehicle ahead of it trying to merge.

The effect of poor ramp geometrics is evident. If in the previous example the critical gap was 5.0 sec, the maximum ramp service volume q_r for the same freeway volume (q = 1,500) would be 50 vph or a total merging service volume for the ramp of 1,550 vph.

The graphs also point out the differences in the macroscopic philosophies of ramp metering (17, 18) and the microscopic approach explained in this paper. In the macroscopic approach, metering would be based on one of the curved lines (one representing the boundary between stable and unstable flow and the other representing possible capacity) regardless of the ramp geometrics or critical gap. This means that, for all conditions except those described as unstable flow on the graph, vehicles would be

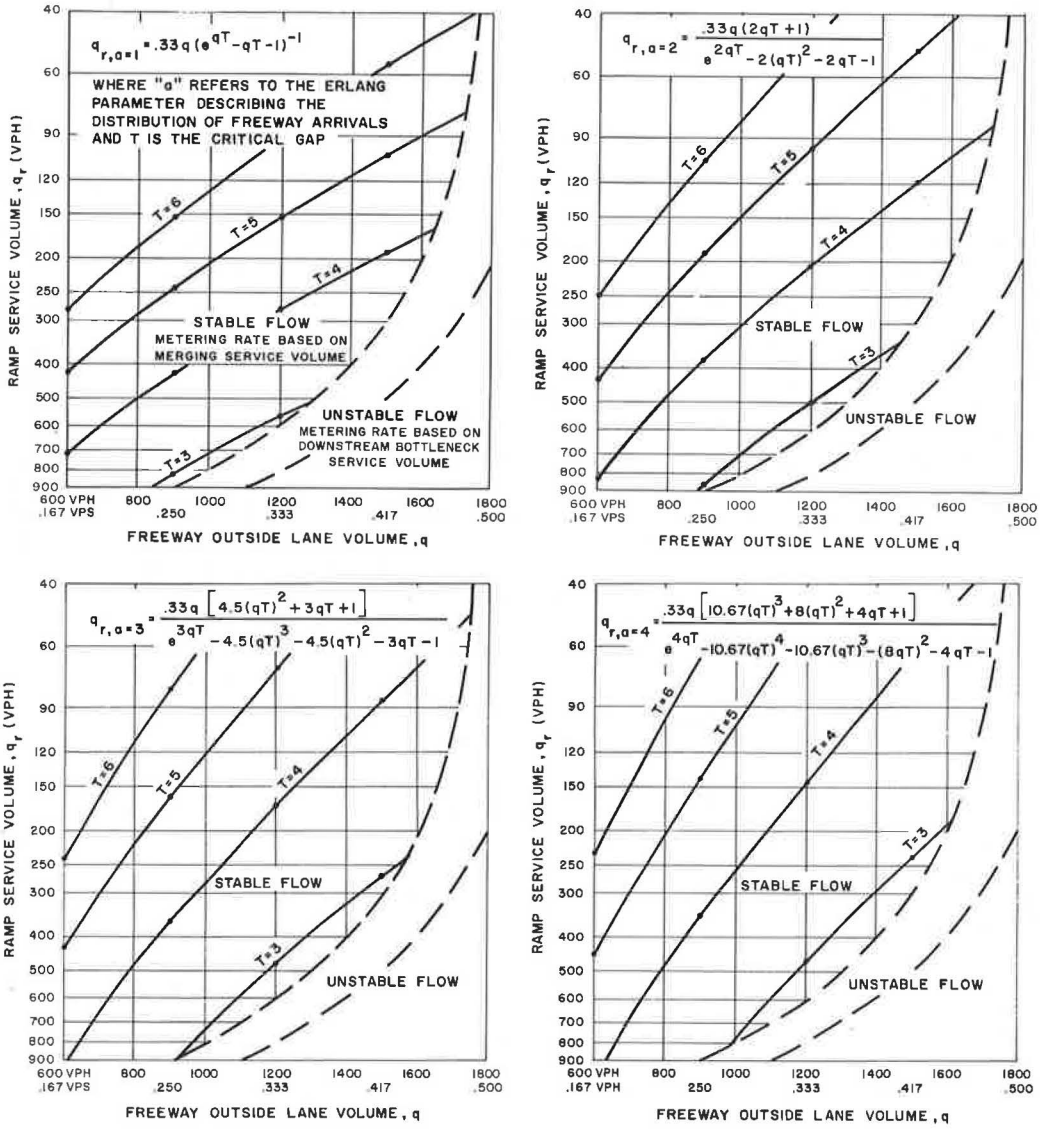


Figure 16. Maximum service volume q_r for a ramp arrival to have a probability of 0.67 of finding no ramp vehicles in merging area.

metered at a faster rate than the service rate at the merging area (available critical gaps) encouraging drivers to either accept smaller gaps than the critical gap or become part of a steadily growing queue at the merging area.

Figure 16 shows the need for a ramp control technique combining both the macroscopic and microscopic approach. For conditions described by "ustable flow," the ramp geometrics do not govern and hence the macroscopic approach based on the downstream bottleneck service volume applies. However, to the left of the 1,800-vph line dividing stable and unstable flow, the critical gap governs since the merging service volume is less than the bottleneck service volume. (It should be explained that the term "service volume" is used here consistent with the level of service concept proposed in the 1965 Highway Capacity Manual, although the word capacity could be substituted in this particular comparison with only the risk in a loss of generality.)

Ramp Design

After the determination of the number of freeway lanes in a rational design procedure, the operating conditions at critical locations of the freeway must be investigated for the effect on capacity and level of service. Unless some designated level of service is met at every point on the freeway, bottlenecks will occur and traffic operation will break down. Critical locations on a freeway are manifest by either sudden increases in traffic demand, the creation of intervehicular conflicts within the traffic stream, or a combination of both.

Entrance ramps represent the third and most serious case since they create two potential conflicts with the maintenance of the adopted level of service of a roadway section. First, the additional ramp traffic may cause operational changes in the outside lane at the merge. This condition, of course, will be aggravated by any adverse geometrics, such as high angle of entry, steep grades, and poor sight distance. Second, the additional ramp volume may change the operating conditions across the entire roadway downstream from the on-ramp. This is particularly true where there is a downstream bottleneck.

The three basic procedures employed in checking capacity for the design of entrance ramps are based on the same philosophies discussed for metering ramps. One method is based on preventing the total freeway volume upstream from the ramp plus the entrance ramp volume from exceeding the capacity of a downstream bottleneck. A second method (19) takes into consideration the distribution of freeway volumes per lane and then limits the ramp volume to the merging capacity less the upstream volume in the outside lane. The third method states that the ramp capacity is limited by the number of gaps in the shoulder lane which are greater than the critical gap for acceptance.

Figure 16 can be useful in the implementation of all three approaches. Thus, if a ramp on a new facility is of a high-type geometric design guaranteeing a low critical gap, methods 1 and 2 are applicable since the merging service volume will exceed any bottleneck service volume. If, however, due to the terrain, spacing of interchanges

Freeway design, like most real world phenomena, involves a series of compromises. Because of the spacing of interchanges on many urban freeways, the fulfillment of desirable entrance ramp design, desirable exit ramp design, and an adequate weaving section between them, offers a dilemma. The alternatives are (a) reduction in the standards of one or more of the features (b) elimination of one of the features (such as one of the ramps), or (c) transferring the weaving from the freeway to the frontage road. These alternatives should be evaluated in terms of their cost, their effect on the freeway and ramp operation at that location, and their effect on adjacent facilities such as adjacent interchanges, and cross-street signalization. The procedure exemplified by Figure 16 enables a designer to evaluate alternatives more rationally and if compromise is needed, to select the element or location where it will be the least objectionable.

SUMMARY AND CONCLUSIONS

Many researchers (4, 5, 9, 10) have defined a parameter called the critical gap to describe the operation of two interacting traffic streams. The critical gap, as a time gap just as likely to be accepted as to be rejected, is by implication a median gap. The advantages of this parameter for describing entrance ramp operation are (a) it provides a single typical figure rather than the whole range of observed gap acceptance characteristics, (b) it gives simplicity in such advanced stages of ramp behavior as delay calculations, (c) it is sensitive to variations in geometrics and, therefore, provides a rational means of evaluating ramp design, and (d) like such parameters as capacity, optimum speed and optimum concentration, it is independent of traffic volumes and is, therefore, invariant for a given location under a given set of environmental and operating conditions. These concepts are verified in this report.

Of particular significance is the development of a distribution of critical gaps which enables a researcher to predict the probability of a driver having a certain critical

gap. Delay values calculated using this distribution (Eq. 19) will always be higher than those calculated assuming that all drivers have the same fixed critical gap (Eq. 8). The fact that the delay obtained using the fixed value is the special case of the delay obtained using the distribution for a variance of zero proves that the mean of the critical gap distribution, not the median, should be used in delay computations.

The delay equations (Eqs. 8, 10, 11, 12) developed in this study are based on the assumption that the distribution of freeway headways is of the form of the Erlang frequency distribution. This represents a generalization of past work (5, 11, 13, 14, 16) in which the distribution of freeway headways was assumed to be of that particular form of the Erlang distribution called the negative exponential distribution.

The distribution of time spent by merging ramp vehicles at the head of the queue may be approximated by a gamma distribution, suggesting that an entrance ramp merging operation may be considered within the context of classical queuing theory. Based on this queuing model the ramp metering curves of Figure 16 were conceived. The metering technique suggested combines a macroscopic and a microscopic approach, the former based on a controlling downstream bottleneck, the latter on the merging area.

In conclusion, it is apparent that additional theoretical research is needed to provide a more complete relationship between the many variables associated with the interaction of vehicles traversing a ramp and merging into a freeway so as to determine the effects of traffic characteristics, merging area geometrics, and environmental elements on merging capacity and operation. This information would be of profound significance in the operation and control of existing facilities, the design of future facilities, and the development of usable distributions of traffic variables for simulation programs.

REFERENCES

1. Drew, D. R. Deterministic Aspects of Freeway Operations and Control. Res. Rept. 24-4, Texas Transp. Inst., Texas A&M Univ.
2. Drew, D. R. Stochastic Considerations in Freeway Operations and Control. Res. Rept. 24-5, Texas Transp. Inst., Texas A&M Univ.
3. Wattleworth, J. A. System Demand-Capacity Analysis of the Inbound Gulf Freeway. Res. Rept. 24-8, Texas Transp. Inst., Texas A&M Univ.
4. Greenshields, B. D., Shapiro, D., and Erickson, E. L. Traffic Performance at Urban Street Intersections. Tech. Rept. No. 1, Bur. of Highway Traffic, Yale Univ., 1947.
5. Raff, M. S., and Hart, J. W. A Volume Warrant for Urban Stop Signs. Eno Found. for Highway Traffic Control, Saugatuck, Conn., 1950.
6. Glickstein, A., Findley, L. D., and Levy, S. L. Application of Computer Simulation Techniques to Interchange Design Problems. HRB Bull. 291, pp. 139-162, 1961.
7. Haight, F. A. Mathematical Theories of Traffic Flow. Academic Press, New York, 1963.
8. Haight, F. A., Bisbee, E. R., and Wojcik, D. Some Mathematical Aspects of the Problem of Merging. HRB Bull. 356, pp. 1-14, 1962.
9. Bissell, H. H. Traffic Gap Acceptance from a Stop Sign. Graduate Res. Rept. (unpubl.) ITTE, Univ. of California, Berkeley.
10. Solberg, P., and Oppenlander, J. C. Lag and Gap Acceptances at Stop-Controlled Intersections. Highway Research Record 118, pp. 48-67, 1966.
11. Blunden, W. R., Clissold, C. M., and Fisher, R. B. Distribution of Acceptance Gaps for Crossing and Turning Maneuvers. Australian Road Res. Board Proc., Vol. 1, Pt. 1, 1962.
12. Fry, T. C. Probability and Its Engineering Uses. D. Van Nostrand Co., Inc., New York, 1928.
13. Adams, W. F. Road Traffic Considered as a Random Series. Jour. Inst. Civ. Engrs., p. 121, Nov. 1936.
14. Tanner, J. C. The Delay to Pedestrians Crossing a Road. Biometrika, Vol. 38, Nos. 3 and 4, Dec. 1951.

15. Kendall, D. G. Some Problems in the Theory of Queues. Jour. Royal Statist. Soc., Series B, Vol. 13, No. 151, 1951.
16. Major, N. G., and Buckley, D. S. Entry to a Traffic Stream. Australian Road Res. Board Proc., Vol. 1, Pt. 1, 1962.
17. May, A. D. Experimentation with Manual and Automatic Ramp Control. Highway Research Record 59, pp. 9-38, 1964.
18. Wattleworth, J. A. Peak Period Control of a Freeway System—Some Theoretical Considerations. Doctoral Dissert., Northwestern Univ., Aug. 1963.
19. Drew, D. R., and Keese, C. J. Freeway Level of Service as Influenced by Volume and Capacity Characteristics. Highway Research Record 99, pp. 1-39, 1965.

Appendix

THE PEARSON TYPE III DISTRIBUTION

A random variably y is said to be distributed as the Type III distribution if its density is

$$f(y) = \frac{b^a}{\gamma(a)} y^{a-1} e^{-by} ; \quad 0 < y < \infty \tag{29}$$

where γ (a) is called the gamma function and is defined by the formula

$$\gamma(a) = \int_0 \quad z^{--} \; e^{-} dz$$

To apply the Type III distribution to traffic phenomena such as space headways in which the distribution curve does not go through the origin (the space headway between successive vehicles in the same lane can never be zero because vehicles possess length), it is necessary to translate the distribution c units from the origin.

Recognizing that the area under f(y) is unity, we obtain the desired generalized distribution by substituting x = y + c for y in Eq. 29:

$$\int_0^\infty f(y)dy = \int_c^\infty \frac{b^a}{\gamma(a)} (x-c)^{a-1} e^{-b(x-c)}$$

giving

$$f(x) = \frac{b^a}{\gamma(a)} (x-c)^{a-1} e^{-b(x-c)} ; \quad c < x < \infty \tag{30}$$

Taking moments about c,

$$\mu_k^c = \frac{b^a}{\gamma(a)} \int_c^\infty (x-c)^{k+a-1} e^{-b(x-c)} dx$$

$$= \frac{b^a}{\gamma(a)} \frac{\gamma(k+a)}{b^{k+a}} \int_c^\infty \frac{b^{k+a}}{\gamma(k+a)} (x-c)^{k+a-1} e^{-b(x-c)} dx$$

$$= \frac{b^a}{\gamma(a)} \frac{\gamma(k+a)}{b^{k+a}} \tag{31}$$

since the integral represents the area under a Type III curve and must equal unity. The first moment about c_1 is obtained by letting k = 1 in Eq. 31

$$\mu_1^c = \frac{a}{b} \tag{32}$$

Adding c to Eq. 32 gives the mean of this distribution

$$\mu = \frac{a}{b} + c \tag{33}$$

The variance is given by

$$\sigma^2 = \mu_2^c - (\mu_1^c)^2 = \frac{a}{b^2} \tag{34}$$

Discussion

PAUL D. CRIBBINS, Associate Professor of Civil Engineering, North Carolina State University, Raleigh—The author is to be commended for his careful investigation of gap acceptance characteristics which comprises a welcome and valuable addition to the literature on capacity-demand relationships on urban freeways. In particular, consideration of the entrance ramp merging operation within the context of classical queuing theory signifies an important step in the evolution of theoretical research related to the ramp-freeway merging maneuver.

This research endeavor, however, would appear to reinforce the implication that a formidable communications barrier exists between the traffic flow theorist and the practitioner faced with problems requiring immediate solutions. An appreciable gap in understanding between these two groups is apparent and efforts must be made by responsible transportation engineers if we are to maximize the utilization of flow theory findings in the real world of traffic operations. In order to narrow the gap between the theorist and the practicing engineer a better appreciation by the flow theorist of the day-to-day problems of the traffic engineer and a greater effort on the part of the latter to obtain and study theoretical findings related to his particular problem are of paramount importance.

Particular consideration has been given in this paper to the design and conduct of an experimental study to validate theoretical models that will eventually be used in an automatic surveillance and control system for the Gulf Freeway in Houston, Texas. Without major design revisions to the freeway, the surveillance and control system ideally will permit an acceptable level of service to be maintained on the freeway during peak demand periods.

Previous operational studies on other freeways have indicated a need for some form of control of the vehicles entering the freeway from the on-ramps during peak demand periods. The first obvious solution to the ramp control problem was complete closure of critical on-ramps, which was often followed by almost complete stagnation of traffic

on the streets surrounding the freeway. Probably the next significant step in the evolution of ramp control techniques was made at the Illinois Expressway Surveillance Project in Chicago where a traffic-adjusted metering scheme was initiated in 1963. If we assume, and I think we must, that the primary objective of automatic ramp metering is to maximize flow on the freeway or expressway at the expense of flow on the surrounding street system, then the initial experience on the Illinois project was very satisfactory. However, recent observation by this writer of peak-period conditions at the demonstration area of the Illinois project indicated that congestion on the arterial streets leading to the on-ramps may become intolerable when the ramp volumes are restricted for prolonged periods.

While it may be argued that such a condition, even though undesirable, is inevitable, there is growing evidence that research in metering and controlling ramp traffic must encompass a comprehensive approach to the total capacity-demand of the freeway and the existing street network treated as a single system. Such an undertaking, even though it involves the expenditure of considerable money and energy, has tremendous potential, and research personnel on the Gulf Freeway Surveillance Project have wisely addressed themselves to the problem in this manner. In doing so they have become aware, and hopefully will make many more of us who are interested in improved traffic flow aware, of the need for a better understanding of the basic freeway characteristics and control parameters.

In attempting to measure delay characteristics at the merging area of the entrance ramps, this investigation fulfills at least two needs. These include the provision of gap acceptance data for ramp simulation inputs and the accumulation of information for calibrating ramp metering models. Perhaps of greater significance, however, is the definition and supporting case that is built for use of a parameter called the "critical gap."

The probability of a driver having a particular critical gap is facilitated by the development of a distribution of critical gaps, and the fact that the mean of the critical was rather than the median should be used in delay computations is amply supported

ways follows the form of the Erlang distribution rather than a more specific form, the negative exponential distribution. This would corroborate earlier work in this area by Haight who has indicated that, while the utilization of a generalized Poisson distribution is satisfactory if the arrival times of vehicles are randomly distributed, many vehicles that appear to have gap distributions in the intermediate area between random and equally spaced distributions can best be described by a Pearson Type III distribution. The author has gone a step further by fitting the Pearson Type III distribution to the observed distributions of delays at each of the observed ramps of the Gulf Freeway and developing an equation for the expected queue length on the ramp.

The major criticism that could be leveled at this paper is very likely beyond the scope of the investigation but concerns its failure to suggest the potential cost savings that could accrue to the motorist from the eventual refinement of a more sophisticated surveillance and control system. Nevertheless, this study of gap acceptance characteristics is a major step toward the dual goals of moving high volumes of traffic at acceptable levels of service and coincidentally offsetting the cost of the control system by increasing the road-user benefits to all drivers on the system.

R. F. DAWSON, Assistant Professor, University of Connecticut, Highway Research Project, Storrs—Gap acceptance characteristics, gap characteristics and merging delay characteristics are essential in the development of ramp metering systems. There is an absolute need for research to formulate general models of these and other essential descriptors of the microaspects of the ramp-freeway queuing system. The work by Drew is a sound initial approach to the problem. There are, however, some aspects that have perhaps been oversimplified. The following discussion will be con-

fined to several of Drew's assumptions in the development of a ramp queuing model without consideration for the application of the proposed queuing model to ramp metering.

Main Components of Drew's Queuing Model

Gap Acceptance Characteristics.—Only a limited amount of gap acceptance research has been conducted, and as a consequence the phenomenon is not well understood. Drew's method for plotting the density function for critical gaps (it is not applicable to lags) is perhaps the best technique that has been proposed to date. The individual observations for which a density function is constructed consist of time range delimited by the largest gap rejected and the gap finally accepted. It is obvious that a driver's critical gap must fall in this range. Drew was successful in fitting an Erlang (a = 5) function (special case of the gamma function) to the density plot, but he presented data from just one ramp situation as evidence of the adequacy of the Erlang function as a general density model.

It will probably be necessary to give some consideration to the development of a technique for weighting individual critical gap density function observations obtained by the proposed Drew method. By their very nature, observations of narrow range would delimit a driver's critical gap more closely, but in the analysis by Drew, observations with wider ranges have greater influence on the shape of the density function. Two simple weighting procedures that might alleviate this undesirable property seem apparent. In the first such procedure individual observations could be given equal weight by recording each observation in the single 0.5-sec interval containing the midpoint of the range, on the assumption that the midpoint is the best estimate of a driver's critical gap. A second approach to weighting the observations would involve the application of a weighting factor that is inversely proportional to the length of the time range. The latter technique has been applied to Drew's data for the Dumble Street on-ramp, and the results are plotted in Figure 17. The Erlang function which has been fitted to

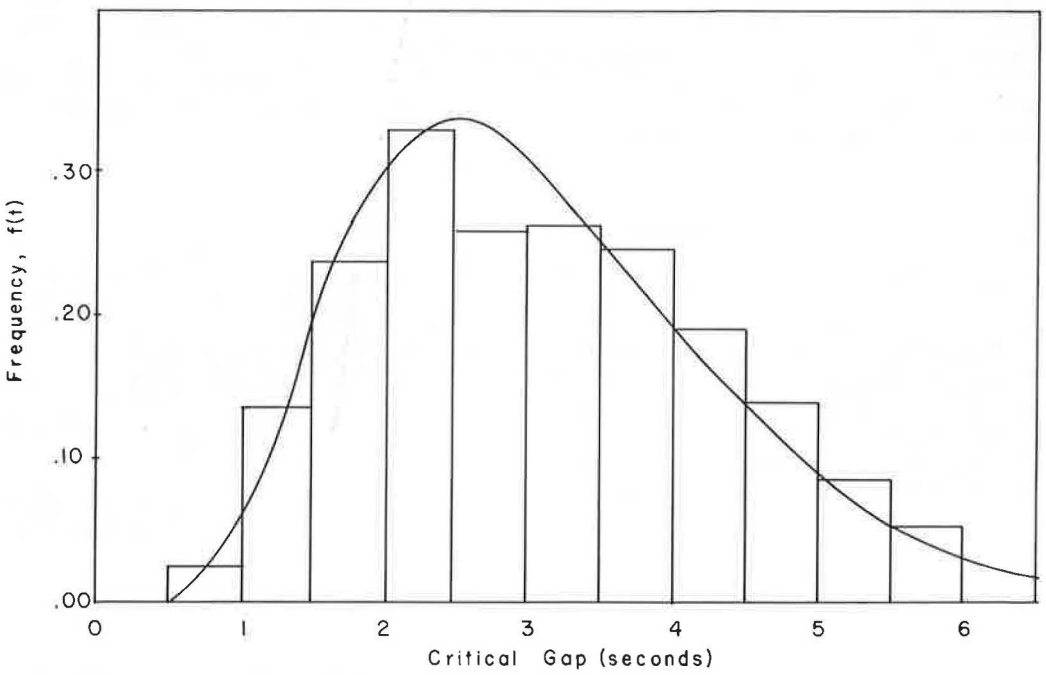


Figure 17. Comparison of weighted critical gap distribution with Erlang density function (a = 4).

these results [an Erlang (a = 4) function with a mean of 2.63 sec and a minimum of 0.50 sec] appears to be a much better fit than that presented by Drew in Figure 11.

Drew also describes a procedure, combining some of the features of the Raff method and some of the features of the Bissell method, that is applicable to the analysis of gap acceptance by both stopped and moving vehicles. The "probability of accepting a gap size t" as computed by this combination procedure, however, is a distinctly different statistic than the "probability of accepting a gap size t" as computed from the cumulative distribution of critical gaps obtained in the proposed Drew procedure. These two statistics do not even appear to have density functions, or cumulative density functions, from the same general family. This is evidenced by a comparison of plots of the two statistics on a standardized graph used for analysis of cumulative functions assumedto have Erlang density functions. If the cumulative function (in this case the probability of accepting a given gap) follows a given a plot on the standardized graph, the density function is an Erlang function of the order a. Failure of the observed cumulative function to follow any a plot is an indication that the density function is not of the Erlang family. The cumulative density function for stopped vehicles, obtained by the Drew method and plotted in Figure 18, is apparently a cumulative plot of the area under an (a = 4) or an (a = 5) function. Plots of the probabilities that given gaps are acceptable, as computed by the Raff-Bissell method, are shown in Figure 19 for stopped vehicles. Since this latter plot does not match an a line on the standardized graph, it can be concluded that the density function for critical gaps implied by this plot is not of the Erlang family.

Gap Characteristics.—As stated by Drew, it is necessary to have a description of gaps in the outside freeway lane, as well as a description of the critical gap distribution, in order to develop a mathematical model for estimating delay due to merging. Drew again proposes that the Erlang function might be the appropriate descriptor. The Erlang distribution, however, is a limited model; it would be useful for describing the availability of gaps in the freeway stream if it could be demonstrated that the distribu-

evidence has been presented, however, indicating that distributions of freeway vehicles tend to range from random distributions to what might be called hyper-random (more than random) distributions. Gap distributions presented in the Highway Capacity Manual are adequately described by a negative exponential (or random) distribution with a shifted axis. In their more recent research studies, Schuhl, Greenshields and May have found that the hyper-exponential distribution is a good descriptor of distributions of time gaps between freeway vehicles.

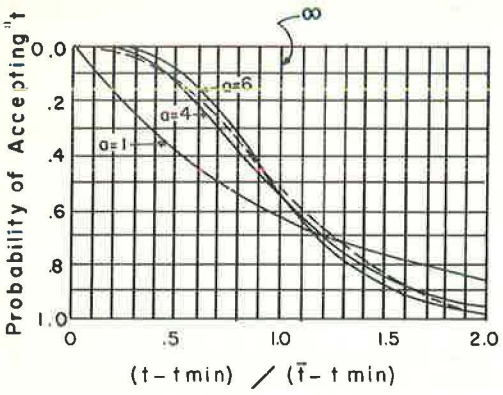


Figure 18. Comparison of drew probability of acceptance with standardized Erlang functions (stopped vehicles).

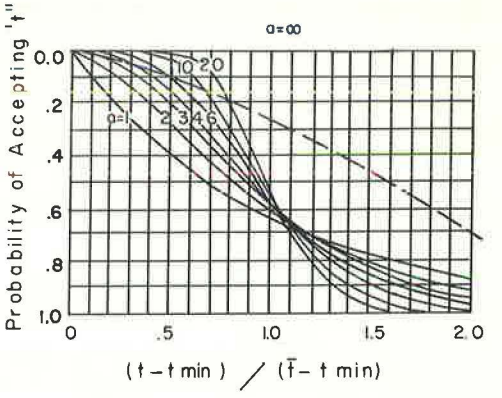


Figure 19. Comparison of "Raff-Bissell" probability of acceptance with standardized Erlang functions (stopped vehicles).

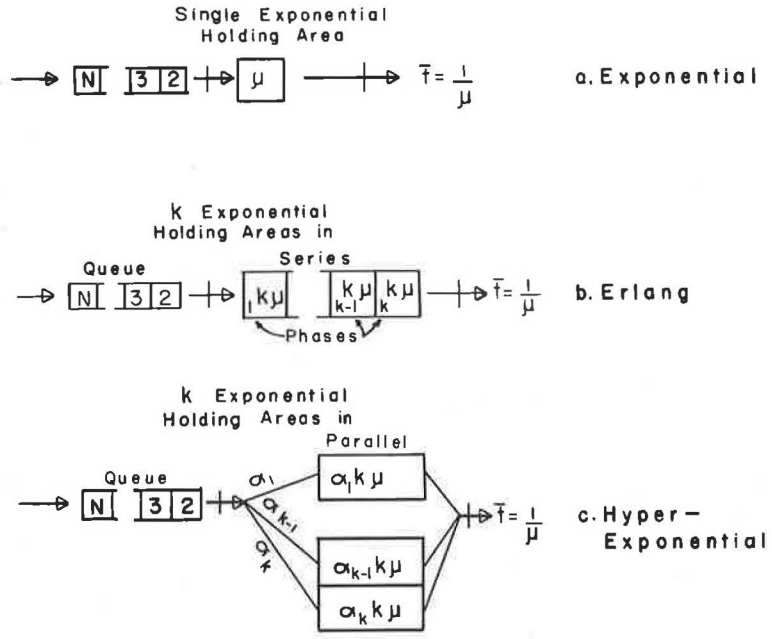


Figure 20. Schematic diagrams of simulators for generating exponential, Erlang and hyper-exponential gap distributions.

Proponents of both the Erlang and hyper-exponential gap distribution models offer rational, but incomplete, bases for their respective models. A more comprehensive description would be afforded by a model combining the properties of the hyper-exponential and Erlang models. Before proposing a specific model it would perhaps be worthwhile to show the physical relationship between the component parts. A schematic diagram of a device, or simulator, to generate completely random time gaps is shown in Figure 20(a). If a continuous queue exists at the entrance to a holding area from which units are allowed to depart at random with a mean rate μ, the times between departures from the holding area will be described by an exponential distribution with a mean of $\bar{t} = 1/\mu$.

Figure 20(b) shows a simulator for generating time gaps that range from the completely random exponential type to the completely uniform type. From a continuous queue at the entrance to the holding area, one unit at a time is allowed to enter the area, which is made up of several phases in series. The entering unit must first go through phase 1 which has an exponential time distribution with a mean rate of completion of $K\mu$. Upon leaving phase 1 the unit goes immediately to phase 2, phase 3, . . ., phase k-1, and to phase k, each with a mean rate of completion of $k\mu$. The distribution of times between departures of the units from phase k, and therefore from the holding area, are described by the Erlang distribution with a mean time $\bar{t} = 1/\mu$. Although the holding area is compounded, it can be considered as a single area because only one unit is allowed in it at a time. That is, the (n + 1) unit cannot enter phase 1 until the n unit has departed from phase k.

Hyper-random or hyper-exponential time gaps are generated by the simulator shown in Figure 20(c). From a continuous queue at the entrance to the holding area, one unit at a time is allowed to enter the area made up of k branches in parallel. As the unit enters it is assigned to one of the k branches at random, but on the average it is assigned to the ith branch a portion α_i of the time. When a unit is held in any one of the branches the entire holding area is considered to be occupied and no other unit can enter any branch until the preceding unit is discharged. The distribution of time gaps

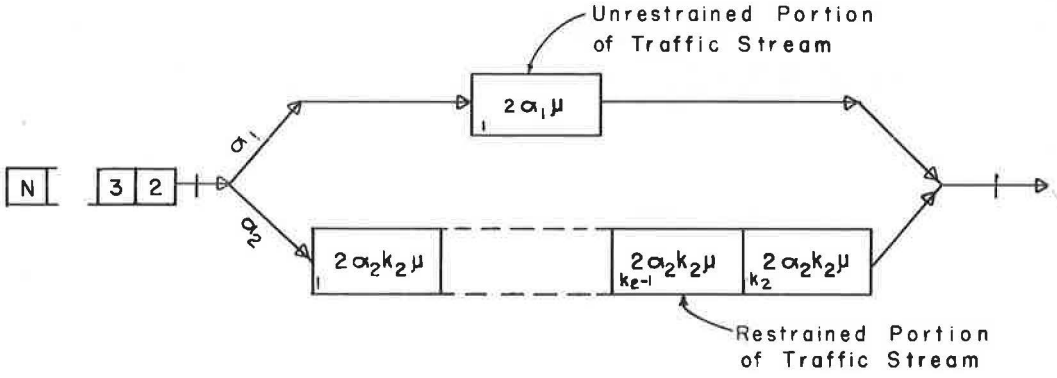


Figure 21. Hyper-Erlang gap distribution simulator for freeways.

between departures from this holding area is described as a hyper-exponential distribution with a mean time between departures of $\bar{t} = 1/\mu$.

At a glance it is apparent that the single exponential holding area is a special case of both the Erlang series arrangement and the hyper-exponential parallel arrangement. It is the case in which there is only one phase in the series arrangement, and the case in which there is only one branch in the parallel arrangement. The exponential distribution is therefore a boundary between Erlang distributions and hyper-exponential distributions.

There is, of course, no real physical device inherent in the outside lane of the freeway that regulates stream flow in such a manner as to generate either hyper-exponential or Erlang type headways. Field observations have shown that the simple exponential

the simple exponential function as a special case.

A more general model for describing gap availability could be constructed by combining the hyper-exponential and Erlang functions into a hyper-Erlang model, shown schematically in Figure 21. The model that might be fitted to gap data would probably consist of an Erlang ($a = k = 1$) function to describe unconstrained vehicles in the stream, in parallel with an Erlang ($a = k \geq 1$) function to describe the constrained vehicles in the stream. In addition, simple shifts of the axes to establish minimum gaps for both the constrained and unconstrained vehicles might be in order.

Delay and Queuing Characteristics

Drew has presented a collection of general, theoretical models for describing delay to the ramp stream as it traverses the merge area. The mere formulation of these models in general form is a significant contribution to the solution of the ramp queuing problem. Using the general equations, and substituting in any desired functions to describe gap availability in the shoulder lane of the freeway and gap acceptance by ramp vehicles, a traffic engineer could obtain estimates of the delay characteristics for given levels of freeway ramp operation. Some care would have to be exercised in applying the models, because they have been structured in such a way that they are not applicable to all ramp situations. In general, Drew's models are applicable to those situations in which all ramp drivers use the same decision model to make gap acceptance decisions. Such is the case on stop-sign controlled ramps where every ramp driver must make his gap-acceptance decision in a stopped condition.

If there is no absolute stop control on the ramp, a driver does not necessarily have to stop and he may make the gap acceptance decision in a moving condition. All drivers refusing a gap while moving would be forced to stop (or at least slow down) and would therefore make subsequent decisions in a stopped condition. Obviously, this complex situation cannot be handled by the proposed delay model unless a gap acceptance function

adequately approximating both decision functions can be obtained. But since Drew himself offers evidence that gap acceptance models for moving and stopped ramp vehicles are different, any such approximation would provide only approximate delay estimates.

There is reason to believe that gap acceptance is directly related to the relative speed between ramp vehicles and shoulder lane vehicles, rather than being just a two-phase phenomenon described separately for stopped and moving ramp vehicles. There is also evidence that gap acceptance characteristics for ramp vehicles merging into a freeway gap in queues are different from characteristics for single vehicles merging into the freeway flow. If these latter aspects of gap acceptance are significant, it will be difficult to adequately describe traffic flow on a ramp with a queuing model. Perhaps the traffic engineer will find that the ramp problem can best be analyzed by a Monte Carlo simulator, in which it is possible to include almost all of the micro aspects that influence overall traffic performance in the ramp-freeway merge area.

JAMES H. KELL, Principal Traffic Engineer, Traffic Research Corp.—Dr. Drew is to be complimented for preparing an excellent paper on a subject which is becoming increasingly critical in freeway operations.

I would like to confine my discussion to gaps and gap acceptance. I realize that my comments may be somewhat out of context with Drew's presentation in that he is primarily concerned with the utilization of gaps in a ramp control scheme. However, it is my belief that more knowledge concerning normal gap acceptance is desirable to develop fully the potential of ramp metering.

Drew defines a gap as the "interval of time between the arrival of two successive shoulder lane vehicles." This is not precisely true for the first gap (or lag, as many authors have chosen to call it) available to a merging vehicle. It is really the time between when the merging vehicle is ready to merge and the arrival of the next shoulder lane vehicle. Measurement of the first gap (or lag) and even succeeding gaps is complicated by the indefinite time when the merging vehicle is ready to accept a gap.

If visibility is good, the merging driver may be able to evaluate an approaching gap quite early and decide to accept it. He then accelerates or decelerates as appropriate to arrive at the merge point at the proper time to accept the gap. On the other hand, if visibility is not good, the driver cannot see the stream in advance and must, therefore, make a more hurried decision.

The technique described in this paper measured gaps in the shoulder lane and ramp arrivals at the nose of the ramp. Ramp vehicles may have already altered their behavior before reaching this point. Therefore, a true measurement of the ramp time is not really obtained.

Turning to the subject of critical gaps, Drew implies that each driver has a specific critical gap. Later on, he divides this into two critical gaps, one for moving vehicles and one for stopped. This is one area where the terminology of lags and gaps might be beneficial.

Assuming for the moment there is such a thing as a critical gap for an individual driver, it is my belief that this value would vary depending on traffic conditions present. The flow rate in the shoulder lane certainly affects the merging driver. With low flows, he is not too concerned if he misses one acceptable gap; another will appear directly. With high flows, he tends to accept shorter gaps since he might have to wait longer for another acceptable gap to arrive.

Speed also affects the entering driver. He will accept a shorter space gap when the freeway is moving slowly. Although I have no data to substantiate my contention, I believe that time gaps also lengthen with higher speeds.

In Bissell's work, which Drew quotes, approxmiately 5 percent of the vehicles observed ultimately accepted a shorter gap than one previously rejected. There is no indication in the paper that Drew observed this in his study, but I have observed it on

freeways, especially under conditions of high flow. Either the driver has not had time to evaluate conditions or he becomes impatient with the delay.

I have not collected data in the field, but have been interested in gaps and merging for quite some time, especially during the last few months since I reviewed Drew's paper. During this period I have attempted to evaluate my own behavior subjectively and have watched others perform the merging operation.

When the freeway is moving freely, I seem to follow a certain rote or procedure. The speed in the shoulder lane seems to be my primary criterion. In many cases, there is a lead vehicle starting the first available gap (orlag) in the shoulder lane. I apparently decide whether I can match his speed, merging behind him in the area available. If this decision is affirmative, I look for the following vehicle. In this case, space seems to be the important criterion. I have frequently accepted space gaps of 50 ft or less. At 60 mph this is a rather short time gap. I do tend to consider the following vehicle's speed and may reject the gap if that vehicle appears to be closing the gap, but not necessarily. In heavy traffic, when the freeway speed is lower, space seems to be the most important item. If there is space available much greater than a vehicle length and I can match the speed of the traffic stream, I will accept the gap.

My behavior is radically altered if there is a vehicle in front of me on the ramp. I tend to hold back to let the preceding vehicle merge and to allow myself maneuvering room to make my own merge.

When traffic is so heavy that I have to stop, especially where I am unable to observe the approach, I may miss acceptable gaps because I have not had time to evaluate the freeway stream. I know there have been numerous such occasions where I finally accepted gaps shorter than ones I had rejected.

Of course, I cannot say that I am an average or typical driver, but I am one charactor in the traffic stream. I have also watched other drivers merge. I have seen some drivers (especially where there was an acceleration lane) drive almost to the end of the land and then turn their heads for the first time to look for a gap. If an acceptable gap is not present right then and there, they stop. This tends to create the

This opens a much discussed topic as to whether driving habits vary in different parts of the country. I believe they do, but this is not the time nor place to pursue that subject.

However, there is one item that is worthy of mention that does occur on California freeways. There is frequently an interaction between the ramp vehicle and what I have called the following vehicle in the shoulder lane. The freeway vehicle will often modify his speed to alter the available gap when a ramp vehicle is present. Occasionally, he will speed up to reduce the gap and prevent the ramp vehicle from entering. More frequently, however, the following vehicle decelerates slightly to increase the gap and then closes up after the merge has been accomplished. There is also a strong tendency at many locations for the shoulder lane vehicles (even heavy trucks) to merge left in advance of a ramp, which is observed even under relatively high volume conditions.

To summarize this portion of the discussion, I believe that freeway speed and volume have a significant effect on gap acceptance. I alsofeel that an individual driver's critical or acceptable gap varies in any given situation as he comes down the ramp. It varies from an initial value for a "flying" merge to some other value (which may change) when he actually stops. To my mind, this becomes critical for a ramp metering scheme where gaps are measured at an upstream location and vehicles are released from the ramp to utilize these gaps.

I dispute Drew's nomenclature, that any vehicle that has rejected one gap is a stopped vehicle. I believe several gaps can be rejected with the ramp vehicle still moving with appreciable speed. This is highly dependent, of course, on the configuration of a particular on-ramp.

There is one other point in Drew's paper on which I would like to comment. In the discussion of stopped vehicles, he points out that a single vehicle can reject a number of gaps but can accept only one. I agree that a bias might be present if all rejected gaps are given the same weight as the one accepted. But, the one accepted gap is also

biased because the same vehicle might have accepted a shorter gap. In Drew's illustration (Fig. 10), the defining example is a vehicle that accepted a gap between 4.0 and 4.4 sec after rejecting a maximum gap of 0.5 to 0.9 sec. Might not this vehicle have accepted a gap of 3 sec, or 2 sec, or more? Conversely, might it not have rejected a gap of 2 sec, or 3 sec, or more? Yet this vehicle contributes eight intervals to the total being used to determine a distribution where other vehicles contribute only one. The resultant distribution is still biased. I believe the whole subject of determining gap acceptance distributions is worthy of more detailed investigation.

In conclusion, Drew is to be commended for adding to the basic knowledge of gap acceptance. Although this research was only an adjunct to their primary objective of freeway surveillance and control, it has provided valuable information in one of the basic areas where knowledge is lacking. I believe it has pointed out the need for more study in this area.

D. R. DREW, Closure—The author is appreciative of the excellent reviews by Dr. Paul D. Cribbins, Dr. Robert F. Dawson and Professor James Kell.

Dr. Cribbins has very aptly defined some of the objectives and problems associated with ramp control. His criticism of the failure in this report to suggest the potential cost savings to motorists attainable through more sophisticated surveillance and control systems is both constructive and timely. There is a definite need for cost-benefit work in this area.

Dr. Dawson has made several important observations that contribute much to the understanding of gap availability and gap acceptance:

1. Either of the weighting procedures he suggests would improve the fit between the observed and theoretical critical gap distributions shown in Figure 11.
2. The critical gap parameters summarized in Table 3 obtained from the procedure utilizing the gap acceptance functions of Figure 7 are, as he states, not the same as those obtainable from the critical gap distribution procedure illustrated in Figures 10 and 11. Actually, the linearity of the plots on the log probability paper in Figure 7 imply cumulative log normal distributions rather than Erlang distributions.
3. The "hyper-Erlang" gap distribution simulator he describes should enable investigators to add considerable realism to future freeway simulation models.

However, the advantage of the Erlang distribution over (a) the log-normal as a gap acceptance model, and over (b) the hyper-Erlang distribution as a gap availability model is dramatized in the derivations in the section on "Determination of Merging Delays" in the report (Eqs. 1-20). The latter distributions defy such analytical treatment.

Professor Kell's comments are directed to the oversimplification of merging as treated in the paper. He states that only gaps, not lags, were considered and that such important aspects of merging as multiple gap acceptance, driver policy, and driver characteristics have generally been ignored. Actually, these were not intended for consideration in this research, not because they are unimportant in explaining the merging maneuver, but because they are largely irrelevant to a ramp-freeway surveillance and control system. For example, lags cannot be detected and measured automatically, whereas gaps can. The purpose of this paper was to gain insight into the calibration of a metering system to be installed on the ramps of the Houston surveillance project.

Again, appreciation is expressed to the discussers for their reviews. The author sincerely feels that the three discussions represent significant contributions in their own right.