THE DETERMINATION OF MERGING CAPACITY AND ITS APPLICATION TO FREEWAY DESIGN AND CONTROL

by

Donald R. Drew
Head of Design and Traffic Department
and Principal Investigator

Johann H. Buhr
Assistant Research Engineer

and

Robert H. Whitson
Research Assistant

Texas Transportation Institute
Texas A&M University
College Station, Texas

Sponsored by
U.S. Bureau of Public Roads
Contract No. CPR-11-2842

ABSTRACT

This paper presents a new approach to the determination of the capacity and service volumes in ramp-freeway merging areas. The capacity of a merging area is based on the critical gap concept and on assumptions regarding the distribution of gaps in the freeway shoulder lane. The service volumes suggested are developed from considerations of the ramp junction as a queueing system. A level of service can then be provided such that a ramp vehicle has a certain probability of finding the merging area empty. Another measure of level of service is the delay suffered by ramp vehicles. This aspect is treated and charts presented for its determination.

The above merging parameters all involve the critical gap of the junction. This critical gap can be estimated from the geometrics of the ramp-freeway junction by a regression equation, developed through the study of a number of entrance facilities, which relates the critical gap to the length of acceleration lane and the angle of convergence. Relationships are also presented of estimating the entire gap acceptance characteristic from these two geometric features.

The paper proceeds to discuss in detail the application of the developed merging parameters in freeway design and control.

CONTENTS

INTRODUCTION

In freeway design, the engineer is often faced with the problem of determining the capacity of a merging area. Existing procedures permit the estimation of outside freeway lane volumes which is then subtracted from fixed control values to give allowable ramp service volumes.[1, 2, 3, 4] These procedures take various traffic characteristics and ramp configurations into consideration but do not account for the effects of the geometrics of the ramp terminal itself. As a result, it is usually assumed that the entrance facility does not suffer from such design deficiencies as a short acceleration lane, a high angle of entry, inadequate sight distance or poor delineation. In practice however, physical limitations will often preclude the development of ideal geometrics so that the designer is forced to use a substandard design. In this paper, the effect on the merging operation of two of these geometric variables, acceleration lane length and angle of entry are analyzed and its application to the design and control of entrance terminals discussed.

Project Objectives

This paper is the fourth in a series of papers on freeway merging resulting from research undertaken by the Texas Transportation Institute and sponsored by the United States Department of Commerce, Bureau of Public Roads. The project, entitled "Gap Acceptance and Traffic Interaction in the Freeway Merging Process" has as its main objective the development of relationships between the many variables associated with the freeway merging process. Such relationships should reflect the effects of traffic characteristics and ramp geometrics on the operation and level of service of the merging area so as to permit their application in the design, operation and simulation of ramp entrance terminals.

Previous Reports

Three earlier reports have already been published. The first report[5] deals primarily with the data collection, reduction and analysis techniques employed. It also presents the study sites selected and discusses some of the data editing and analysis programs written for the project. The qualitative effects of various geometric elements on the operation as mirrored by the traffic parameters of volume, density, speed and acceleration are discussed and presented in the form of contour diagrams. This paper further serves to demonstrate not only the nature of the data available but also the vast quantity of data involved.

The second report[6] deals primarily with the gap acceptance behavior of entering drivers. Theoretical models developed for describing the merging process include the derivation of the mean and variance of the delay to a ramp vehicle and the treatment of the variability of critical gaps and gap acceptance among driver through the identification of the representative forms for both critical gap distributions and gap acceptance functions. Using probit analysis techniques, the gap and lag acceptance characteristics of single and multiple entry merges are determined as well as the effects of ramp speeds and relative speeds on these characteristics. This was done for several films taken at each of thirty-two ramps and thus reflects diverse operating, geometric, geographic and environmental conditions.

The third report[7] deals primarily with the effects of entrance ramp geometrics on the driving practice of entering drivers as reflected by the distribution of ramp speeds. The speeds considered are: ramp speeds at the nose, ramp merging speeds, relative speeds at the nose, relative speeds of merge and speed changes on the acceleration lane. The distributions of these speeds are presented in such a manner that qualitative comparisons can be made and the effects of acceleration lane length and angle of convergence illustrated. No qualitative analyses are attempted. Other criteria used to illustrate the effects of ramp geometrics are the number of gaps rejected before acceptance and the distribution of entry points along the acceleration lane.

These three reports are effectively summarized in a paper presented at the Bureau of Public Roads Research and Development Program Review Meeting in December of 1966.[8]

This paper reaches further towards the objectives of the overall project in that it relates ramp geometrics to merging capacity and level of service through considerations of the gap acceptance practices of drivers, suggesting further the utility of these relationships in freeway design and control.

DETERMINATION OF MERGING PARAMETERS

Merging Capacity

The efficiency of traffic movement on the through lanes of an urban freeway are directly affected by the adequacy of the associated ramps. The proper design and placement of ramps on high-volume freeways is therefore imperative if those facilities are to afford fast, efficient and safe operation. The development of such suitable designs depends to a large extent on the accurate determination of the capacity at the ramp junction, heretofore referred to as the merging capacity.

In the merging situation, the maximum number of ramp vehicles that can be accommodated in the shoulder lane is equivalent to the number of ramp vehicles that will use each available gap, assuming a continuous backlog of vehicles on the ramp.[9] The concept of gap acceptance is therefore of major importance in considerations of the capacity of a merging area.

Consider a single, inexhaustible queue waiting to enter a random shoulder-lane traffic stream. If the passing time headway, t, less than the critical gap, T, no ramp vehicle enters; if t is between T and T + T' one vehicle enters; if t is between T + T' and T + 2T' two vehicles enter, etc. The ability of the outside freeway lane to absorb ramp vehicles per unit of time becomes

$$q_r = q \sum_{i=o}^{\infty} (i+1) \, P(T + iT' < t < T + (i+1)T') \qquad (1)$$

where q is the shoulder lane flow.

If the distribution of gaps in the shoulder lanes f(t) is given by the negative exponential distribution

$$f(t) = qe^{-qt} ,$$

then it follows from (1) that

$$q_r = q\,[e^{-qT} - e^{-q(T+T')}] + 2q\,[e^{-q(T+T')} - e^{-q(T+2T')}] + \ldots$$

$$= qe^{-qT} + qe^{-q(T+T')} + qe^{-q(T+2T')} + \ldots$$

$$= qe^{-qT} (1+e^{-qT'} + e^{-2qT'} + \ldots)$$

$$= \frac{qe^{-qT}}{1-e^{-qT'}} \qquad \frac{qe^{-1}}{1-e^{-1}} = \frac{1/3 \cdot .368}{.632} \tag{2}$$

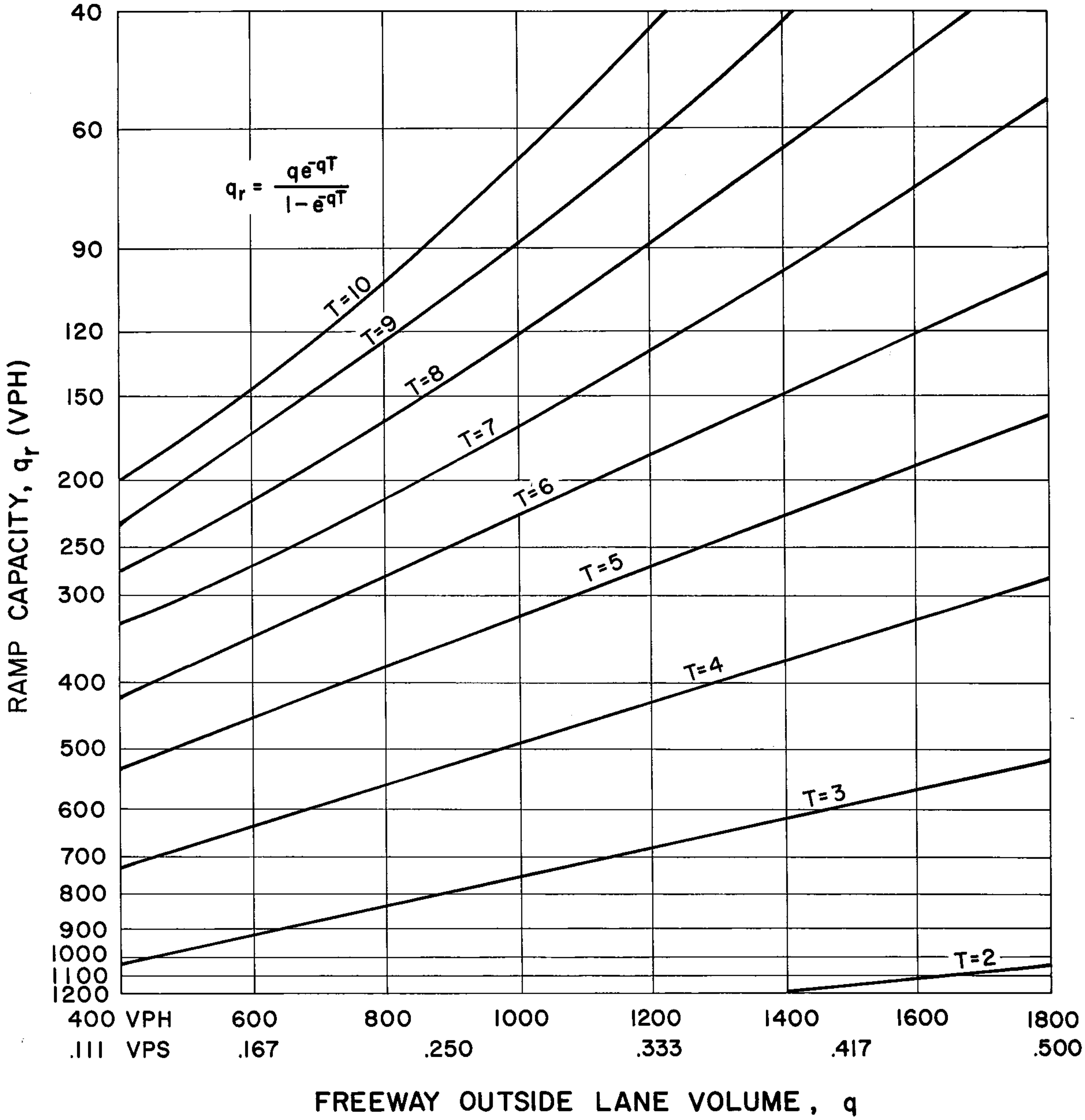
In the previous report[6] dealing with multiple entries it was concluded that double entries are more "sensitive" than single entries, and triple entries more "sensitive" than double entries to differences in gap sizes. Here "sensitivity" means the data shows that the percent acceptance curves are steeper for triple entries than for double entries and steeper for double than for single entries. The percent acceptance curves show that at the 50 percentile acceptance level (the critical gap), T and T' are approximately equal. Therefore, the expression for ramp capacity in (2) may be simplified, becoming

$$q_r = \frac{qe^{-qT}}{1-e^{-qT}} \tag{3}$$

Equation (3) is illustrated in Figure 1. In order to use the graph, it is necessary that the shoulder lane volume q and the critical gap T be known. If an existing design is being evaluated, q, q_r and T can be measured. In the case of a proposed design, methods of estimating the percent of the total freeway volume in the shoulder lane are well documented.[1] Variables which have been found to significantly affect q are the total freeway volume, the entrance ramp volume q_r, upstream ramp volume, downstream exit ramp volume, and distance to downstream exit ramp. Figure 2 is illustrative of the relationship between q and two of these variables - total freeway volume and entrance ramp volume, q_r.

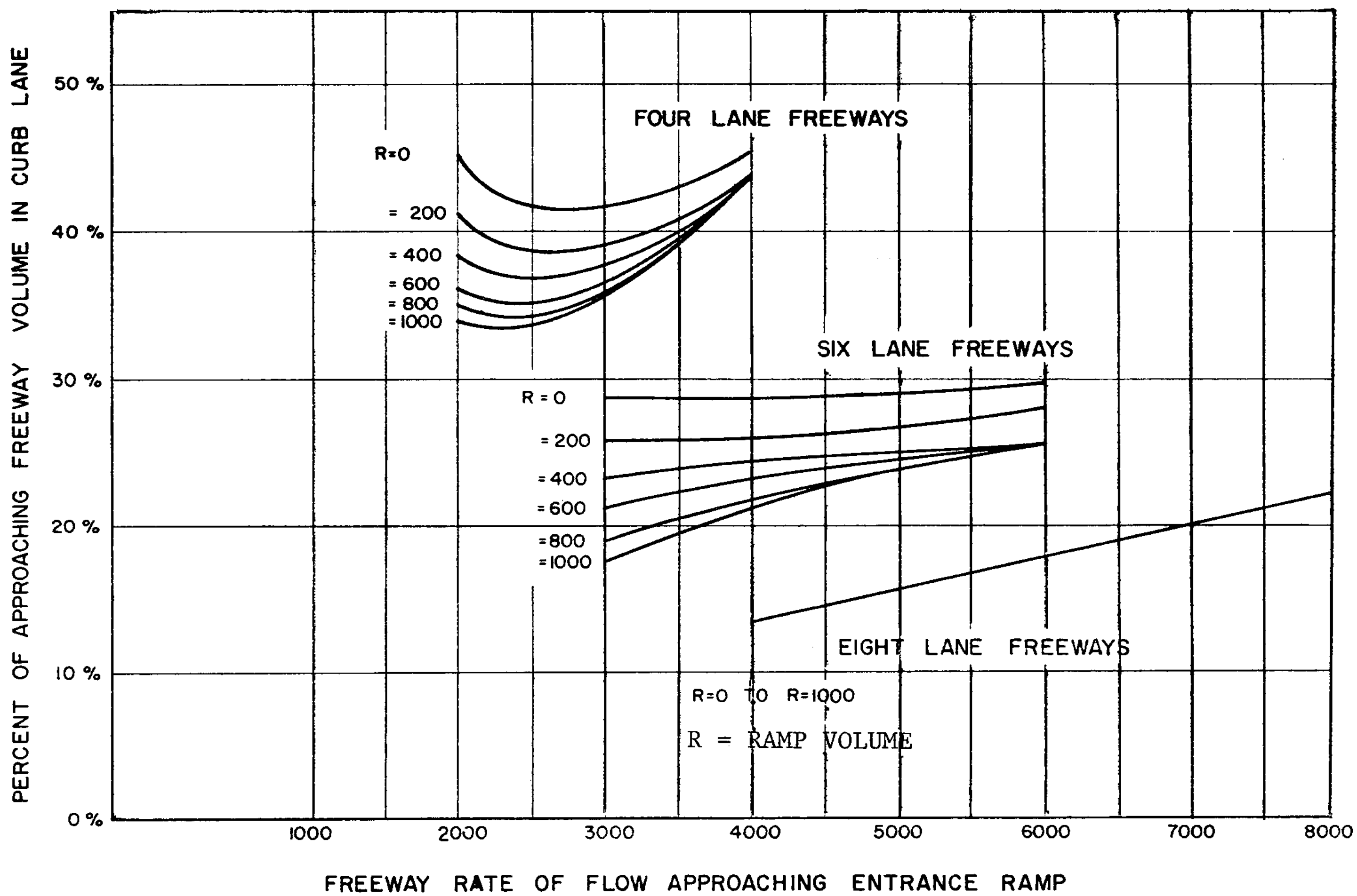
Merging Service Volumes

As defined in the previous section, the merging capacity is the maximum number of vehicles that can be accommodated with a continual backlog or queue of ramp vehicles. As such, it is a possible capacity. Whenever the opportunity occurs for r vehicles to enter the shoulder-lane stream there must, of course, be at least r vehicles queued on



POSSIBLE CAPACITY OF MERGING AREAS

Figure 1



RELATIONSHIP BETWEEN PERCENT OF TOTAL FREEWAY VOLUME IN THE OUTSIDE LANE AND FREEWAY AND RAMP VOLUMES

Figure 2

the ramp to utilize this capacity potential. Although the delay or queue lengths associated with such a traffic condition could be excessive, they were not considered in the capacity analysis. However, in order to provide a certain level of service, the determination of merging service volumes must take delays or queue lengths into account.

In an earlier report in this series[6], expressions were derived for the mean and variance of the delay suffered by ramp vehicles in position to merge. They are

$$E(t)_a = \frac{e^{aqT} - \sum_{i=0}^{a} \frac{(aqT)^i}{i!}}{q \sum_{i=0}^{a-1} \frac{(aqT)^i}{i!}} \qquad (4)$$

$${}^2(t)_a = \frac{a+1 \left[e^{aqT} - \sum_{i=0}^{a+1} \frac{(aqT)^i}{i!} \right]}{aq^2 T \sum_{i=0}^{a-1} \frac{(aqT)^i}{i!}} + E^2(t)_a \qquad (5)$$

where a is the parameter of the Erlang Distribution denoting the distribution of headways in the shoulder-lane stream.

Considering the ramp junction as a queueing system, the entrance ramp vehicles arrive at the junction at a rate q_r and are obliged to yield to the freeway traffic, thus forming a single line waiting for successive vehicles at the head of the queue to merge. If the moments of the distribution of time, f(t), spent by vehicles at the head of the queue are given by (4) and (5), the expressions for some useful queueing parameters may be developed.

Let n_0, n_1 denote the ramp queue lengths immediately after two successive ramp vehicles C_0, C_1 have merged. Let t be the service time of C_1 and 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

$$n_1 = n_0 + r - 1 + \qquad (6)$$

It is to be noted from the definition of δ that

$$\delta^2 = \delta$$

and that

$$n_0 \delta = 0,$$

and hence, from (6), on taking expected values, we obtain

$$E(n_1) = E(n_0) + E(r) - 1 + E(\delta). \qquad (7)$$

If the system is assumed to be in a state of statistical equilibrium, then

$$E(n_1) = E(n_0) = E(n)$$

and

$$E(r) = q_r \; E(t)_a$$

Thus, substituting in (7),

$$E(\delta) = 1 - q_r \; E(t) \qquad (8)$$

However, we also know that by definition

$$E\,(\delta) = \sum_{\delta=0}^{1} \delta P(\delta) = P(\delta=1) = P(n_0=0) = P_0 \qquad (9)$$

Equating (8) to (9), one finds that

$$q_r = (1-P_0)/E(t)_a \qquad (10)$$

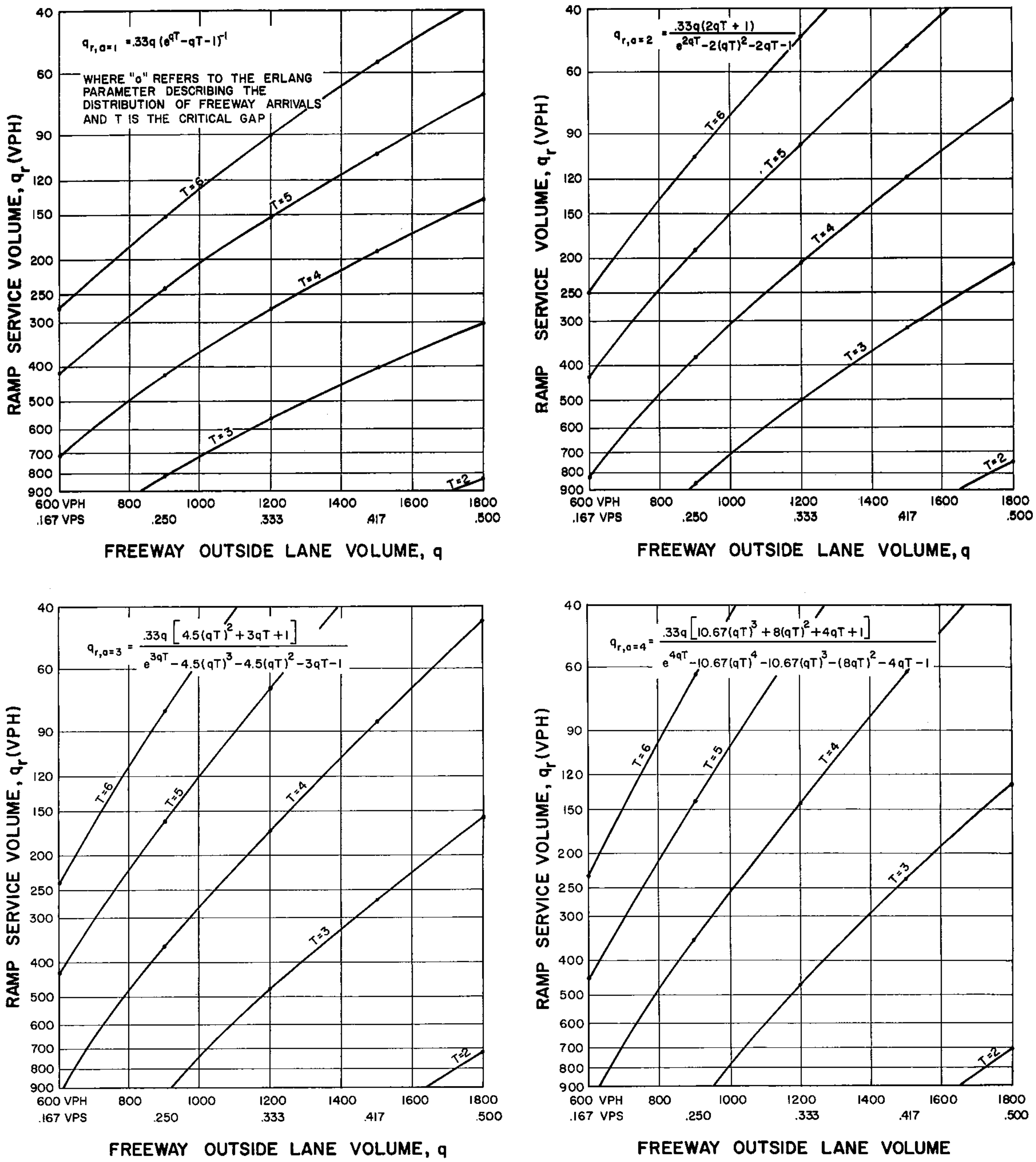
The ramp flow q_r may be interpreted as a ramp service volume as opposed to ramp capacity. P_0 is the probability of a ramp arrival finding the merging area empty, and as such affords a measure of level of service. Substituting (4) in (10) and arbitrarily allowing P_0 to equal .67 provides the basis for the ramp service volume curves in Figure 3. For example, for a=1, the ramp service volume is given by

$$q_r = .33q\,(e^{qT} - q^{T} - 1)^{-1} \qquad (11)$$

Figure 3 illustrates the relationship between the freeway shoulder lane volume q, the headway distribution of the shoulder-lane stream as defined by the Erlang parameter a, the critical gap T, and the ramp volume q_r. Consider an entrance ramp operating with a critical gap of 4.0 seconds and with the distribution of freeway traffic conforming to an Erlang Distribution with a=2. It is apparent that the sum of the coordinates of any point on the line T=4 in the graph a=2 of Figure 3 describes the merging service volume for that ramp. For example, the point described by q=1500 and q_r=120 tells us that the merging service volume is 1620 and that under these operating characteristics, a ramp arrival has a 67% chance of finding the ramp empty or a 33% of finding a vehicle ahead of it trying to merge.

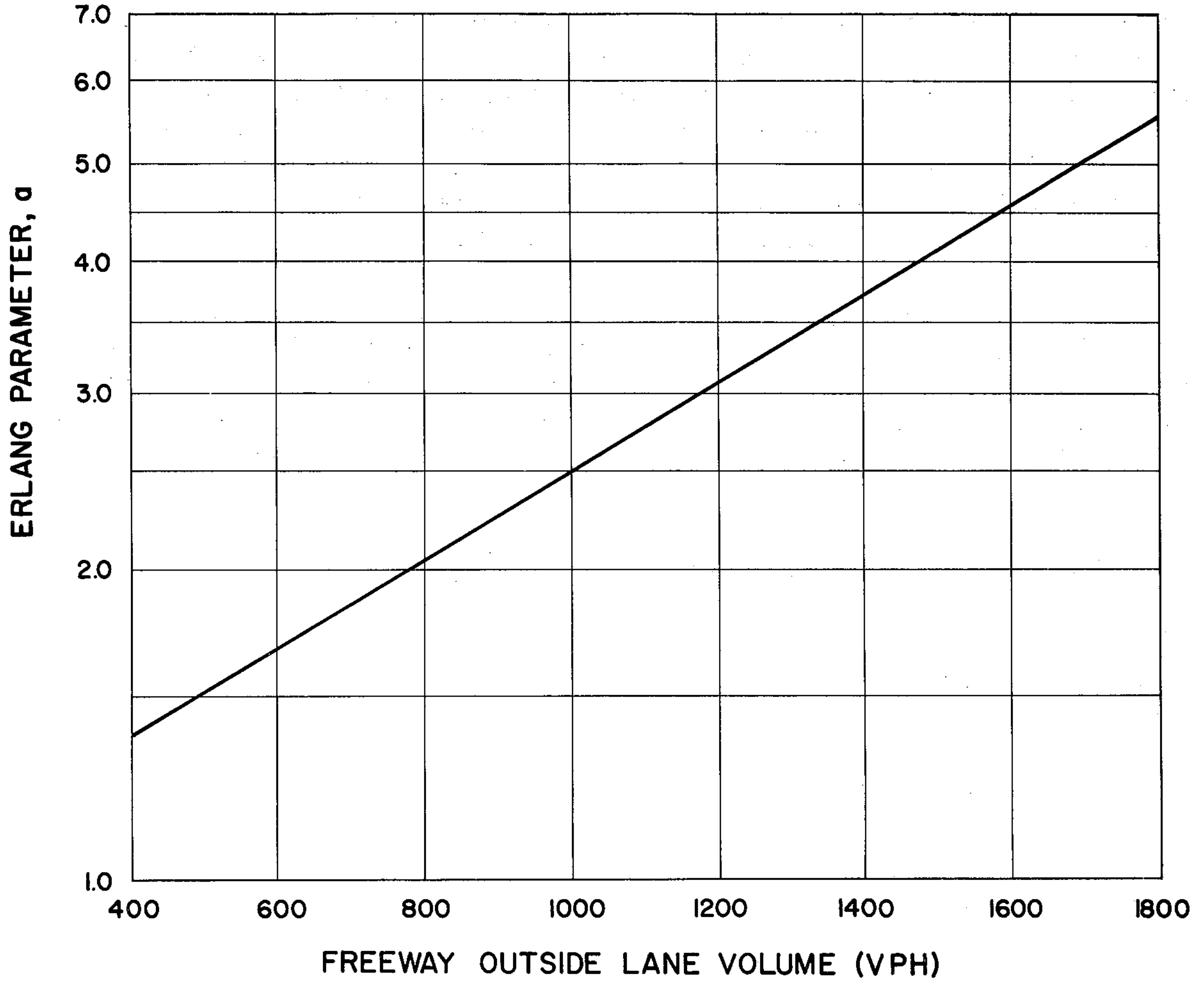
In the design of a new facility however, the engineer is confronted with asset of assigned volumes and have no knowledge of the value of T and a, and is therefore at a loss as to which curve or even which set of curves to use for determining the service volume of a proposed design. Towards this end, the dollected data was analyzed with the objective of formulating relationships that would be useful to the designed is that it should allow the prediction of the Erlang parameter a and the initial gap T, from information that would normally be available.

The Erlang parameter is largely affected by the volume level. To be sure, there are several other variables such as alignment, grade and other environmental elements which affect the value of a. However, in the absence of any knowledge of these variables, the curve shown in Figure 4 has been found to approximate the value of a as related to the



MAXIMUM SERVICE VOLUME q_r FOR A RAMP ARRIVAL TO HAVE A PROBABILITY OF .67 OF FINDING NO RAMP VEHICLES IN THE MERGING AREA

Figure 3



APPROXIMATE VALUE OF ERLANG a AS RELATED TO THE FREEWAY OUTSIDE LANE VOLUME

Figure 4

freeway volume. This curve can be used in conjunction with the relationships shown earlier in Figure 3 to develop the set of curves shown in Figure 5 which relates the ramp service volume to the outside freeway lane volume and the critical gap T, eliminating the need for a knowledge of the Erlang parameter a of the freeway traffic time headway distribution. The critical gap T can be estimated from the geometrics of the entrance terminal. Before this is treated in detail, some other measures of level of service will be discussed.

The mean queue length confronting an arriving ramp vehicle and its delay (time in the system) are additional measures of the level of service afforded ramp traffic. Expressions for these parameters are obtainable using the techniques in Equations (6) through (11). Squaring both sides of (6) and taking expected values as before, leads to

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

which reduces to

$$E(n_0) = \rho + \frac{E(r^2) - \rho}{2(1-\rho)} \qquad (12)$$

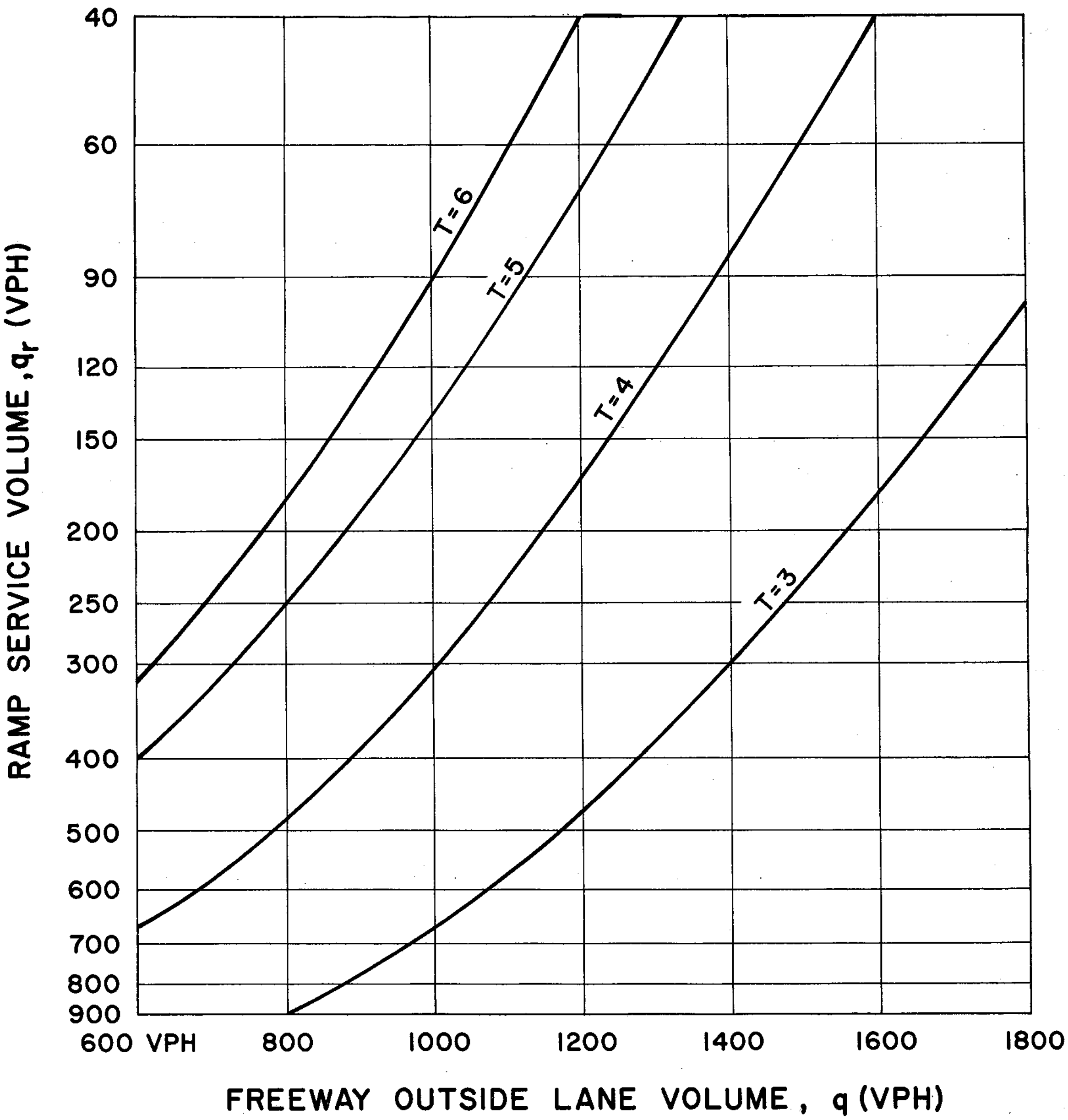
where $\rho = q_r E(t)$.

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, we have

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

and, considering the relationship between the first two moments and the variance

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



MAXIMUM SERVICE VOLUME q_r FOR A RAMP ARRIVAL TO HAVE A PROBABILITY OF .67 OF FINDING NO RAMP VEHICLES IN THE MERGING AREA

Figure 5

then

$$E(r^2) = \rho + \rho^2 + q_r^2 \sigma^2(t) \tag{13}$$

Substituting (13) in (12) gives the expected queue length on the ramp as

$$E(n) = \frac{q_r^2 \sigma^2 + \rho^2}{2(1-\rho)} \tag{14}$$

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.

$$E(n) = q_r E(t+w)$$

and

$$E(w) = [E(n)/q_r] - E(t) \tag{15}$$

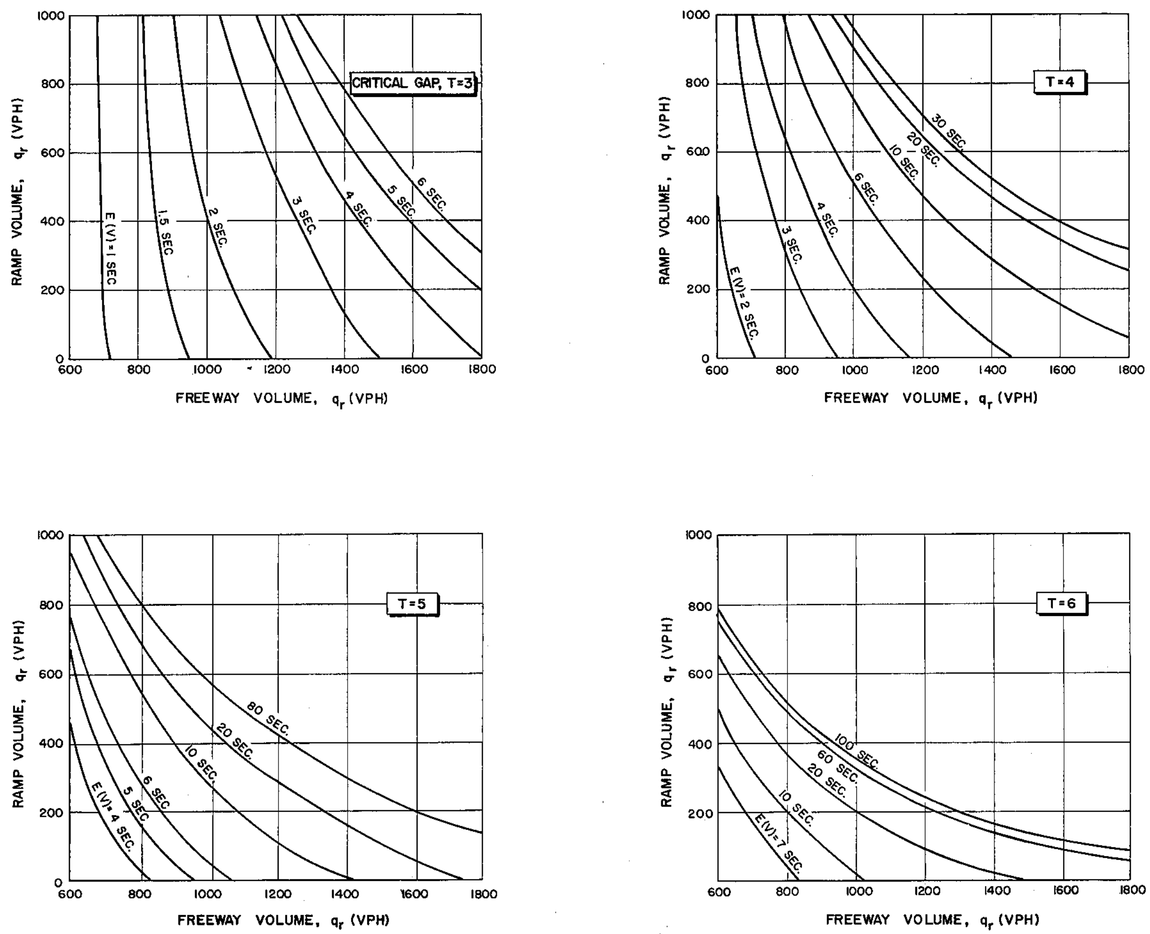
It follows that the mean wait in the system for a ramp vehicle is

$$E(v) = E(n)/q_r \tag{16}$$

The curves for Equation (16) are plotted in Figure 6 in terms of the shoulder-lane volume q and the ramp volume q_r. The distribution of headways on the shoulder-lane is assumed to be random (a=1) with critical gap values of T=3, 4, 5 and 6 seconds.

Estimation of the Critical Gap

As indicated above, the critical gap T must be estimated in order to find the capacity and service volume of a proposed design. This estimation must, of course, be based on the geometrics of the merging area.



AVERAGE DELAY OF RAMP VEHICLES WITH RANDOM FREEWAY ARRIVALS

FIGURE 6

The elements of good design for an entrance ramp junction are well documented. They are, to name a few: (1) adequate length for drivers to accomplish merging; (2) a flat angle of approach which aligns the driver along an easy and natural path into the acceleration lane; (3) good visibility to allow the entrance ramp driver to judge and accept a freeway gap with a minimum of indecision; and (4) a clearly marked and delineated entrance ramp which would eliminate any confusion in distinguishing between the entrance ramp elements and the main freeway lanes.

Translated into geometric variables, the critical gap depends primarily on the length of acceleration lane, L, and the angle of entry, θ . This effect was clearly evidenced by the study data as demonstrated in Figures 7 and 8. Figure 7 shows the effect of the length of acceleration on the gap acceptance behavior of drivers for ramps having convergence angles from 3 to 6 degrees. The effect of the angle of convergence on percent acceptance is shown in Figure 8 for ramps having acceleration lanes between 650 and 800 feet. It can be seen that 50% of the drivers accepted gaps less than 1.5 seconds at the entrance ramp with a 3 degree angle of convergence whereas the 50% percentile gap is 3.5 seconds for an 11 degree angle.[8]

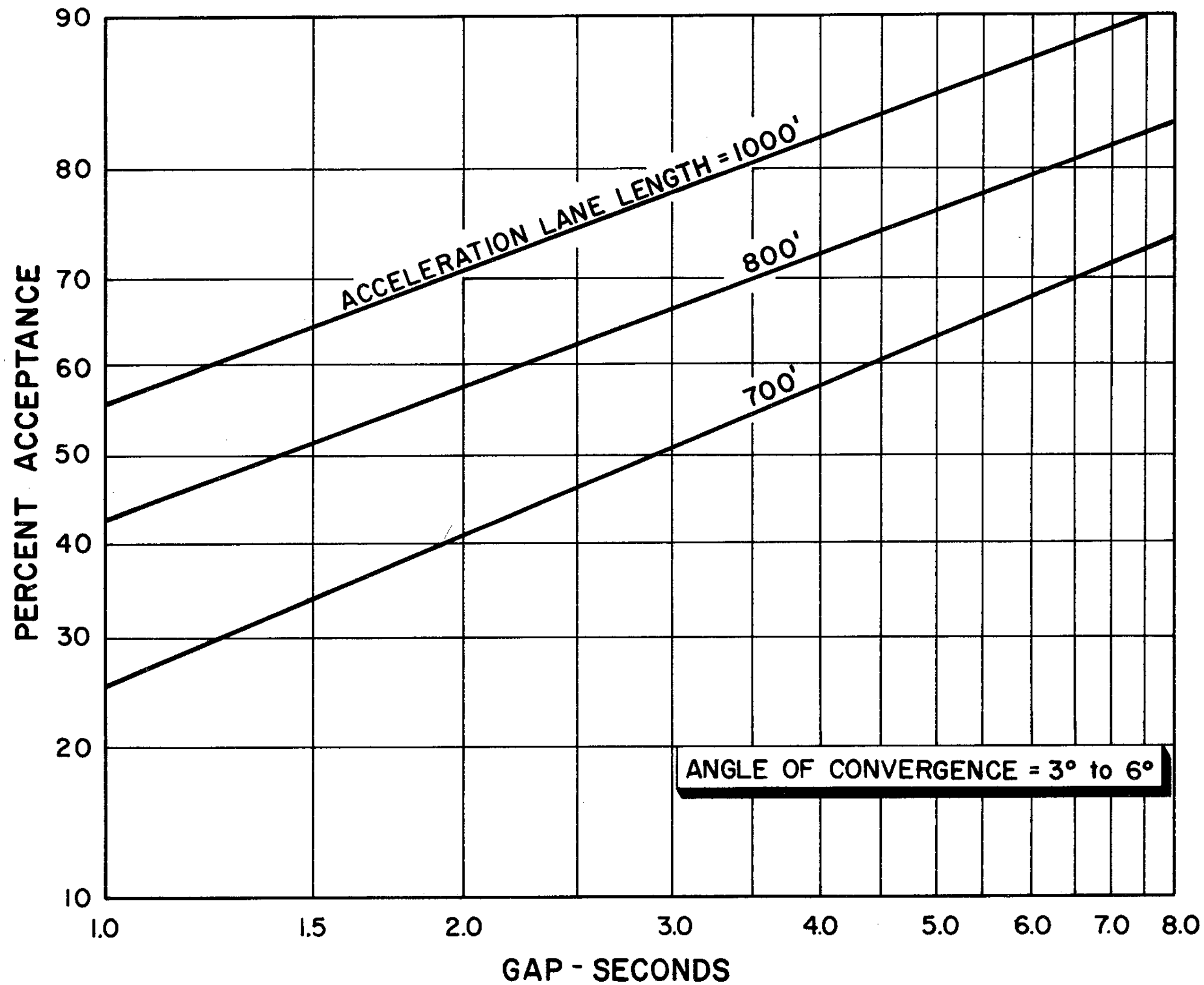
In order to determine the effects of acceleration lane length and angle of converge on the critical gap and to develop a relationship between the geometric variables and the gap acceptance characteristic of an entrance ramp, two sets of regression analyses were performed. Using as input data the gap acceptance characteristic developed from each data film[8], regression equations were found for the 50 percentile or critical gap and for the slope of the gap acceptance line, using a stepdown procedure.

The critical gap is given by

$$T = 5.547 + 0.828\theta - 1.043L + 0.045L^2 - 0.042\theta^2 - 0.874S$$

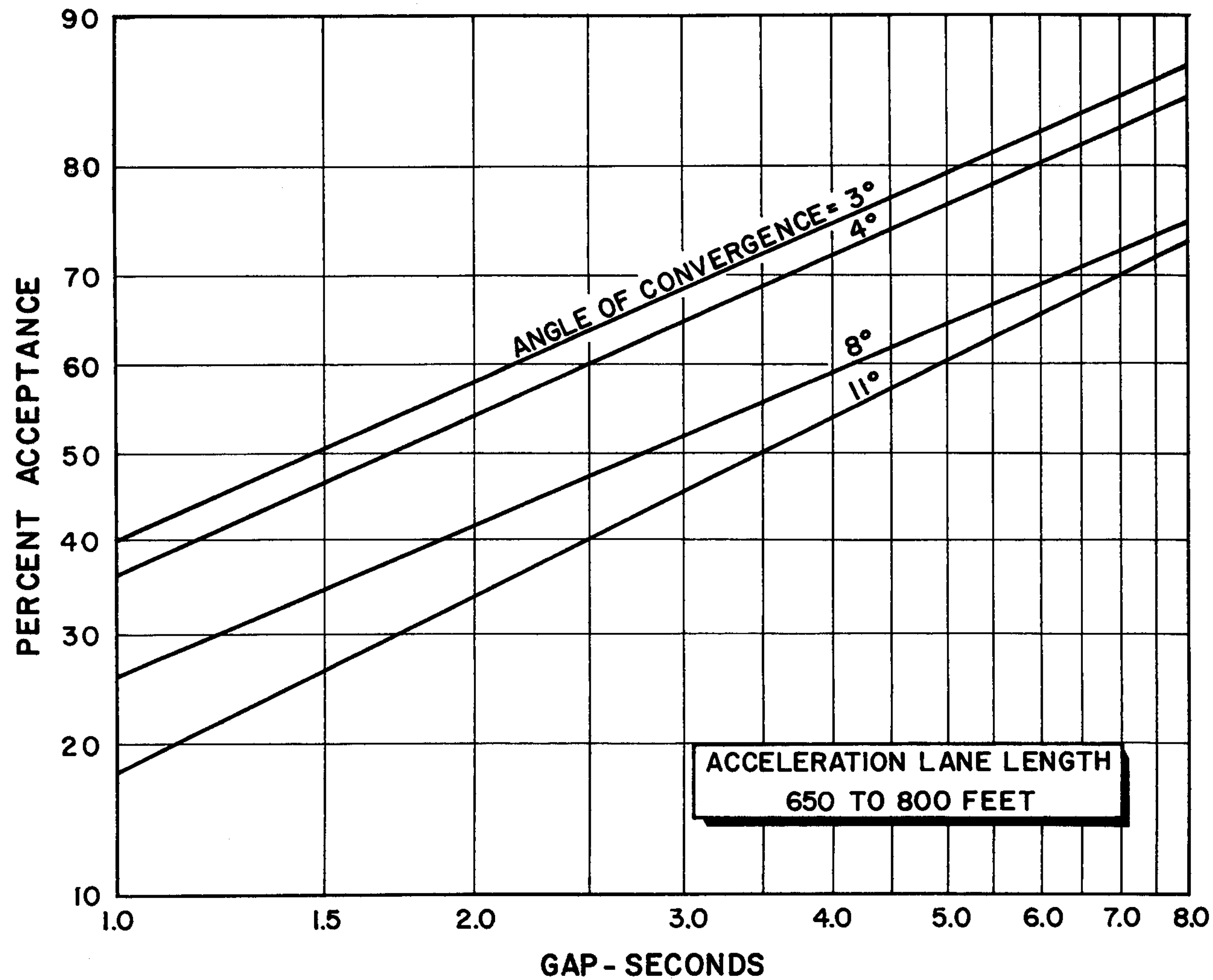
where θ = angle of convergence in degrees
L = length of acceleration lane in station
S = shape factor = 1 for taper type
= 0 for parallel type

From the equation, it can be seen that increasing the angle, increases the critical gap while increasing the length, decreases the critical gap. As expected then, a lower angle of convergence and a



EFFECT OF LENGTH OF ACCELERATION LANE ON GAP ACCEPTANCE CHARACTERISTICS

FIGURE 7



EFFECT OF CONVERGENCE ANGLE ON GAP ACCEPTANCE CHARACTERISTICS

FIGURE 8

longer acceleration lane are desirable in the design of the ramp junction. Although this is not new, the significance of the above equation is that it quantifies these effects so that the designer can determine what is gained or lost by varying θ and L. Curves for estimating the critical gap from the length of accerlation lane and angle of entry are shown in Figure 9.

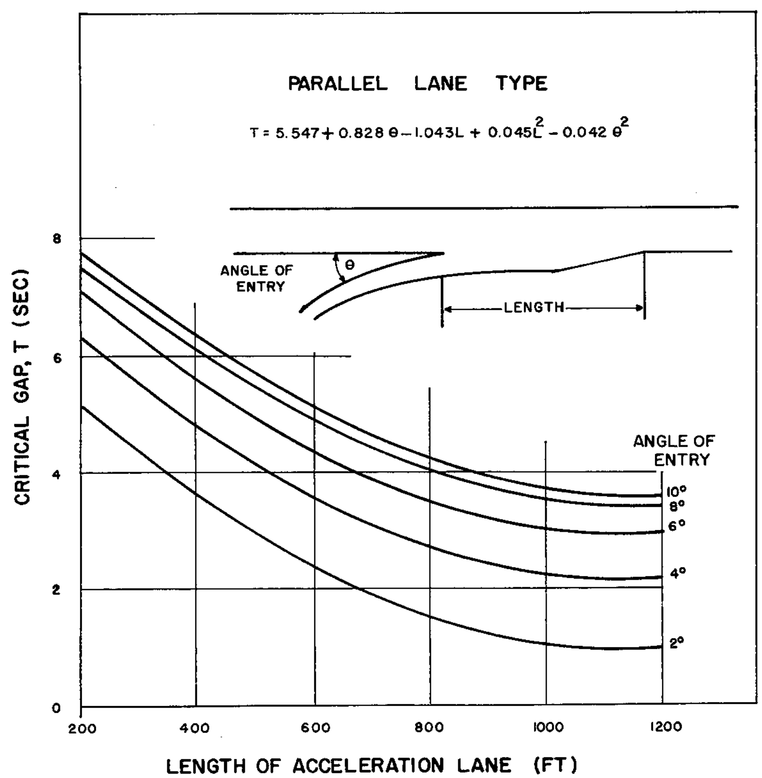
Another interesting aspect of this analysis is that it shows, for the first time in the literature, a difference between a taper type and a parallel lane type of acceleration lane - a topic that has given rise to considerable controversy. According to the analysis which is based on observations of the operation of 13 taper type and 16 parallel lane type acceleration lanes, a tapered entrance terminal will, on the average, have a critical gap that is about 0.9 second smaller than that of a parallel lane type acceleration lane with the same length and the same angle of convergence. This finding is not to be construed as an unconditional endorsement of the taper type junction. Other factors such as grade, ramp length and curvature and environment also plays an important part. The limitations of the data and the above mentioned analysis should be kept in mind. Nonetheless, based on the study data, it appears than under identical conditions the taper type acceleration lane has, on the average, a more favorable gap acceptance characteristic than the parallel lane type.

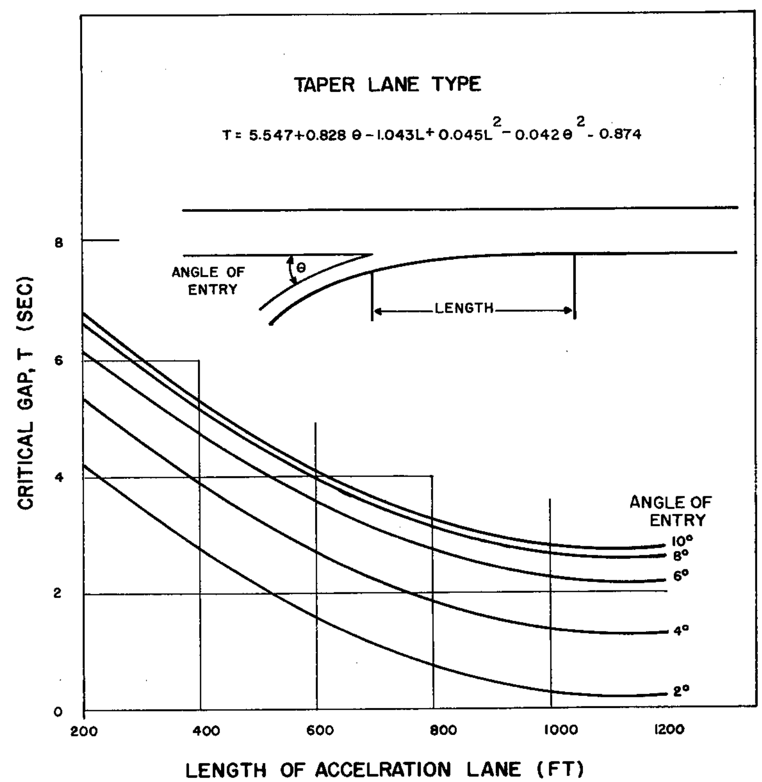
The stepdown regression analysis with the slop of the gap acceptance line as the independent variable, yielded the following equation.

$$B_1 = 1.394 + 0.289\theta - 0.027L\theta$$

where θ = angle of convergence in degrees
L = length of acceleration lane in station
B_1 = slope of the gap acceptance line as expressed by $Y=A+B_1X$, Y being the probit and X the logrithm of the gap size.

An increase in the angle increases the slope of the gap acceptance line and thus decreases the variance of the critical gap distribution while the acceleration lane length has the opposite effect for a fixed angle of entry. Note that the shape of the acceleration lane does not effect the variance of the critical gap distribution.

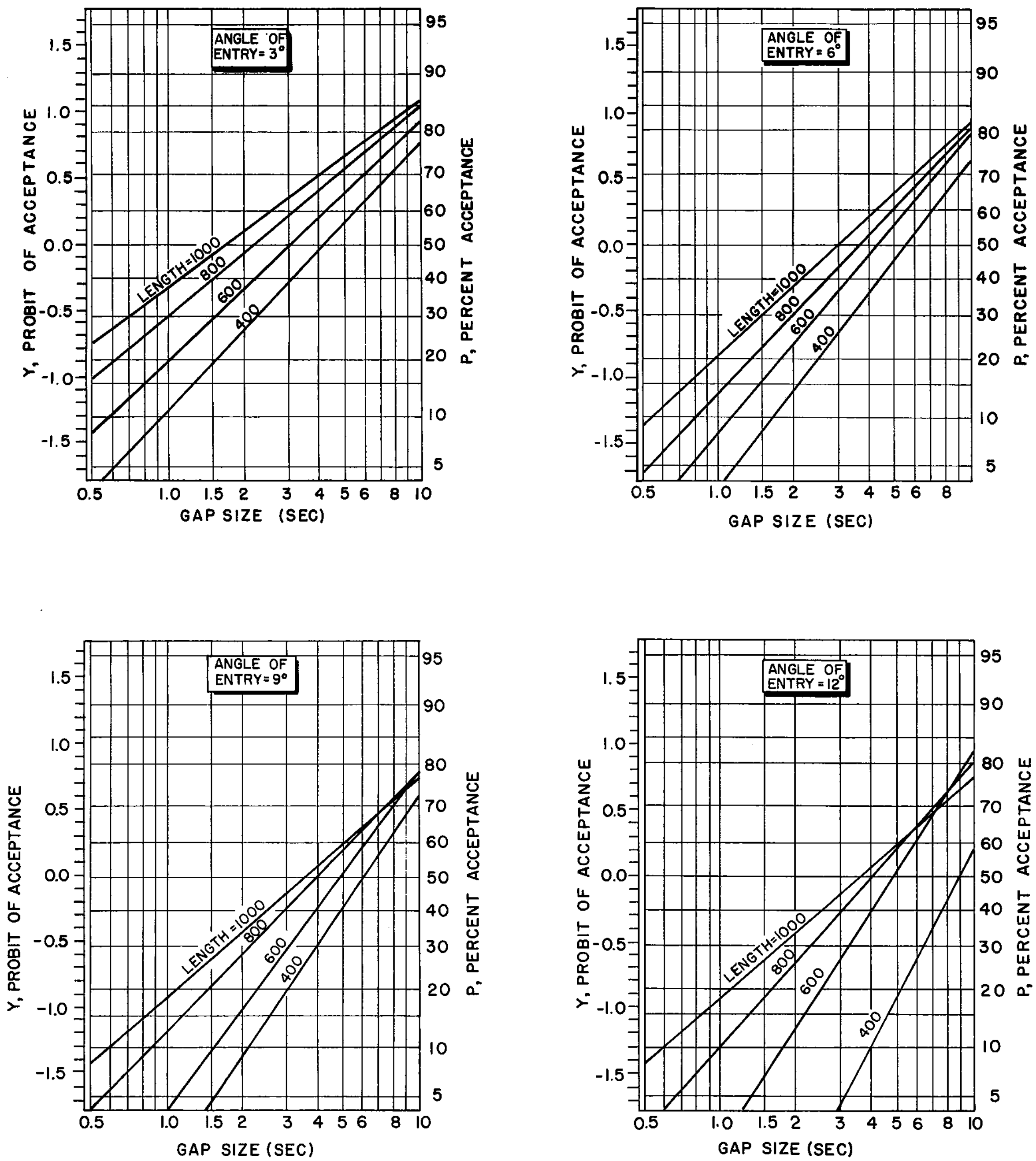




CRITICAL GAP, T, BASED ON ENTRANCE RAMP GEOMETRICS

Figure 9

The gap acceptance characteristics based on the above two regression equations are shown in Figure 10 for different angles of convergence and different lengths of parallel acceleration lane.



GAP ACCEPTANCE CHARACTERISTICS AS RELATED TO ENTRANCE RAMP GEOMETRICS

Figure 10

APPLICATION TO DESIGN

The Capacity-Demand Concept

The most vigorous attempt to eliminate traffic congestion has been the development of the freeway. The freeway provided a new approach; it released the roadway from old alignments, from abutting property, from intersections at grade, from outmoded design standards, and from old right-of-way limitations.

The first generation freeways - those constructed shortly after World War II - left much to be desired. Rollercoaster grade lines and poor lane alignments were features typical of this generation. In many instances, ramps were spaced too closely, were poorly designed, unwisely located and some were without speed change lanes to afford the driver a smooth merge with the freeway traffic. Operational studies were unknown, with the result that designers had little feedback of operational information and that there was little knowledge as to theories of traffic flow. Designers also had very little background information on which to base design. There were no accurate traffic projections available and no information on the type of traffic or the speeds at which vehicles would travel.

The transition from the first to the second generation freeway began with the development of the Interstate System. Designers improved design concepts and practices, and most important, they began to concern themselves with the capacity of facilities and projected traffic volumes. They discovered that they needed, not just four lanes, but six, eights, and sometimes even ten lanes. In defining the second generation freeway, it might be said that it came about from a need or a desire to resolve the capacity-demand problem.

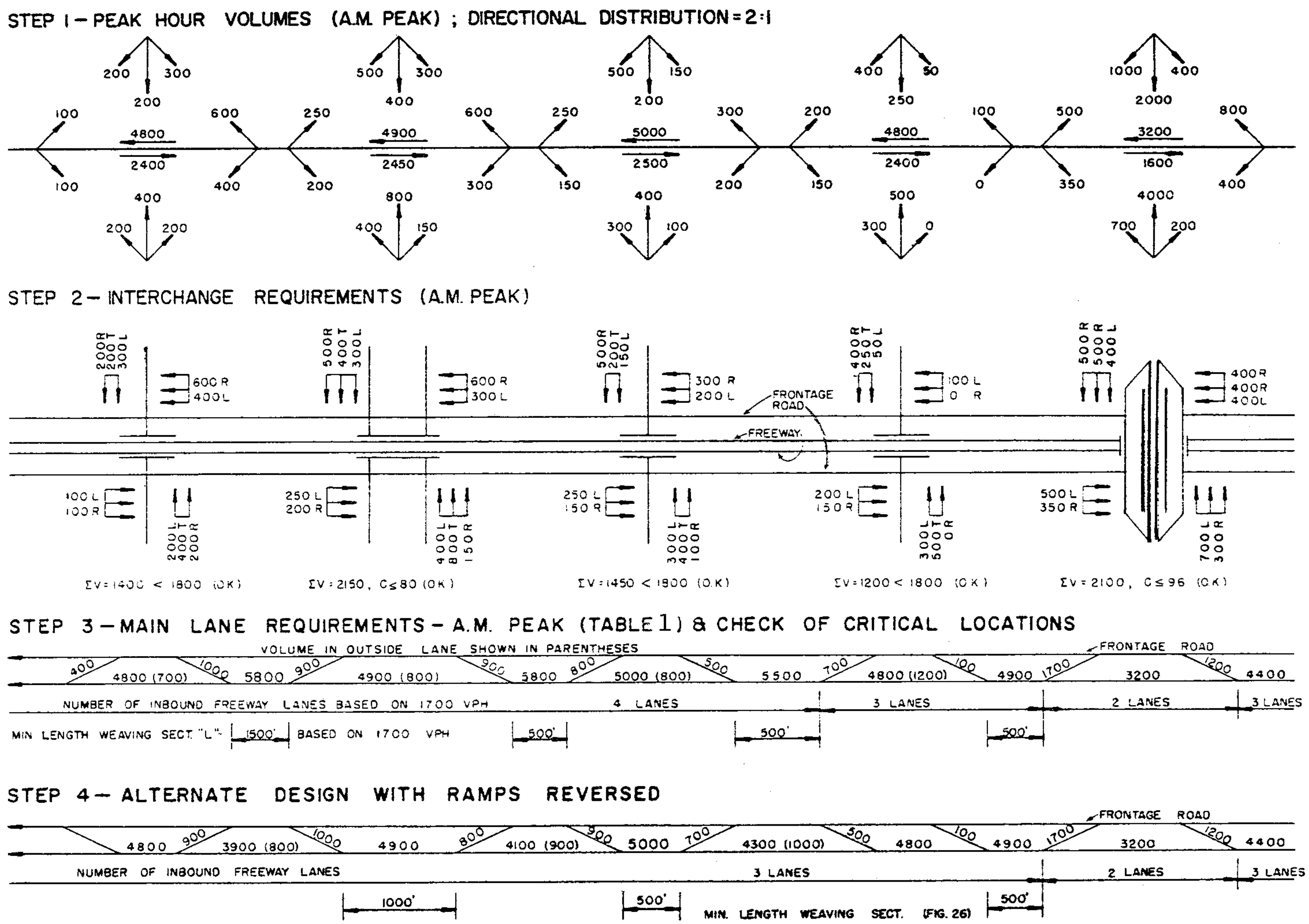
Freeway design is an engineering function - not a handbook problem. The engineer is faced with the problem of predicting traffic demands in future years and providing facilities that will accommodate that traffic under a selected set of operating conditions or level of service. Freeway design, just as any other engineering design problem may be described as a systematic attempt to resolve a capacity-demand relationship at an acceptable level of service. In the design of a layered pavement system, it is attempted to build enough strength into the materials 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 contract also affect the driver's ride and, as such, are level of service factors that must be considered. Although the strength of the materials in a pavement can probably not be estimated as accurately as the capacity of a freeway lane the loads (demand) on the facility are controlled, and in most states limited by law.

The traffic engineer's problem of resolving a capacity-demand relationship is basically similar to that of pavement designer. He must be able to either measure the parameters defining capacity and demand accurately enough to design for them, or he must be able to control them after the facility is designed. It is, of course, impossible to predict traffic demand accurately for some date in the future. Therefore, if urban freeways are to operate at the levels of service for which they were designed, design procedures must be improved and/or the demand on these facilities must also be regulated.

Freeway Ramp Design Procedure

The location of a highway and its design elements, though influenced to some degree by topography, physical features, and land use of the area traversed, should reflect anticipated traffic patterns and travel desires. Consider some average daily traffic volumes and turning movements along a hypothetical route, presumable provided by the Planning Survey Division of the State Highway Department. Figure 11 shows the design hourly volumes for the peak periods, obtained by multiplying the ADT volumes by a given "K" factor (ratio of peak hour to daily traffic) and the directional requirements by multiplying by a 67 percent "D" factor (distribution of peak freeway traffic by direction). The high traffic volumes, urban nature of the route location, and economic considerations suggest a design speed of 70 mph as a basis for controlling all the design elements toward achieving an efficient balanced level of service.

The full control of access feature of freeways automatically requires grade separations for intersecting streets, and interchanges where turning movements are to be provided for. Interchanges may be classified according to the number of legs, a direct or indirect, and as signalized or unsignalized. The three principal types are, of course, the cloverleaf, the directional, and the diamond interchange. The capacities of fully directional and cloverleaf interchanges are essentially determined by the capacities of their ramps. Diamond interchanges are always signalized in urban areas and their capacities therefore depend upon the individual intersections or the coordinated system of intersections.



FREEWAY DESIGN PROCEDURE

FIGURE 11

Though not generally appreciated it is important that the interchange requirements for the freeway schematics be determined before the freeway main lane requirements are investigated, because the number of ramps depends on the choice of interchange. Thus, a cloverleaf interchange and a directional interchange may have one of two entrance ramps and one or two exit ramps in each direction, whereas diamond interchanges have one entrance and one exit ramp in each direction. If the interchange is to be signalized, a capacity check is made to see if the planned facilities will handle the traffic with reasonable cycle lengths. Should a facility be apparently underdesigned, additional approach lanes may be added or a higher type interchange be substituted in its place. For example, if a conventional diamond will not work a three-level diamond should be tried.

The next step is the determination of the number of main lanes based on an analysis of the estimated peak hour demand and the service volume value chosen as the design capacity. The freeway design service volumes in Table 1 enable the designer to judge what level of service can be expected for a given service volume based on the probability of obtaining various types of flow conditions during the peak five-minute period.[10] For example, if the population of the metropolitan area in the design year is taken to be in the 1,000,000 range, then, based on a possible capacity of 2000 vph per lane, Table 1 tells us that a freeway design service volume of 1700 would give a rate of flow of 2000 vph during the peak five-minute period.

After the determination of the number of freeway lanes, 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 inter-vehicular 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.

TABLE 1

FREEWAY CAPACITY WITH CONFIDENCE LIMITS

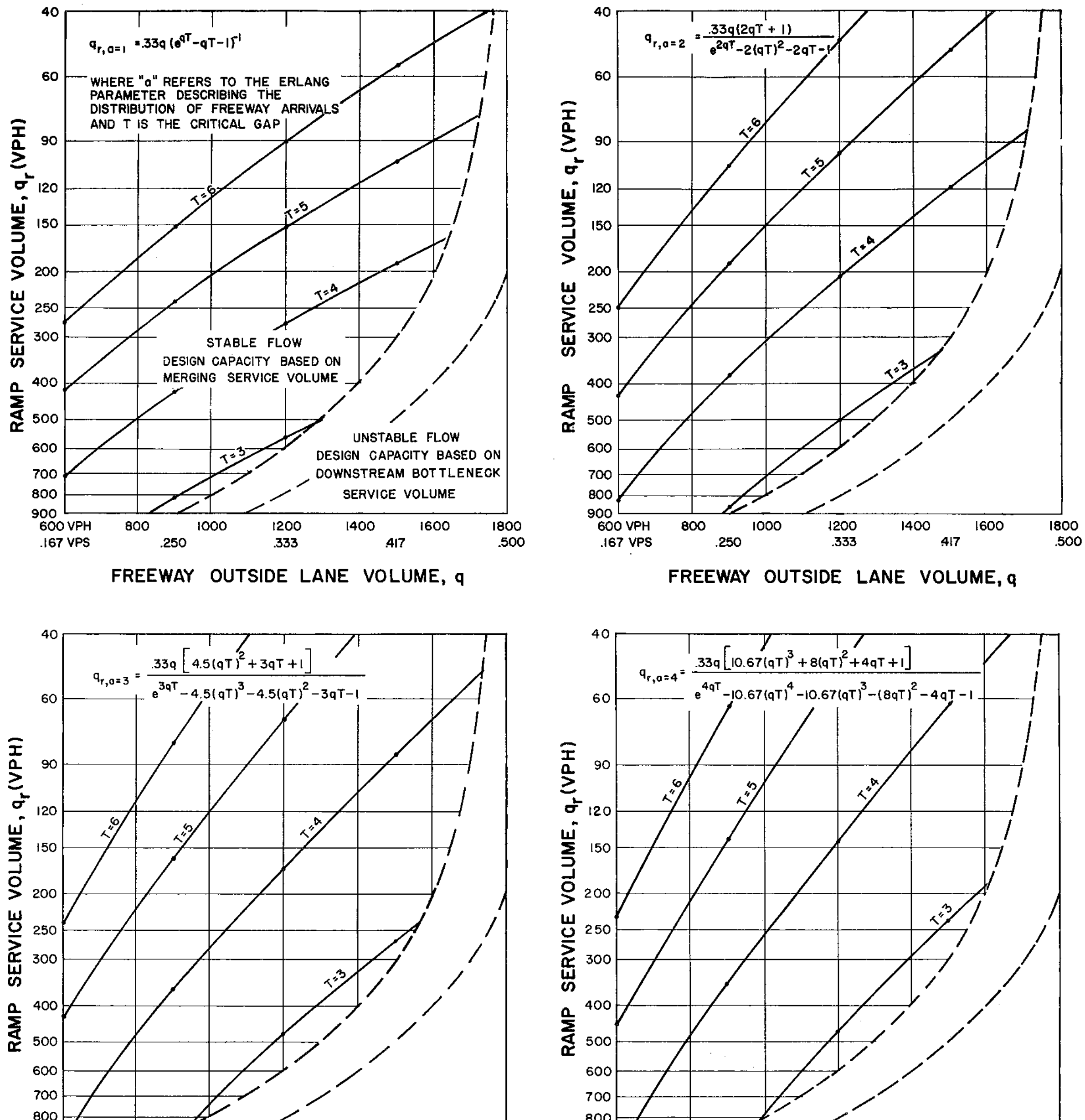
Peak 5-Min Flow (VPH)	Approx. Probabilities of Various Types of Flow in Peak 5-Min			Freeway Design Service Volume (Total hourly vol. /lane) Population of Metropolitan Area			
	Stable	Unstable	Forced	100,000	500,000	1,000,000	5,000,000
1500	1.00	0.00	0.00	1100	1200	1300	1300
1600	0.98	0.02	0.00	1200	1300	1300	1400
1700	0.85	0.15	0.00	1300	1400	1400	1500
1800	0.50	0.48	0.02	1400	1500	1500	1600
1900	0.15	0.69	0.16	1500	1600	1600	1700
2000	0.03	0.47	0.50	1500	1600	1700	1800

There are three basic procedures employed in checking capacity for the design of entrance ramps. One method is based on preventing the total freeway volume upstream of the ramp plus the entrance ramp volume from exceeding the capacity of a downstream bottleneck. A second method 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 discussed in this report 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 12 which is a modification of Figure 3 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. However, if due to the terrain, spacing of interchanges or ramp configuration, some compromise in the geometric design of the ramp-freeway merging area is necessary, then the third method should be employed.

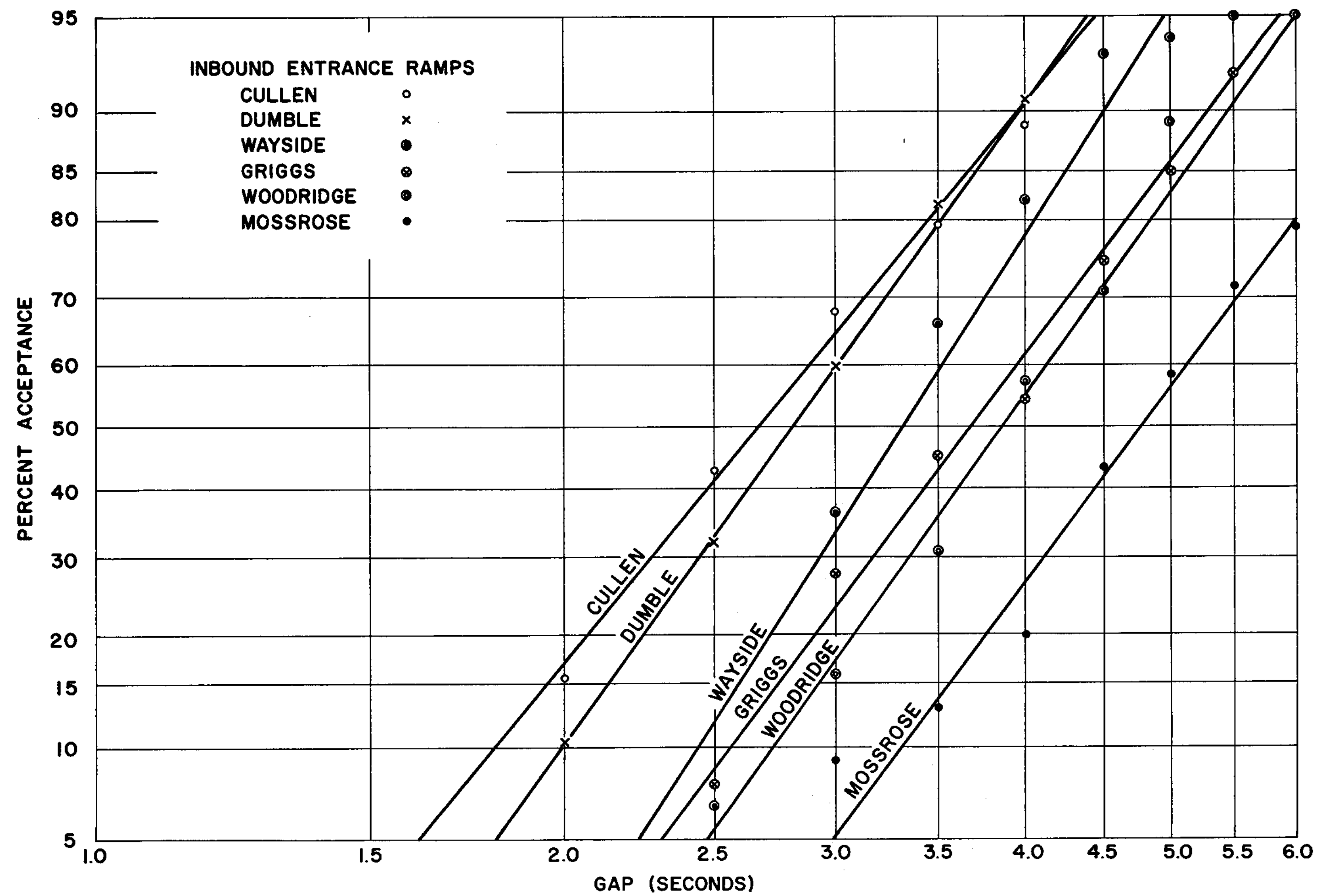
The effect of poor ramp geometrics is evident. Consider the differences in ramp service volumes for a shoulder-lane flow of 1200 vph (a=3 from Figure 4) as the critical gap T increases from 3 to 4 seconds. From the lower left hand graph in Figure 12, one sees that q_r drops from 480 vph to 160 vph. To have used some arbitrary merging service volume (say 1800 vph) in the capacity check for this freeway would have been a dangerous over simplification. Actually, the entrance ramp design capacity curves in Figure 12 or Figure 5 should be employed in order that the individuality of each ramp junction be considered. The values of the critical gap T needed to enter the curves is, of course obtained from Figure 9. The graph of the percent acceptance for merging vehicles at six inbound ramps of the Gulf Freeway in Figure 13 helps dramatize the inportance of considering merging geometrics in a freeway design procedure.

Freeway design is, as are most real world phenomena, 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 the provision for an adequate weaving section between them often offer a dilemma. The alternatives are: (1) reduction in the standards of one or more of the features, (2) elimination of one of the features (such as one of the ramps) or (3) transferring the weaving from the freeway to the frontage road. These alternatives should be evaluated in terms of their cost, their effect on adjacent facilities such as adjacent interchanges, cross street



ENTRANCE RAMP DESIGN CAPACITY CURVES

FIGURE 12



VARIATIONS IN PERCENT ACCEPTANCE FOR MERGING VEHICLES AT DIFFERENT RAMPS

FIGURE 13

signalization, etc. The procedure described in this paper 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.

APPLICATION TO CONTROL

Freeway Surveillance and Control

The term "surveillance" has developed in the highway terminology primarily in the last decade and denotes the observation of conditions in time and space. Initially, urban freeway surveillance was limited to moving police patrols. Recently, helicopters have been used for freeway surveillance in many metropolitan areas. Efficient operation of high density freeways is, however, more than knowing the location of stranded vehicles or the qualitative description of the degree of congestion by high flying disk jockeys. Television surveillance became an operational reality in the late 1950's both in the U.S. and Europe. The Port of New York Authority utilized closed circuit television for monitoring traffic in the Hudson River Tunnels and in Germany, a well publicized TV system was developed to monitor traffic at a major, complex urban intersection.

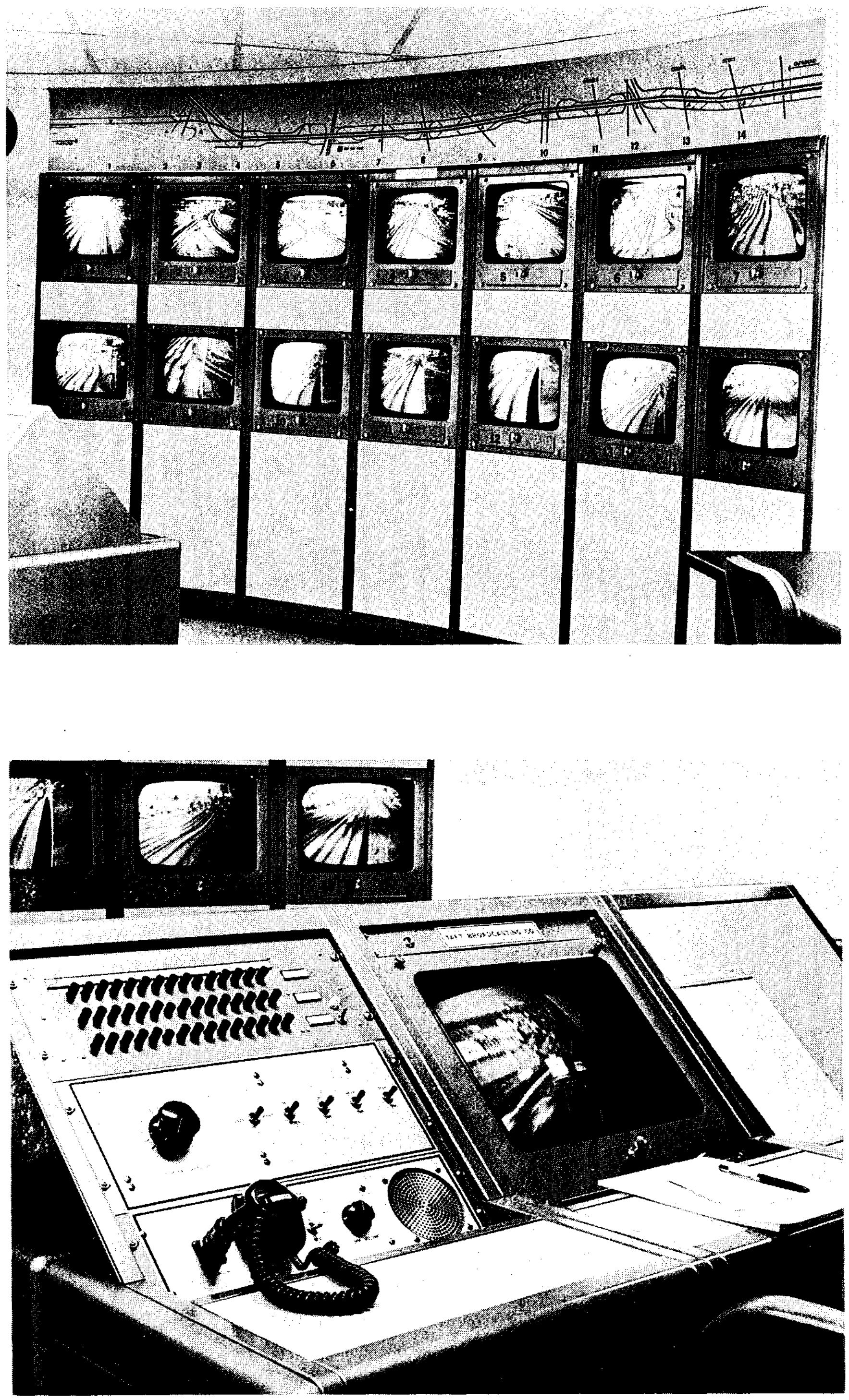
Experimentation with closed circuit television as a freeway surveillance tool was initiated on a three-mile section of the John C. Lodge Freeway in Detroit. This offered the opportunity of seeing a long area of highway in a short, almost instantaneous period of time made possible by spacing cameras along the freeway so that a complete picture could be obtained of the entire section of the roadway. The system was put into use in the summer of 1961.

A similar closed circuit television system now exists on a six-mile section of the Gulf Freeway in Houston. It permits complete surveillance of the traffic flow as well as the expedient handling of accidents or stalled vehicles on the freeway. The television monitors are housed in a central control center shown in Figure 14.

Making better use of traffic facilities has long been a basic concept of the traffic engineer. However, the installation of access controls on freeways to obtain better traffic flow was not originally conceived for these facilities. The rapid growth of traffic demand in our urban areas, coupled with the long-term construction requirements for building an extensive urban freeway system, has required the application of a control concept to freeway operation.

The Evolution of Ramp Control Criteria

When demand exceeds or sometimes only approaches the capacity



MONITOR ARRANGEMENT IN CENTRAL CONTROL CENTER

FIGURE 14

of a system, there is a self-aggravating deterioration of operation and build-up of congestion. In such cases, classical control systems are employed to either make the facility flexible enough to accommodate fluctuations in demand or to reduce the magnitude of the demand fluctuations. Freeway surveillance and control projects are necessarily limited to the latter. One approach, pioneered by the Detroit Project, is to inform the motorist of traffic conditions by utilizing lane controls and variable speed messages. A second and more positive approach is exercised at the point or points of ingress such as the entrance in the case of tunnel control or the on-ramps in the case of freeway control.

Metering, the process of controlling the amount of entering traffic, was developed by the Port of New York Authority. The first step was the identification of the bottleneck at the foot of the tunnel upgrade. Secondly, a mathematical model[11] was formulated to describe the behavior of vehicular traffic in the tunnel. The significant feature of the model was its prediction of shock waves upstream of the bottleneck. The remedy consisted of metering traffic at the entrance of the tunnel to (1) prevent the development of instability by keeping traffic density below some critical value and (2) keeping the traffic demand below the bottleneck capacity.

Based on the success of metering in the tunnel, a similar control plan was formulated for the Eisenhower Expressway by the staff of the Chicago Project. Two bottlenecks on the outbound facility were identified within the study area.[12] The one fartherst upstream is caused by a reduction in the number of lanes from four to three without a corresponding reduction in traffic demand. The second bottleneck, further downstream and the last bottleneck, on the outbound expressway, is caused by fairly heavy on-ramp traffic and is located at the top of an approximate 1000 ft., three percent upgrade.

Two metering techniques were developed. One technique utilized a point density or occupancy measurement on the freeway just upstream of the entrance ramp to be metered; the other utilized a volume measurement on the freeway about one-half mile in advance of the entrance ramp, and an exit ramp volume between the freeway volume measurement and the entrance ramp. After further study, the technique based on occupancy was selected in which a value of fifteen percent occupancy on the center lane was used as a control parameter for initiating metering. From a relation established between the center lane occupancy and the maximum safe ramp volume, a metering rate was established for various levels of occupancy.

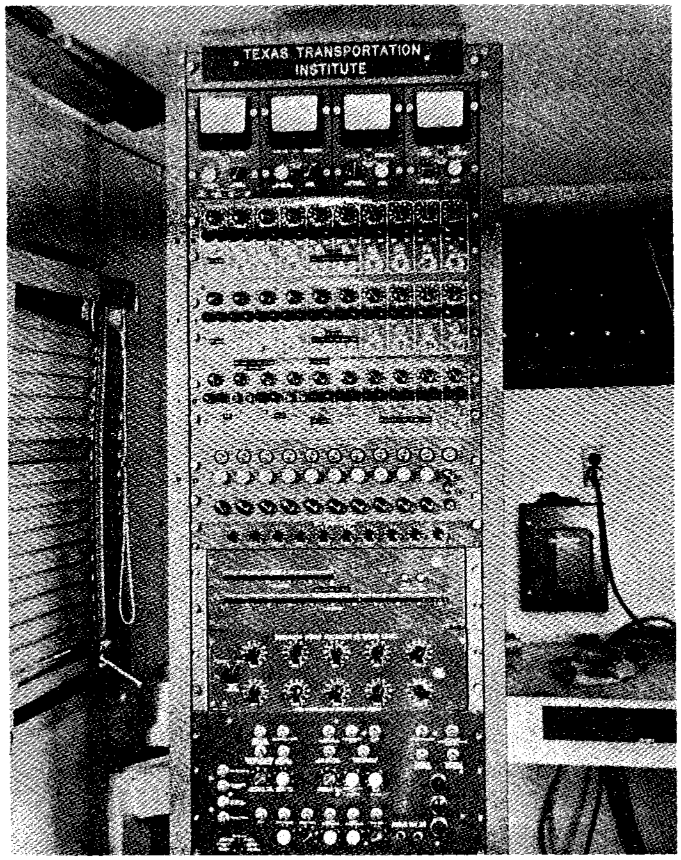
Some researchers who followed the Chicago experiments were more impressed by the use of a freeway capacity-demand relationship as a control parameter for ramp metering. Wattleworth[13] has championed this "capacity-demand" criterion in which an individual ramp would be metered according to the difference between the upstream freeway demand and downstream freeway bottleneck capacity. He has also developed a linear programming model in which several entrance ramps in a freeway system would be metered so as to maximize the output of the system subject to constraints assuring that the demand will not exceed the total directional capacity at each freeway bottleneck.[14]

In a paper presented in 1963, Drew[15] describes a "moving queues" model based on coordinating ramp metering with the detection of acceptable gaps in the outside freeway lane. An acceptable gap is defined as one equal to or larger than the critical gap (that gap for which an equal percentage of ramp traffic will accept a smaller gap as will reject a larger one). Moving queues or platoons occur when the time headway or gap between successive vehicles is less than an arbitrary queueing headway. Since the queueing headway is taken as the critical gap, the number of ramp vehicles to be metered in some time-constant equals the number of moving queues detected. The average number of vehicles per moving queue, as the reciprocal of the probability of a gap larger than the critical gap, provides a rational index of freeway operation. The model has the flexibility of metering a single ramp vehicle per available acceptable gap on the freeway or metering ramp vehicles in bunches or platoons using a "bulk service" technique[16].

Automatic Ramp Control

Before an automatic ramp metering system is designed, its purposes and objectives should be considered. Assuming the proposed ramp metering system is both a research and an operational tool, it should involve the continuous sampling of basic traffic characteristics for interpretation by established parameters, in order to provide a quantitative knowledge of operating conditions necessary for immediate rational ramp control. In short: the system should be traffic responsive, and it should be automatic.

Functional specifications were developed by the Texas Transportation Institute on a companion project sponsored by the Texas Highway Department and Bureau of Public Roads for the controller shown in Figure 15. This controller called the "Gap Acceptance Mode" or Mode I



AUTOMATIC RAMP CONTROLLER

FIGURE 15

detects gaps (or time spacing between vehicles) in the outside lane of the freeway upstream of an entrance ramp and evaluates the size of these gaps with regard to their ability to accommodate a vehicle entering from the ramp. When a desirable freeway gap is detected, it is projected downstream by means of a delay circuit to a point where a waiting vehicle on the entrance ramp can be merged into the gap. At this time, the signal on the ramp (see Figure 16) turns green and releases a ramp vehicle for a smooth merge into the freeway as shown in Figure 17. The functional process followed by the controller is illustrated in Figure 18.

Operation of the Gap Acceptance Mode

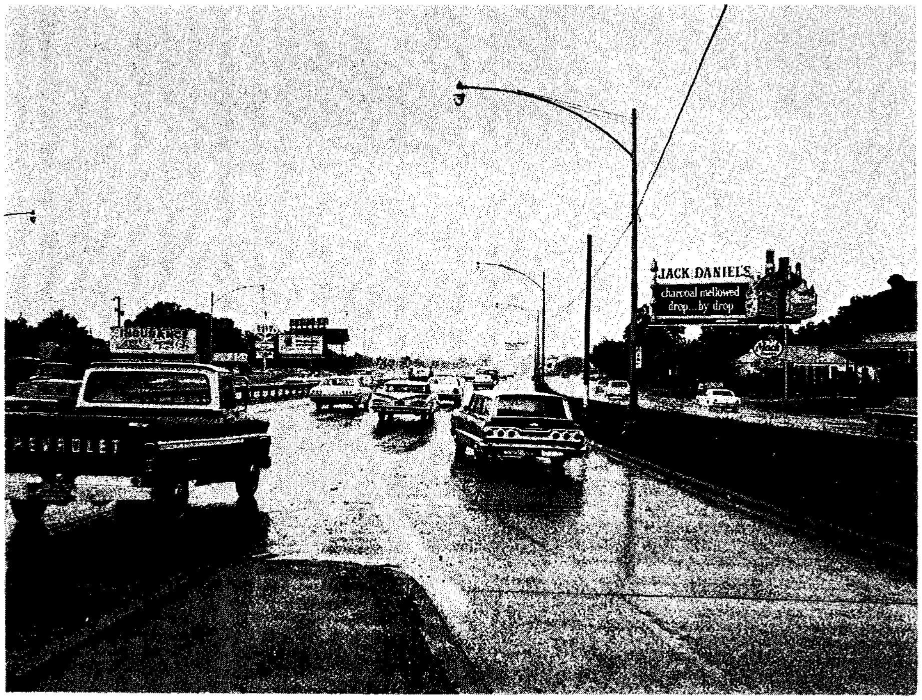
The Gap Acceptance Merging Control Model, designated Mode I, was installed in March 1966 on the Telephone Raod inbound entrance ramp of the Gulf Freeway. The control of the signal is completely automatic. Loop detectors on both sides of the signal provide the calls for the green and red signals. Control is designed for either single vehicle or multi-vehicle entry. The detectors, speed and volume computers and signal controller are rackmounted in the Surveillance Center. The closed circuit television system in the Surveillance Center is used to observe the operation of the signal.

The control of the ramp signal is basically by the detection and projection of acceptable gaps. However, because of the nearby intersection, the length of queue waiting at the signal has a control function. As a safeguard against long delays of the signal due to slowdowns on the freeway lanes, the speed of traffic in the outside lane is a second basis for control. There is also a provision for keeping the ramp area from the signal to the freeway clear. These functions are explained in the following paragraphs[17]. The time-space diagram in Figure 19 has been prepared to complement this description.

Gap Projection - A sonic detector is mounted in a side fire position on a luminaire standard about 950 feet upstream of the ramp nose. (see Note 1 in Figure 19) The detector measures all gaps in the outside lane and calculates the speed of traffic flow (see Note 2 in Figure 19). When a gap is detected that is equal to or greater than the designated acceptable gap size, it is projected in the signal controller at a rate defined by the vehicle speed in the outside lane. If a ramp vehicle is waiting at the ramp signal, a call for the green signal is made when the projected gap reaches the position in time, designated the decision point, at which the travel time of the gap to the merge area is the same as the travel time of the ramp vehicle from the signal to the

RAMP CONTROL SIGNAL

FIGURE 16



RAMP VEHICLE MERGE

FIGURE 17

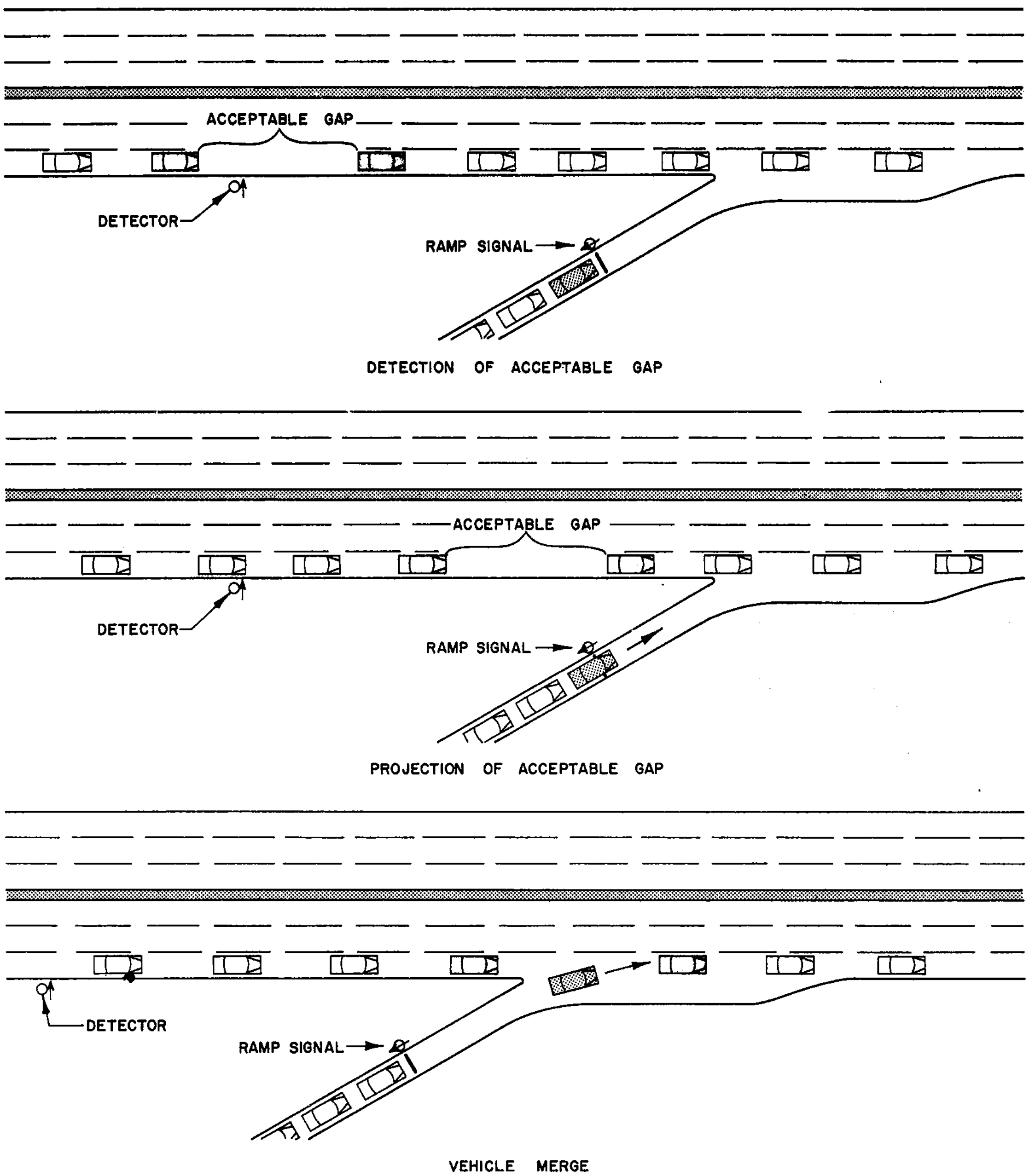
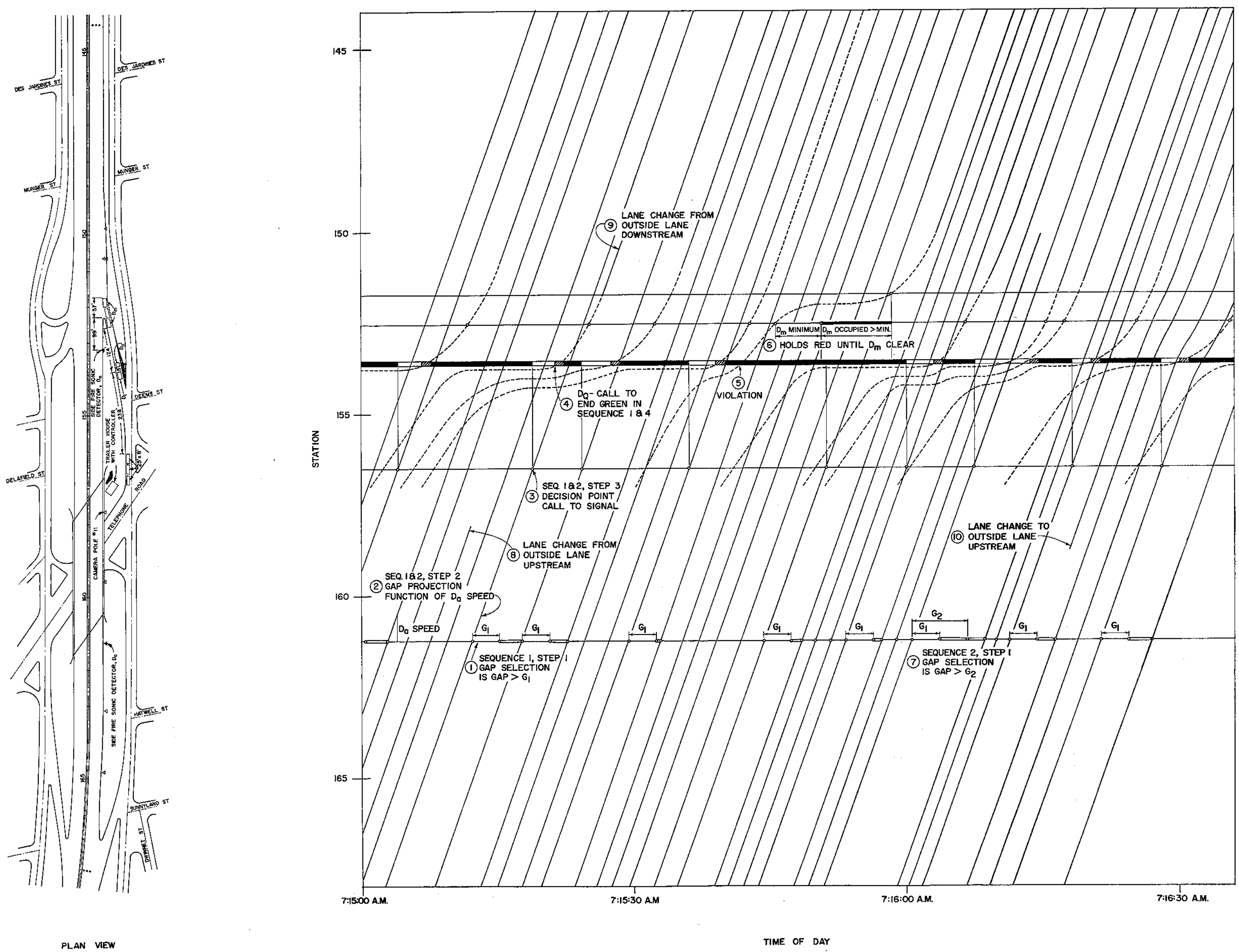


ILLUSTRATION OF GAP ACCEPTANCE MODE OF RAMP CONTROL
FIGURE 18



TIME-SPACE DIAGRAM ILLUSTRATING OPERATION OF GAP ACCEPTANCE RAMP METERING MODE

Figure 19

merge area (see Note 3 in Figure 19). However, the green signal will not be called if there is a ramp vehicle over the merge detector (see Note 6 in Figure 19).

If the gap is equal to or greater than the designated acceptable gap size for more than one vehicle, the controller holds the green signal until the gap passes the decision point (see Note 7 in Figure 8).

Speed of Outside Lane Traffic - A sonic detector mounted in a side fire position on a luminaire standard at the nose of the entrance ramp measures the speed of traffic flow, which is used to select the size of the acceptable gap. When speeds in this area drop below a preset speed, a background cycle rate, set on a fixed rate control, is put into effect. The signal continues to release vehicles when acceptable gaps are available, but it also releases vehicles after a specified waiting time. The difference in this rate and the rate called by the queueing detector described in the next paragraph is that this rate is a minimum setting, and is called when the freeway is in a very slow and congested condition. This control overrides the queueing detector. The fixed rate setting is usually in the range of 150 to 200 vehicles per hour.

Length of Queue - A loop detector is placed in the pavement of the left lane of the inbound frontage road near the Telephone Road intersection. If the queue at the ramp signal is greater than 14 or 15 vehicles, the detector is actuated, and a background cycle rate, set on a guaranteed rate control, is put into effect. The signal continues to releases vehicles after a specified waiting time. This rate stays in effect only as long as the queueing detector is timed out. The guaranteed rate setting is usually of the order of 500 to 600 vehicles per hour.

Occupancy of the Merge Area - A loop detector is placed in the pavement of the ramp just upstream of the merging area. All vehicles entering the freeway from the ramp will actuate the detector. If a vehicle stops on the ramp in this area, blocking the entrance to the freeway, the detector will time out and the controller will hold the signal in red until the detector is cleared (see Note 6 in Figure 19).

Ramp Metering vs Merging Control

The interest of research lies in the subtle blending of theory and experiment. Scientific theories excite the curiosity; they are always useful, occasionally they may be even beautiful. Based on theory, we predict as precisely as possible what should happen in some new experiment. While carrying out the appropriate test, we are hoping that the theory will work, and our moment of triumph occurs when our theory has predicted some new phenomenon accurately for the first time. On the other hand, if the theory is not ours but a rival theory, there is, added to our own curiosity in setting up the experiments to test it, a sense of rivalry. We now plan and carry out the experiments hoping to disprove this theory; similarly, the rival researcher is planning to try to disprove our theory. Attempting to disprove theories in this way is a very important part of scientific endeavor which can be compared to the role of the opposite party in government. This common interest of many researchers in the same phenomenon causes some new theories to be reputed and others, after repeated testing, to be accepted and as such is necessary to the growth of science.

The significance of the Gap Acceptance Merging Control Model lies in its conceptual appeal. Note the use of the term "merging control" rather than "ramp control" in describing the model. The Gap Acceptance Mode is the only metering system that aids the ramp driver in the merging maneuver. This is important and shall be explained in more detail.

When the volume of traffic on a freeway begins to approach capacity, the merging driver is placed in an extremely difficult position. The number of acceptable gaps in the freeway stream decrease sharply as the freeway volume increases. At these higher volumes, the merging driver cannot always defer his decision to merge until he is on the acceleration lane. Instead he must detect the location of gaps in the oncoming stream before he reaches the acceleration lane. Operating in this manner, he must project the progress of a gap onto the acceleration lane in order to decide whether or not it will be available to him. This in turn requires that he estimate his own speed and acceleration as well as the speed and size of the gap in order to decide whether there will be sufficient space for the merging maneuver to be completed successfully within the limit of the acceleration lane.

Michaels and Weingarten[18] do not think it is possible for the driver under these circumstances to reliably solve the appropriate equations of motion. They state:

> "It is obvious that as the main stream volume approaches capacity, the merging driver's task becomes for all practical purposes impossible. Thus, effective ramp metering will require the equations of motion to be solved automatically whenever a vehicle enters a ramp. Mathematically, the problem is

quite simple requiring a knowledge of the location of gaps and their speed. Knowing something about the accelerating capability of the ramp vehicle and the length of the ramp and acceleration lane, a perfectly determinate solution is possible. Instrumentation to carry out these operations is well within the state of the art of existing electronic technology."

The gap oriented system installed at the Telephone Road interchange locates freeway gaps and their speeds, compares these gaps to a "critical gap", takes into account the accelerating capability of the ramp vehicles and the length of the ramp and acceleration lane and solves the equations of motions automatically before the metering signal is actuated to allow a vehicle, or vehicles, to make a smooth merge. In addition to the increased efficiency so obtained, other factors such as safety and higher ramp capacity are improved for a comparatively low installation cost.

Considering safety, any speed differential at a point in the traffic stream in either a longitudinal or transverse direction is dangerous. A vehicle that stops in a travelled lane is in particular danger; it is a safety hazard to the remaining traffic, to its driver, and to its occupants. This is indicated by by the high percentage of accidents that are of the rear-end type occurring at induced stop and yield locations such as in the freeway merging area. The Gap Acceptance Mode virtually eliminates ramp vehicles stopped in the merging area, thereby contributing greatly to safe operation. In addition, the system affords the opportunity for increased ramp capacity over other metering models. In systems which meter ramp vehicles one at a time, the ramp capacity is obviously a function of the ramp cycle length. Since it takes about four seconds to go through the ramp signal cycle - the maximum metering is at a rate of one vehicle every four seconds or 900 vph. The Mode I system can meter at a faster rate because it has the flexibility to meter more than one ramp vehicle whenever large freeway gaps are detected.

In conclusion, the Gap Acceptance Mode provides the merging driver with the necessary information to know that a sifficient gap is available. Second, because of its nature, it is also a metering system. Such a dynamic merge aiding and metering technique appears to be a very attractive and inexpensive way of maintaining high efficiency of flow on a freeway and at the same time of obtaining maximum ramp capacity and merging safety.

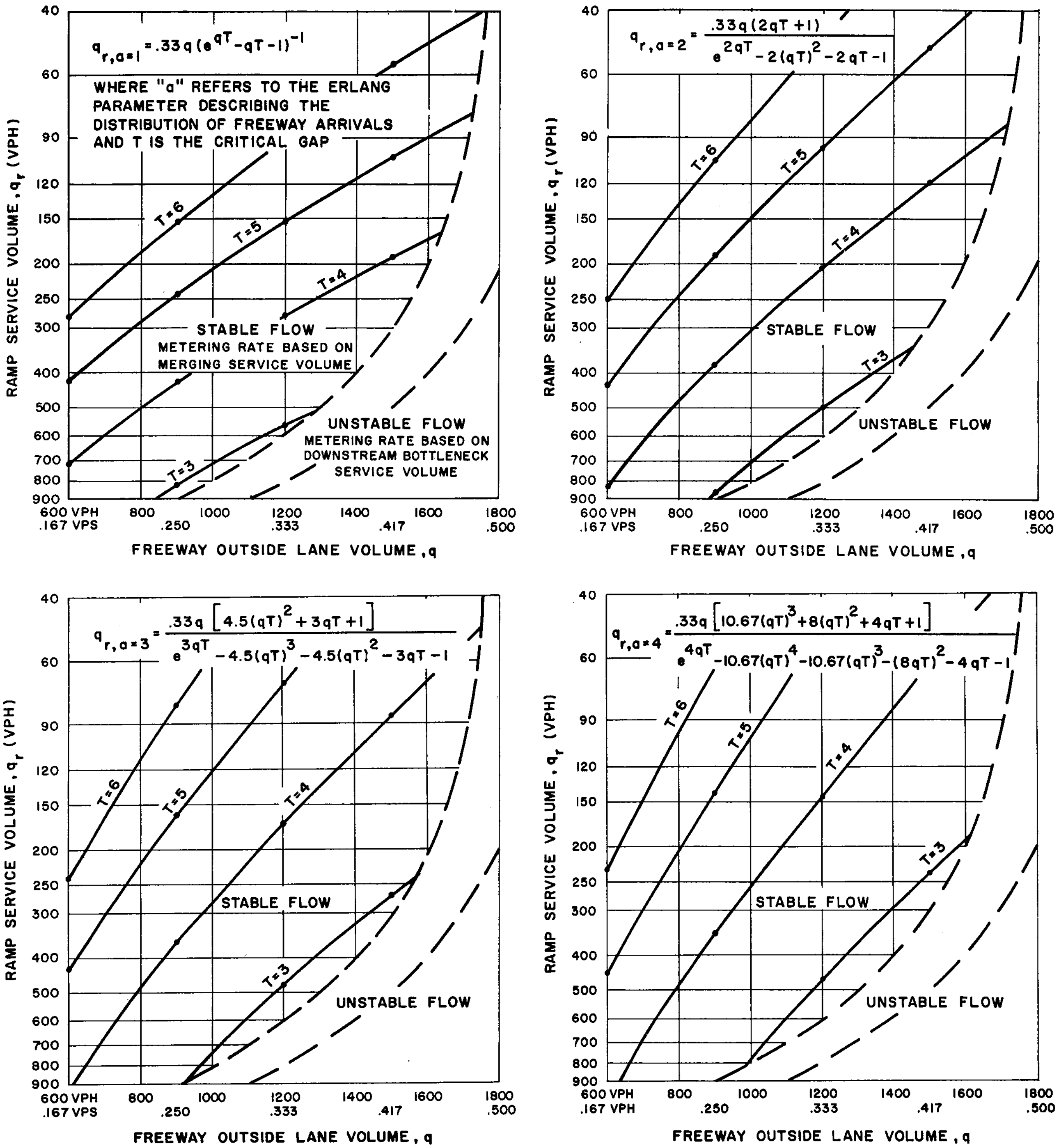
Integrated Freeway Control

With the exception of the gap acceptance or merging control model, all the ramp metering techniques in use today may be classified as macroscopic in nature. For example, in the capacity-demand criterion of ramp metering, a ramp is metered according to the difference between the upstream freeway demand and downstream bottleneck capacity. In other words, steady state stability is maintained as long as the demand-capacity ratio is less than unity. On the other hand, the merging control criteria is microscopic in nature since it considers each freeway gap and each ramp vehicle individually.

The graphs in Figure 20 point out the differences in the macroscopic philosophies or ramp metering and the microscopic merging control 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 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 20 clearly illustrates the need for a ramp control technique combining both the macroscopic and microscopic approach. For conditions described on the graph by "unstable 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 1800 vph line dividing stable and unstable flow, the critical gap governs since the merging service volume is less than the bottleneck service volume.

Looking ahead, it is well known that optimization of a part of a system or subsystem does not necessarily lead to the optimum solution for the entire system. Similarly, optimizing the operation of a single merging control system may not necessarily lead to the optimization of the overall system. The entrance ramp control curves in Figure 20 afford the flexibility of controlling all the ramps in a freeway system according to either the individual merging areas or the downstream bottlenecks.



ENTRACE RAMP CONTROL CURVES

Figure 20

REFERENCES

1. "Highway Capacity Manual 1965," Highway Research Board, Special Report 87, 1965.

2. Hess, J.W., "Capacities and Characteristics of Ramp-Freeway Connections." Highway Research Record, No. 27, 1965.

3. Hess, J. W., "Ramp-Freeway Terminal Operation as Related to Freeway Lane Volume Distribution and Adjacent Ramp Influence," Highway Research Board No. 99, 1965.

4. Moskeowitz, K., and Newman, L., "Traffic Bulletin No. 4 - Notes on Freeway Capacity," Highway Research Record No. 27, 1963.

5. Buhr, J. H., Drew, D. R., Wattleworth, J.A., and Williams, T. G., "A Nationwide Study of Freeway Merging Operations," Texas Transportation Institute Research Report 430-1, Jan. 1967.

6. Drew, D.R., LaMotte, L.R., Buhr, J.H., and Wattleworth, J.A., "Gap Acceptance in the Freeway Merging Process," Texas Transportation Institute Research Report 430-2, Jan. 1967.

7. Wattleworth, J.A., Buhr, J.H., Drew, D.R., and Gerig, F.A., "Operational Effects of Some Entrance Ramp Geometrics on Freeway Merging," Texas Transportation Institute Research Report 430-3, Jan. 1967.

8. Drew, D.R., Wattleworth, J.A. and Buhr, J.H. "Gap Acceptance and Traffic Interaction in the Freeway Merging Process," Proceedings of the Bureau of Public Roads Research and Development Program Review Meeting, Washington, D.C., December, 1966.

9. Major, N.G. and Buckley, D.J., "Entry to a Traffic Stream," Australian Road Research Board Proceedings, Vol. 1, Part 1, pp. 206-228, 1962.

10. Keese, C.J., Pinnell, C., Drew, D.R., "Highway Capacity the Level of Service Concept," Highway Research Record 99, 1965.

11. Greenberg, H., "An Analysis of Traffic Flow," Operations Research No. 7, 1964.

12. May, A. D., "Experimentation with Manual and Automatic Ramp Control," Highway Research Record No. 59, 1964.

13. Wattleworth, J . A., "Peak-Period Analysis andControl of a Freeway System," Texas Transportation Institute, Research Report 24-15.

14. Wattleworth, J. A., "Peak-Period Control on a Freeway System-Some Theoretical Considerations," DoctoralD issertation, Northwestern University, 1963.

15. Drew, D . R., "Theoretical Approaches to the Study and Control of Freeway Congestion," Texas Transportation Institute, Research Report 24-1.

16. Drew, D. R., "Gap Acceptance Characteristics for Ramp-Freeway Surveillance and Control," Texas Transportation Institute Research Report, 24-12.

17. McCasland, Wm. R., "Operation of Gap Projection Controller Mode I," Unpublished.

18. Michaels, R. M., Weingarten, H., "Driver Judgment in Gap Acceptance," U.S. Bureau of Public Roads, Washington, D. C., Unpublished.