Chapter 3

QUEUEING THEORY APPROACHES

D. E. Cleveland
Texas A&M University

and

D. G. Capelle
Automotive Safety Foundation

Contents

Section

Chapter 3
QUEUEING THEORY APPROACHES

3.1 INTRODUCTION

Highway and traffic engineers are charged with many responsibilities. They must work to reduce motor vehicle accidents. But they also design and operate highway systems which minimize delay for the traveling public.

Delay is a direct product of congestion. Therefore, a fundamental understanding of delay is necessary to obtain the greatest efficiency from existing and planned highway systems.

An observer of traffic on a highway network cannot help but be impressed by the variability which he sees. Vehicles of different types operated by drivers with different desires and characteristics are seen in varying numbers. The action of any one driver can create minor or serious congestion problems. It is extremely difficult to take into account all the information needed to predict the detailed operation of such a system.

Variable phenomena of this type are called "stochastic" phenomena, and the methods of probability and statistics provide a means by which it is possible to predict some delay characteristics. For example, knowledge of the characteristics of arrival of main-street traffic and pedestrian crossing demand can be used to predict delays to pedestrians, thus helping to establish improvements and warrants for the installation of traffic control devices.

Probability models of congestion can vary in complexity. Some simple models do a rather poor job, which is not surprising. On the other hand, there are simple models, for which solutions are readily derived, which do surprisingly well in predicting delays observed in the field. As the models are made more complex to account for such things as driver variability, the solutions become more difficult. It must always be remembered that mathematical descriptions of system operations rarely account fully for observed behavior and that the results of mathematical analysis must be viewed critically.

The purpose of this chapter is to present some of the results of studies of probability models of traffic delay. Section 3.2 briefly describes some fundamental characteristics of variable processes, as well as the important assumptions governing the arrival of streams of traffic at a given point and the variability of gap acceptance of drivers and pedestrians attempting to cross a traffic stream. Section 3.3 presents a brief summary of some elements of queueing or waiting-line theory, that branch of mathematics dealing with congested systems. Sections 3.4, 3.5 and 3.6 present summaries of the most significant published works relative to delays at signalized and stop-sign controlled intersections, passing on a two-lane roadway, and a number of special topics such as multiple queues, parking, and one-lane bottlenecks.

The original papers upon which this chapter is based are generally available in journals found in the collection of a good university library. The interested reader can obtain these for further study. The chapter necessarily avoids detailed mathematical development, but does present the theorist's assumptions and some results of interest.

Those interested in studying probabilistic approaches to traffic flow theory should have access to the work of Haight, of the Institute of Transportation and Traffic Engineering, University of California, Los Angeles, who recently published a book on mathematical theories of traffic flow (*27*). The reader is referred to that source for further development.

3.2 TRAFFIC DISTRIBUTIONS

Highway traffic characteristics are statistical rather than deterministic in nature. Therefore, traffic variables, such as volume, speed, delay, and headways, can be described by probability distributions. Examples of "discrete" probability distributions which occur frequently in traffic applications have been given special names such as "binomial distribution," "Poisson distribution" and "geometric distribution." Similarly, familiar examples of "continuous" probability distributions are the "exponential distribution" and "normal distribution." Some fundamentals of probability distributions are discussed in Section 3.2.1. Several important traffic flow distributions are described in Section 3.2.2. Section 3.2.3 presents information on gap acceptance distributions for pedestrians and drivers waiting to cross or merge with a conflicting traffic stream.

3.2.1 Fundamentals

Probability distributions can be described in terms of three important parameters:

(a) The frequency function $f(t)$.
(b) The mean $\bar{t}$ or $E(t)$.
(c) The variance $Var(t)$.

The Poisson distribution is frequently used as a model to determine the distribution of vehicular traffic on a highway. Outlined in the following are a few generalized mathematical relationships describing this distribution.

If $P(n|q\,T)$ is the probability of exactly n arrivals in T seconds and q is the traffic flow (see Fig. 3.1),

$$P(n|q\,T) = \frac{(q\,T)^n\,e^{-qT}}{n!} \qquad (3.1)$$

The probability of no arrivals $(n = 0)$ in time T becomes

$$P(0|q\,T) = e^{-qT} \qquad (3.2)$$

If there are no arrivals in a particular interval T, there must be a time gap or headway of at least T seconds between the last previous arrival and the next arrival. In other words, $P(0|q\,T)$ is also the probability of a headway equal to or greater than T, or

$$P(h \geq T) = e^{-qT} \qquad (3.3)$$

The probability of a headway less than or equal to any time t is (see Fig. 3.2)

$$P(h < t) = 1 - e^{-qt} \qquad (3.4)$$

usually called the "cumulative distribution function" of the variable t. The function $f(t)$, defined when the cumulative distribution function is differentiable, is called the "probability density function" of t. Thus, differentiating Eq. 3.4 gives the frequency function or probability density function for the exponential distribution (see Fig. 3.3):

$$f(t) = q\,e^{-qt} \qquad (3.5)$$

Some immediate consequences for any variable t with probability density function $f(t)$ are

$$\int_{-\infty}^{\infty} f(t)\,dt = 1 \qquad (3.6)$$

or, in other words, the summation of all probabilities is unity.

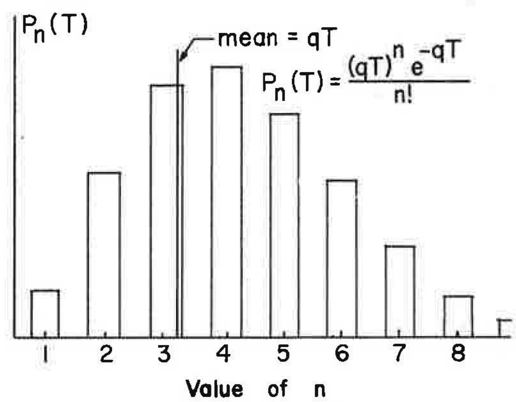


Figure 3.1. Poisson distribution.

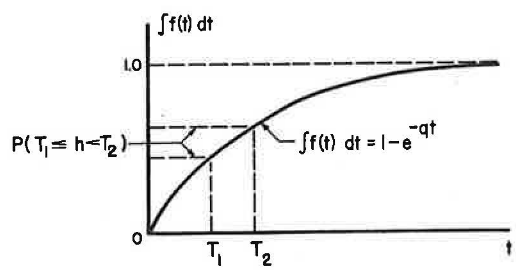


Figure 3.2. Cumulative exponential distribution.

A probability density function which describes the chances that a headway will lie in any range of values between T_1 and T_2 is

$$P(T_1 < h < T_2) = \int_{T_1}^{T_2} f(t)\,dt \tag{3.7}$$

and, for the exponential distribution, substituting Eq. 3.5 in Eq. 3.7 gives (Fig. 3.2)

$$P(T_1 < h < T_2) = \int_{T_1}^{T_2} q\,e^{-qt}\,dt \tag{3.8}$$

Eq. 3.7 may be extended to any variable, such as delay, and any probability distribution, such as the normal distribution.

The general expressions for mean and variance for any distribution are

$$\bar{t} = E(t) = \int_{-\infty}^{\infty} t\,f(t)\,dt \tag{3.9}$$

and

$$\mathrm{Var}(t) = \int_{-\infty}^{\infty} (t-\bar{t})^2\,f(t)\,dt \tag{3.10}$$

Substituting Eq. 3.5 in Eqs. 3.9 and 3.10, the mean and variance for the exponential distribution are

$$\bar{t} = 1/q \tag{3.11}$$

and

$$\mathrm{Var}(t) = 1/q^2 \tag{3.12}$$

These parameters have significance as measures of central tendency and dispersion, respectively. However, these are incomplete descriptions of a probability distribution, and the frequency function or cumulative distribution function is needed to describe completely the characteristics of the variable.

3.2.2 Gap Distributions

The Poisson distribution is the main theoretical instrument for determining the distribution of vehicular traffic on a highway. The assumption leading to a Poisson distribution is that the total number of arrivals during any given time interval is independent of the number of arrivals that have occurred prior to the beginning of the interval. It can be shown that when the Poisson theory is applied to the distribution of time spacings, h, between adjacent vehicles, the exponential distribution results are

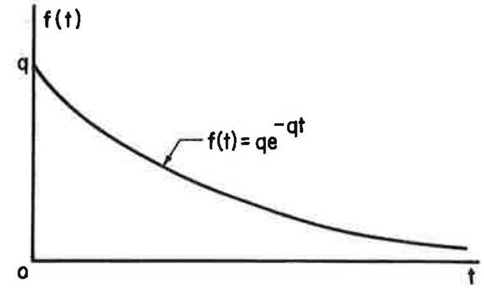


Figure 3.3. Exponential distribution.

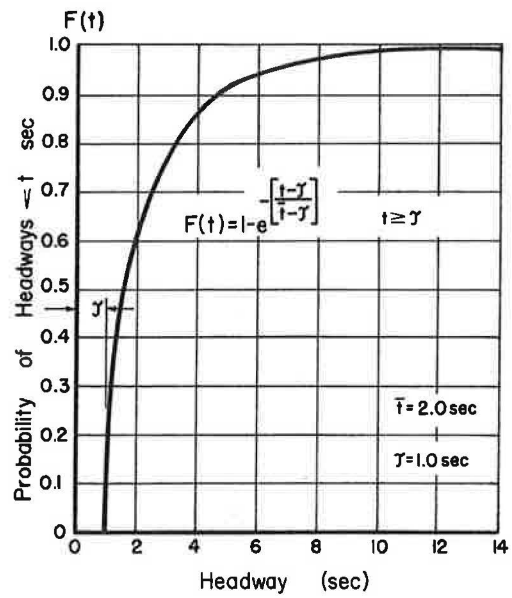


Figure 3.4. Shifted exponential distribution.

$$P(h \geq t) = e^{-t/\bar{t}} \tag{3.13}$$

and

$$P(h < t) = 1 - e^{-t/\bar{t}} \tag{3.14}$$

Although the results yielded by these equations agree well enough with actual observations for low free-flowing traffic volumes, they differ greatly from observations of high-volume conditions for the following reasons:

(a) Vehicles are not points; they possess length and must follow each other at some minimum safe distance.
(b) Vehicles cannot pass at will.

The first difficulty can be partially overcome by shifting the exponential curve to the right by an amount equal to a certain minimum headway τ. This, in effect, states that the probability of a gap between successive vehicles of less than τ is zero, or (Fig. 3.4)

$$P(h \geq t) = \exp[-(t-\tau)/(\bar{t}-\tau)] \qquad h \geq \tau \qquad (3.15)$$

and

$$P(h < t) = 1 - \exp[-(t-\tau)/(\bar{t}-\tau)] \qquad (3.16)$$

In considering the second difficulty regarding passing, Schuhl (*66*) proposed that the traffic stream be considered as composed of a combination of free-flowing and constrained vehicles each of which conforms to a Poisson behavior. This traffic stream is described by

$$P(h \geq t) = (1-\alpha)\exp(-t/\bar{t}_1) + \alpha\exp[-(t-\tau_2)/(\bar{t}_2-\tau_2)] \qquad (3.17)$$

and

$$P(h < t) = (1-\alpha)[1-\exp(-t/\bar{t}_1)] + \alpha(1-\exp[-(t-\tau_2)/(\bar{t}_2-\tau_2)]) \qquad (3.18)$$

in which $\bar{t}_1$ is the average headway of free-flowing vehicles, $\bar{t}_2$ is the average headway of constrained vehicles, τ_2 is the minimum headway of constrained vehicles, and α and $(1-\alpha)$ are the fractions of total volume made up of constrained and free-flowing vehicles, respectively. Figure 3.5 represents Schuhl's plots of Eqs. 3.17 and 3.18 for a total volume of 900 vehicles evenly distributed between free-flowing and constrained vehicles, using arbitrary values of $\tau_2 = 0.5$, $\bar{t}_2 = 2.0$ sec, and $\bar{t}_1 = 6.0$ sec.

Kell (*39*) has generalized Eqs. 3.17 and 3.18 by assuming that a minimum headway τ_1 exists for free-flowing vehicles as well as for the constrained vehicles. This leads to

$$P(h \geq t) = (1-\alpha)\exp[-(t-\tau_1)/(\bar{t}-\tau_2)] + \alpha\exp[-(t-\tau_1)/(\bar{t}_2-\tau_2)] \qquad (3.19)$$

A theoretical distribution for the entire traffic stream, which is essentially a summation of two subdistributions, has been referred to in the literature as the composite Poisson or composite exponential distribution. Morse (*50*) termed the special case of the distribution, described by Eq. 3.19, in which there is no shift ($\tau_1 = \tau_2 = 0$), the "hyper-exponential distribution." One of its discrete distributions was named the "hyper-Poisson."

Haight (*29*) suggested that gaps less than the minimum headway, τ, should be considered improbable, but not impossible. The exponential and hyper-exponential distributions, on the other hand, represent curves which find their maximum probabil-

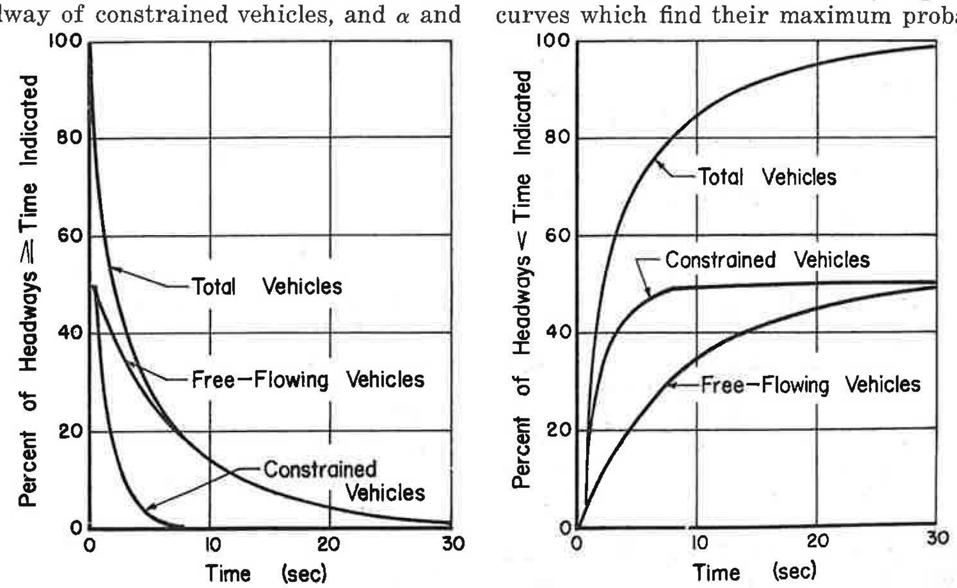


Figure 3.5. Schuhl's composite exponential distribution.

ity at the origin and then decline as t approaches infinity. This implies, erroneously, that the smaller the gap the more likely it is to occur. To overcome these difficulties, the Pearson Type III gap distribution is proposed. This distribution is sometimes called the Erlang or gamma distribution, a two-parameter generalization of the exponential family obtained by multiplying the function in Eq. 3.5 by some appropriate power of t. Thus (Fig. 3.6),

$$f(t) = \frac{t^{K-1}}{(K-1)!} q^K e^{-qt} \qquad (3.20)$$

If $K = 1$, Eq. 3.2 is obtained. As K goes to infinity the variance approaches zero, which suggests a constant rate of flow corresponding to high volumes of traffic. Thus Eq. 3.20 represents the distribution of vehicles for all cases between randomness and regularity. The associated discrete distribution is called the generalized Poisson distribution. It states that the probability of no arrivals in the interval T is the sum of the first K terms of some Poisson series; that the probability of one arrival is the sum of the next K terms of the same Poisson series; etc. Stated mathematically,

$$P(n|q\,T) = \sum_{j=nK}^{(n+1)K-1} \frac{(qT)^j e^{-qT}}{j!} \qquad (3.21)$$

In order to apply Eq. 3.21, one must decide on a value of K. This estimation of parameters, as well as a more complete treatment of the generalized Poisson distribution, has been discussed by Haight (*29, 22*).

It is apparent that the correspondence between gap (continuous) and counting (discrete) distributions has great practical significance, as it is much easier to count vehicles than it is to measure gaps. There are two techniques for measuring the counting distribution in the field. In the usual procedure, traffic counts are started and terminated at given clock times independent of traffic flow. This is referred to as the *asynchronous* case. The second technique, the *synchronous* case, occurs when the counting period starts immediately following the arrival of a vehicle. Except for the case of random flow, the two counting distributions are never the same. The synchronous counting distribution is often referred

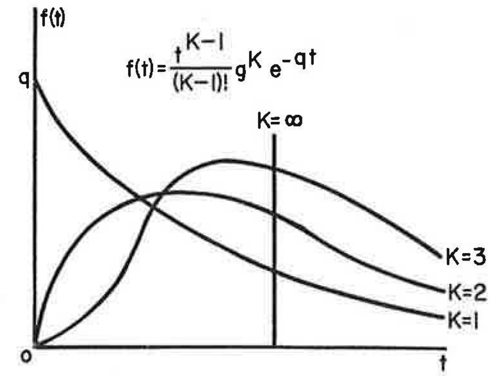


Figure 3.6. The Erlang gap distribution.

to as the generalized Poisson (*22*) and has also been studied by Goodman (*19*) and Oliver (*58*). The asynchronous distribution was studied by Morse (*50*) and has been discussed by Jewell (*36*). A comparison of the two is given by Whittlesey and Haight (*76*).

Figure 3.7 is a time-space diagram illustrating the synchronous counting procedure. Two locations are considered: location A at a point downstream from a signalized intersection timed such that arrivals at A can be assumed to be regularly spaced, and location B far enough downstream from A so that arrivals are random (Poisson). This illustrates that the mean rates of arrivals at A and B are equal to the number of arrivals n divided by total time T. Thus,

$$q = q_a = q_b = n/T \qquad (3.22)$$

Because the chance of occurrence of an arrival at B is independent of the time of the preceding arrival according to the assumptions of a Poisson distribution, the probability of no arrivals in time t is the same for both the synchronous and asynchronous cases, and is

$$P_0(t) = e^{-qt} \qquad (3.23)$$

However, at location A the probability of no arrivals in the counting interval t for the synchronous case depends on whether or not t is less than or equal to and greater than $\bar{t}$,

$$P_0(t) = 1 \qquad (t < \bar{t}) \qquad (3.24)$$

$$P_0(t) = 0 \qquad (t \geq \bar{t}) \qquad (3.25)$$

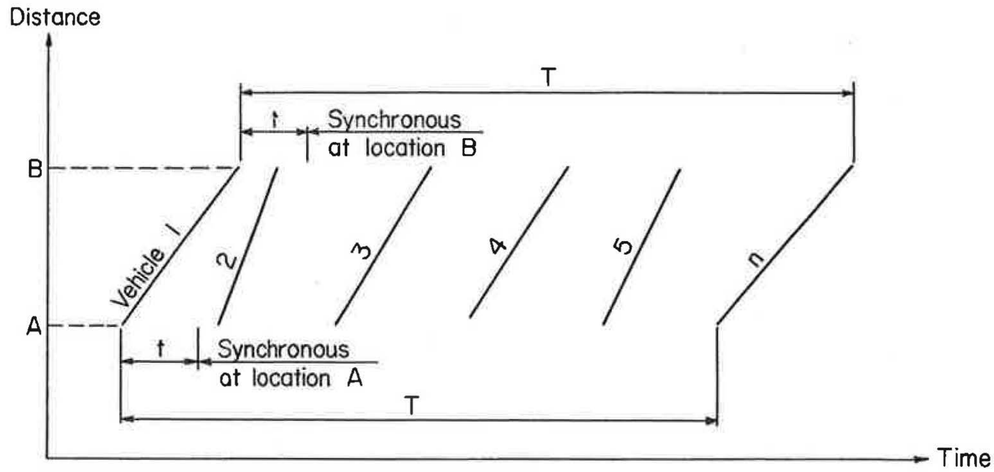


Figure 3.7. Time-space diagram illustrating synchronous counting procedures at two locations: A, regular arrivals, and B, Poisson arrivals.

On the other hand, if at point A the counting period t is chosen at random (asynchronous case), the probability of no arrivals is

$$P_0(t) = 1 - (t/\bar{t}) \quad (t < \bar{t}) \qquad (3.26)$$

$$P_0(t) = 0 \qquad (t \geq \bar{t}) \qquad (3.27)$$

This information is summarized in Table 3.1.

The queueing approach provides a distinct method for explaining the bunching tendency of constrained vehicles. In a queueing process, with random arrivals at a rate q per unit time and constant service time B (see Section 3.3), the probability that n units will be served during some period P_n follows a Borel distribution (*58*):

$$P_n = \frac{e^{-nBq}(nBq)^{n-1}}{n!} \qquad (n = 1, 2, \ldots .) \qquad (3.28)$$

Tanner (*68*) extended this concept to the general case to show that the distribution of the number of units served in a busy period starting with an accumulation of r units is

$$P(n \mid r) = \frac{e^{-nBq}(nBq)^{n-r}}{(n-r)!}\left(\frac{r}{n}\right)$$

$$n = r, r + 1, \ldots \qquad (3.29)$$

This is close to the Poisson form of Eq. 3.1. If constrained vehicles on a highway are considered as platoons or queues, the Borel-Tanner distribution can be used as a

Table 3.1 Comparison of the Synchronous and Asynchronous Counting Procedures Applied to Two Distributions of Arrivals

Location	Distribution of Arrivals	Probability of No Arrivals in Counting Interval			
		Synchronous Case		Asynchronous Case	
		$t < \bar{t}$	$t \geq \bar{t}$	$t < \bar{t}$	$t \geq \bar{t}$
A	Regular	$P_0(t) = 1$	$P_0(t) = 0$	$P_0(t) = 1 - (t/t)$	$P_0(t) = 0$
B	Random	$P_0(t) = e^{-qt}$	$P_0(t) = e^{-qt}$	$P_0(t) = e^{-qt}$	$P_0(t) = e^{-qt}$

model for the distribution of the queue lengths of constrained vehicles. This model is obtained if one starts with a random positioning (or set of arrival times) of vehicles, considers all vehicles within a distance B of the one ahead as queueing, and then moves these queueing vehicles back so that they are exactly B distance apart. At the same time, additional vehicles within B distance of the end of the queue are included. The probability that a queue has exactly n vehicles is given by

$$\mathrm{P}_n = \frac{n^{n-1}}{n!} r^{n-1} e^{-rn} \qquad (3.30)$$

In this derivation, the parameter r is given by $r = B\,k$, with k being the concentration of vehicles.

In a similar treatment, Miller (*46*) assumed that vehicles may be considered as traveling in platoons or queues, where a queue may consist of only one vehicle and where queues are independent of each other in size, position, and velocity. His criteria for determining queues were:

(a) The time interval between queueing vehicles should be less than 8 sec.
(b) The relative velocities of queueing vehicles should be within the range -3 to $+6$ mph.

The gaps between queues are assumed to be exponentially distributed and a one-parameter continuous distribution has been fitted to the number of vehicles in a queue as follows:

$$\mathrm{P}_n = (m+1)\,(m+1)!\,\frac{(n-1)!}{(m+n+1)!} \qquad (3.31)$$

in which m is the parameter of the Beta distribution. Miller stated that the distribution of Eq. 3.31 fits observed frequencies of queue lengths about as well as the Borel-Tanner distribution given by Eq. 3.30.

3.2.3 Gap Acceptance

In using mathematics to estimate delay when two streams of traffic interact, it is necessary to make assumptions regarding the time required for vehicles in the minor stream to cross or merge. It is assumed that the waiting driver or pedestrian measures each time gap, h, in the traffic on the major highway. He crosses (accepts the gap if $h \geq \tau$) or waits (rejects the gap if $h < \tau$). The value of τ was assumed to be a single constant value by early theorists. The interval from the arrival of the side street pedestrian or vehicle to the arrival of the next main street vehicle is not the same as the headway for that main street vehicle. Raff (*64*) used the term "lag" to describe both this first time interval and successive main street gaps. The critical lag, τ, was defined by Raff as that value of lag which has the property that the number of accepted lags shorter than τ is the same as the number of rejected lags longer than τ. This is shown in Figure 3.8.

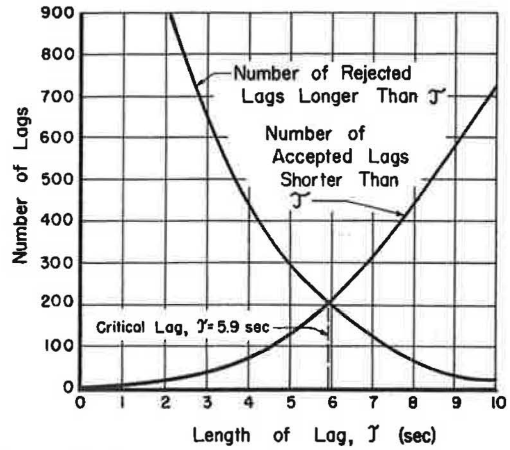
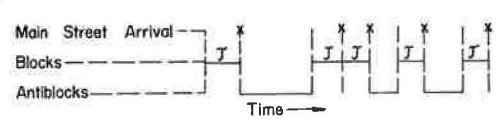


Figure 3.8. Typical distribution of accepted and rejected lags.

A stream of traffic can be thought of as a succession of gaps or succession of queues. In considering the pedestrian desiring to cross the street, it is practical to divide the traffic stream into intervals during which one cannot cross and intervals in which one can cross. In an early treatment of this concept Raff (*64*) referred to these intervals as blocks and antiblocks, respectively. A block is defined as the time preceding the passage of a main street vehicle by the critical lag τ. The time which is more than τ before the passage of the next car is considered to be in antiblocks as shown in the following:



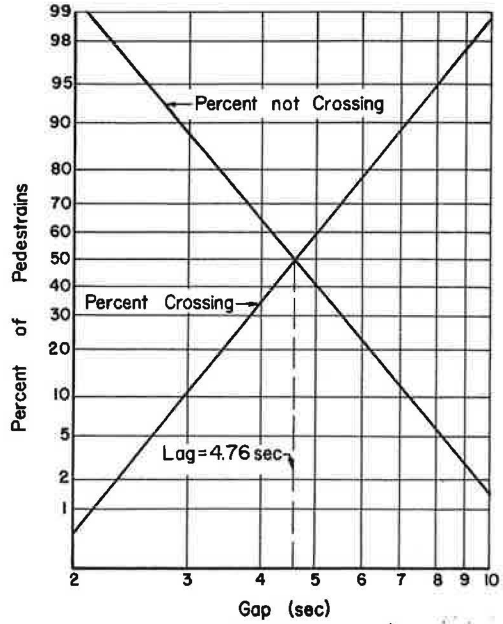


Figure 3.9. Pedestrian gap acceptance.

Additional treatment and applications to the interaction of two traffic streams at uncontrolled intersections are presented in Section 3.4.1.

Tanner (*67*) discussed the gap acceptance problem as applied to pedestrians and showed that an Erlang distribution of gap acceptances could be used to predict delays.

Cohen, Dearnaley and Hansel (*10*) observed pedestrian behavior on a main road in an English city and formulated a "criterion of risk" accepted by pedestrians. Gap acceptance data were separated into groups according to age and sex and a cumulative logarithmic normal distribution was fitted to gap acceptance, as shown in Figure 3.9.

The speed of the merging vehicle is important in considering the distribution of gaps that is acceptable to the merging driver at a freeway entrance ramp. The Midwest Research Institute (*18*) analyzed gap acceptances for moving and stopped vehicles. The data used were gathered by the Texas Transportation Institute on several Texas freeways. Figure 3.10 shows the results. A more complete discussion of merging delays is contained in Section 3.6.1.

Several writers have proposed more realistic models, which associate with each time gap, h, a gap acceptance probability, $F(h)$. This says that there exists a certain chance or probability, $F(h)$, that a driver or pedestrian when faced with a gap of duration h will accept it and cross the street. In a controlled series of experiments conducted at the General Motors Research Laboratory, Herman and Weiss (*30*) showed that for stopped cars the form of $F(h)$ can be approximated by a translated exponential distribution

$$F(h) = 0 \qquad (h < \tau)$$

$$F(h) = 1 - \exp\left[-\lambda\,(t - \tau)\right] \qquad (h \geq \tau)$$

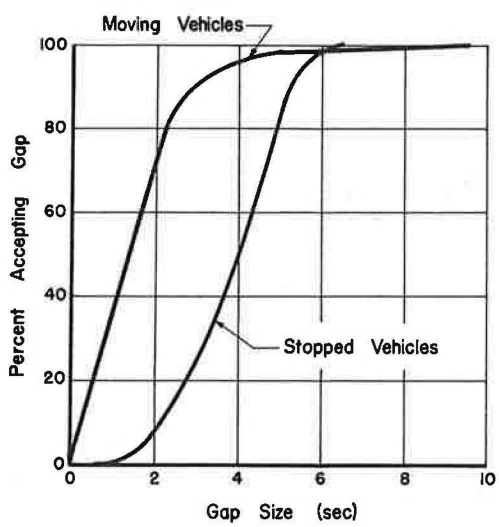


Figure 3.10. Gap acceptance of merging vehicles.

in which τ and λ are the parameters of the translated exponential distribution. τ is the minimum acceptable gap, and λ is $1/(\bar{t} - \tau)$, where $\bar{t}$ is the average gap accepted.

A graph of $1 - F(h)$ versus h for $\tau = 3.3$ sec and $\lambda = 2.7$ sec^{-1} is shown in Figure 3.11.

Weiss and Maradudin (*75*) developed a method of treating gap acceptance delay which accounts for driver impatience. They postulated that the size of acceptable gap is reduced as delay increases. Instead of a constant size of acceptable gap, τ, they state that the probability of a driver accepting a gap of size H after the ith vehicle has passed is $F_i(H)$, and the driver's impatience would be reflected by the case where

$$F_0(H) \leq F_1(H) \leq \ldots\ldots F_i(H)$$

The previous discussion has been predicated on the crossing of a single lane of traffic. In considering the N-lane highway from the waiting driver or pedestrian point of view, the individual waiting may elect to regard a gap as the time between the arrival of two cars at the intersection, regardless of which lane they occupy, or to cross on the basis of the gaps in each of the lanes.

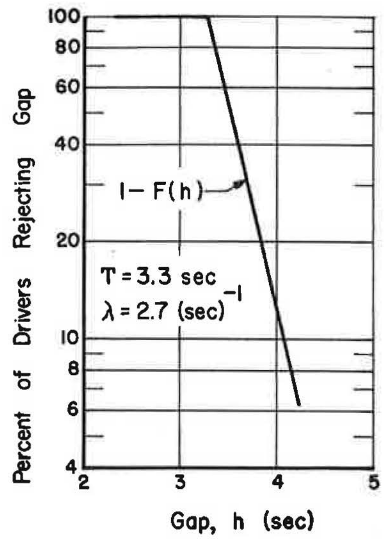


Figure 3.11. Crossing gap rejection.

3.3 ELEMENTARY QUEUEING THEORY

In most traffic engineering problems the first step is to provide adequate capacity for the *average* flow of vehicles in the system. If this is not done, there will be constant congestion. Even if the capacity is adequate for average flow, congestion can occur because the flow or capacity fluctuates. Queueing or waiting-line theory is concerned with describing these fluctuations and predicting quantitative operating characteristics of the system.

Theoretical research into the properties of congested systems began in the 20th century in connection with problems in the design of telephone exchanges. However, it was not until about 1950 that waiting-line theory was extensively applied to other congestion problems.

Most operational systems can be broken into elements, each of which has a basic behavior pattern. Items arrive at some facility which services and eventually discharges each item. Arrival of vehicles at a toll booth would be an example of such a system. In some cases, such as traffic flow through signalized intersections, items must pass through a sequence of servicing operations.

If the demand for service occurs at equal intervals of time, if the servicing rate of the system is constant, and if the serving capacity of the servicing facility is greater than the demand, each item entering the system will experience the same delay. However, in almost all situations involving human actions there are irregularities in demand and service which result in varying levels of congestion. If these irregularities can be specified mathematically, the important congestion characteristics can be obtained.

There are two fundamental approaches to describing the operation of a queueing system. From the customer's viewpoint, such characteristics as the average delay in the system, the percentage of customers delayed, and the percentage of customers delayed longer than a given period are important. From the serving facility's viewpoint, the degree of utilization or idleness of the facility becomes important. The optimum solution for each approach does not lead to the same system configuration. In almost all practical cases, the minimization of customer delay results in poor utilization of the service facility. Efficient use of a serving facility usually means substantial delay to items in the system.

In order to predict mathematically the characteristics of a queueing system, it is necessary to specify the following system characteristics and parameters:

(a) Arrival pattern characteristics:
  (1) Average rate of arrival.
  (2) Statistical distribution of gaps.

(b) Service facility characteristics:
  (1) Service time average rates and distribution.
  (2) Number of customers which can be served simultaneously, or number of channels available.

(c) Queue discipline characteristics, such as the means by which the next customer to be served is selected; for example, "first come first served," or "most profitable customer first."

Assume that items arrive at a location where they are to be processed. The time between the arrival of consecutive items is called the inter-arrival time or gap. Both the inter-arrival and service times required at the service centers are of varying lengths. This variation is statistical. That is, the probability of the occurrence of a given time interval is described by a probability distribution. Items unable to be served at once form a waiting line or queue and are served in turn when the service channels are free. If the arrival and service intervals are independent, a description of the system at any time t depends on the inter-arrival and service time distributions, together with knowledge of the situation at time zero.

The fundamental quantities characterizing a waiting line are the *states* of the system. The system is said to be in state n if it contains exactly n items (this includes all items being served). The value of n may be either 0 or some positive integer.

The queue will behave differently under the following two conditions:

(a) The average arrival rate is less than the mean service rate.
(b) The average arrival rate exceeds the mean service rate.

If the average arrival rate is called λ, the average interval between arrivals is $1/\lambda$. If the service rate of the system is called μ, the average service time is $1/\mu$. The ratio $\rho = \lambda/\mu$, sometimes called the traffic intensity or utilization factor, determines the nature of the various states. If $\rho < 1$ (that is, $\lambda < \mu$), and a sufficiently long time elapses, each state will be recurrent. This means that there is a finite probability of the queue being in any state n. If, on the other hand, $\rho > 1$, every state is transient and the number in the system will become longer and longer without limit. A fundamental theorem states that the queue will be in equilibrium only if $\rho < 1$.

An understanding of the characteristics of queueing systems can be obtained from simple cases. Consider the case of a single-channel queueing system with a mean random arrival rate of λ customers per unit of time and where service times are independent and distributed exponentially with a mean rate μ. Let $P_n(t)$ be the probability that the queueing system has n items at time t. Consider the situation at time $t + \Delta t$ where Δt is so short that only one customer can enter or leave the system during this time. There are three ways in which the system can reach state n at time $t + \Delta t$ (when $n > 0$):

(a) The system was in state n at t and no customers arrived or departed in Δt.
(b) The system was in state $n - 1$ at t and one customer arrived in Δt.
(c) The system was in state $n + 1$ at t and one customer departed in Δt.

The probability of the system being in state n at $t + \Delta t$ is

$$\begin{aligned} P_n(t + \Delta t) = {} & \\ & P_n(t)\,[(1 - \lambda\,\Delta t)(1 - \mu\,\Delta t)] + \\ & P_{n-1}(t)\,[(\lambda\,\Delta t)(1 - \mu\,\Delta t)] + \\ & P_{n+1}(t)\,[(1 - \lambda\,\Delta t)(\mu\,\Delta t)] \end{aligned} \tag{3.32}$$

After developing the expression for $[P_n(t + \Delta t) - P_n(t)]/\Delta t$, and letting $\Delta t \to 0$,

$$\frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) + \mu P_{n+1}(t) - (\lambda + \mu)\,P_n(t) \tag{3.33}$$

in which $n = 1, 2, 3, \ldots$.

When $n = 0$,

$$\frac{dP_0(t)}{dt} = \mu P_1(t) - \lambda P_0(t) \tag{3.34}$$

These fundamental equations can be expressed as differential-difference equations whereby the steady-state solutions are obtained by setting

$$\frac{dP_n(t)}{dt} = 0 \tag{3.35}$$

The resulting equations are of the form

$$\mu P_{n+1} + \lambda P_{n-1} = (\lambda + \mu) P_n, \quad (n > 0)$$

$$\text{and } \mu P_1 = \lambda P_0, \quad (n = 0) \tag{3.36}$$

in which P_n is the value of $P_n(t)$ as $t \to \infty$.

The first few equations are as follows:

$$\lambda P_0 = \mu P_1 \tag{3.37}$$

$$\lambda P_0 + \mu P_2 = (\lambda + \mu)P_1 \qquad (3.38)$$

$$\lambda P_1 + \mu P_3 = (\lambda + \mu)P_2 \qquad (3.39)$$

Noting that $P_1 = \rho P_0$, and substituting this in Eqs. 3.37, 3.38, and 3.39, gives

$$P_2 = (\rho + 1)\, P_1 - \rho P_0 = \rho^2 P_0 \qquad (3.40)$$

$$P_3 = (\rho + 1)\, P_2 - \rho P_1 = \rho^3 P_0 \qquad (3.41)$$

$$P_n = \rho^n P_0 \qquad (3.42)$$

Because the sum of all probabilities is 1,

$$\sum_{n=0}^{n=\infty} P_n = 1$$

$$1 = P_0 + \rho P_0 + \rho^2 P_0 + \ldots.,$$

$$1 = P_0 (1 + \rho + \rho^2 + \rho^3 + \ldots..)$$

$$1 = P_0 \left(\frac{1}{1-\rho}\right)$$

and

$$P_0 = 1 - \rho \qquad (3.43)$$

The traffic intensity, ρ, then can be seen to express the fraction of time that the system is busy (P_0 is the probability that the system is empty and $1 - P_0$ is the probability that it is occupied).

The average number of customers in the system is

$$E(n) = \sum_{n=0}^{n=\infty} n\, P_n$$

$$= 0 + P_1 + 2P_2 + 3P_3 + \ldots..$$

$$= P_0 (\rho + 2\rho^2 + 3\rho^3 + \ldots\ldots\ldots)$$

$$= (1-\rho)\left(\frac{\rho}{(1-\rho)^2}\right)$$

$$= \frac{\rho}{1-\rho} \qquad (3.44)$$

This relationship, shown in the upper curve of Figure 3.12, illustrates a characteristic of most queueing systems. The average number in the system increases slowly until

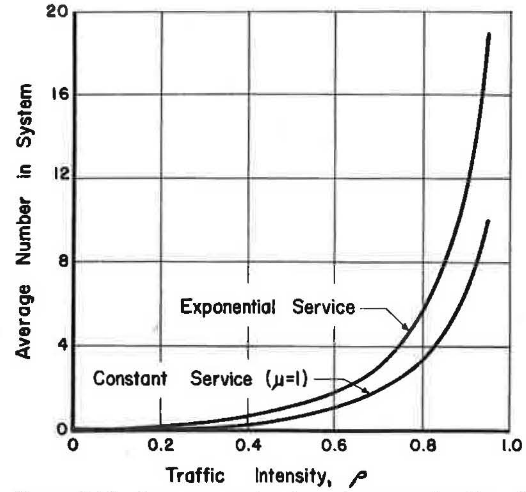


Figure 3.12. Average number in system as a function of traffic intensity.

a ρ of approximately 0.8 or more is reached and then increases rapidly.

The variance of the number in the system is

$$\text{Var}(n) = \sum_{n=0}^{n=\infty} [n - E(n)]^2 P_n = \frac{\rho}{(1-\rho)^2} \qquad (3.45)$$

This relationship, plotted in Figure 3.13,

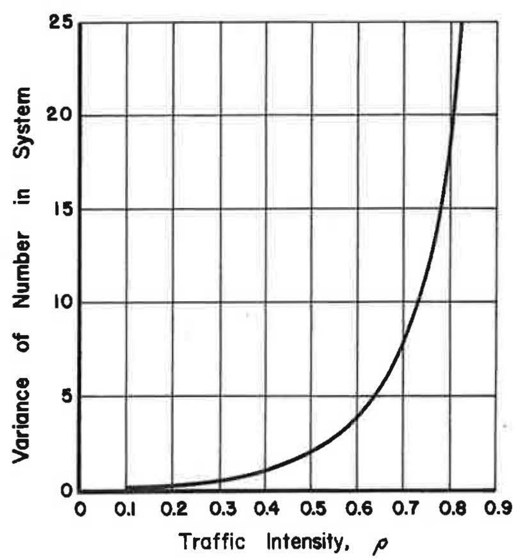


Figure 3.13. Variance of number in system as a function of traffic intensity.

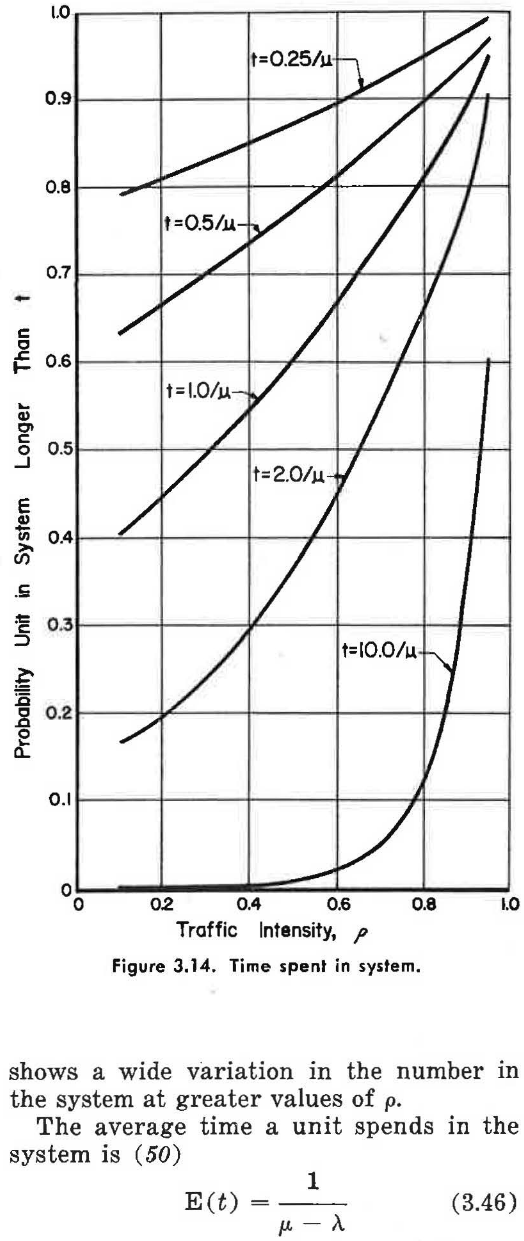


Figure 3.14. Time spent in system.

shows a wide variation in the number in the system at greater values of ρ.

The average time a unit spends in the system is *(50)*

$$E(t) = \frac{1}{\mu - \lambda} \tag{3.46}$$

The probability that a unit is in the system longer than some multiple of the average service time, $1/\mu$, is shown in Figure 3.14.

In the special case where the time required to serve each customer is constant, the average number in the system is less than when service is exponentially distributed. The equation is

$$E(n) = \frac{\rho(2 - \rho)}{2(1 - \rho)} \tag{3.47}$$

which is plotted as the lower curve of Figure 3.12.

After a complex system has been described in mathematical terms, the analyst may develop the equations describing the operating characteristics of the system. If the problem is complex it may be necessary to resort to simulation (see Chapter 4).

3.4 DELAYS AT INTERSECTIONS

Traffic theorists have developed several probabilistic approaches to the problem of analyzing delays at an intersection of two streets. This section summarizes some of their findings.

Tanner *(71)* (see Section 3.6.2) has presented an explicit formulation for the uncontrolled, low-volume intersection where the vehicle occupying the intersection has the right-of-way. Usually, flows under these conditions are not high enough to warrant a study of delay, and this situation will not be considered further in this section.

The cases of interest are those where the traffic flow on the main street is of sufficient magnitude that side-street traffic encounters delay in crossing. Stop-sign or traffic-signal controls are generally used in these circumstances. This section deals with delays at intersections with these two types of control.

At the stop-sign controlled intersection, it is assumed that the side-street traffic waits for an adequate gap in the main-street traffic before crossing.

The problem of crossing the main street will be considered for both pedestrians and vehicles. There is a fundamental difference between these two cases. Pedestrians arrive at the crossing and accumulate at the curb until an opportunity to cross presents itself. The entire group then crosses together, independent of the number of pedestrians waiting. On the other hand, later vehicular arrivals cannot cross the main stream until the first vehicle in line has departed. If side-street vehicular flow is so low that two vehicles will rarely be waiting, the delays to individual vehicles will be the same as those for individual pedestrians.

The problem of pedestrians crossing at a pre-timed signalized intersection is insignificant when conflicts with cross-street turning traffic are ignored. Under these conditions delays to these pedestrians can be easily determined from knowledge of the pedestrian arrival distribution and the traffic signal timing. The unsignalized and signalized intersection delay problems are treated in Sections 3.4.1 and 3.4.2, respectively.

The intersectional delay problem presents a type of queueing problem different from the typical situation described in Section 3.3. In the typical situation, delay results from servicing items. In the intersectional delay problem, delay results from a combination of the gap acceptance characteristics of the crossing traffic and the passage of gaps inadequate for crossing.

3.4.1 Unsignalized Control

Pedestrian delay at an unsignalized intersection was first treated by Adams (*1*) in 1936 in one of the earliest theoretical traffic papers. He assumed that pedestrian and vehicle arrivals are random and made field observations which generally justified the assumption. If it is assumed that the main-street flow is q and that an interval τ (the critical gap) is required between successive arrivals on the main street for a pedestrian to cross safely, several delay relationships can be derived.

The probability that pedestrians will be delayed is

$$P_d = 1 - e^{-q\tau} \tag{3.48}$$

which is plotted in Figure 3.15 with main-street flow expressed in vehicles per minimum acceptable gap.

The average delay for all pedestrians is

$$E(t) = \frac{1}{q\, e^{-q\tau}} - \frac{1}{q} - \tau \tag{3.49}$$

which is plotted in Figure 3.16 with delay in terms of the minimum crossing gap required.

Also plotted in Figure 3.16 is the average delay for those pedestrians delayed, which is expressed as

$$E_d(t) = \frac{1}{q\, e^{-q\tau}} - \frac{\tau}{1 - e^{-q\tau}} \tag{3.50}$$

Adams observed pedestrian delay at five London locations and computed τ from observed values of the dependent variables in the three relationships shown in Eqs. 3.48, 3.49, and 3.50. The average value of τ was found to be approximately 4 sec, with a variability of less than 0.5 sec at most locations. This indicates that the assumption of Poisson traffic in deriving these relationships is reasonably satisfactory.

If a pedestrian or vehicle wishes to cross the main street and must yield the right-of-way to main-street traffic, there is a period of time, τ, which must be available for the crossing to be made safely. Raff (*64*) called this time the "critical lag." Traffic on the main street generates a succession of time periods when crossing is alternately possible and impossible for the side-street traffic. The periods when crossing is impossible are called *blocks* and those when crossing is possible are called *antiblocks* (see Section 3.2.3).

Raff (*64*) developed the probability distribution of block lengths by considering the distribution of waiting times for crossing vehicles. He showed that the cumulative distribution of block lengths, $B(t)$, is related to the cumulative distribution of waiting times, $F(t)$, in the following manner:

$$B(t) = 1 - \left(\frac{1}{q\, e^{-q\tau}}\right)\left(\frac{dF(t)}{dt}\right) \tag{3.51}$$

in which q is the main-street flow and τ is the critical lag.

Using Garwood's (*16*) expression for $F(t)$, Raff evaluated Eq. 3.51. The percentage of waiting times less than several multiples of τ is shown in Figure 3.17, in

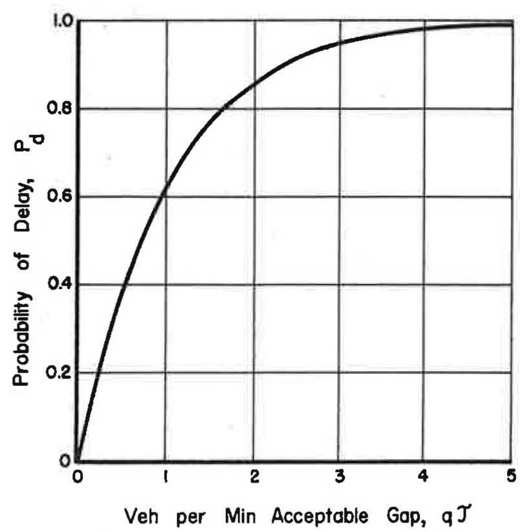


Figure 3.15. Probability of pedestrian delay.

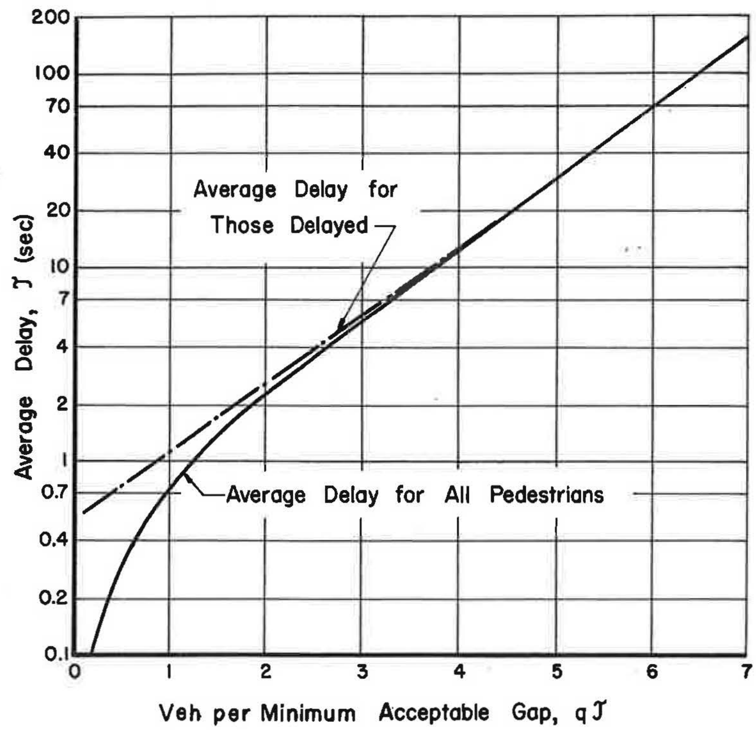


Figure 3.16. Average delay to pedestrians.

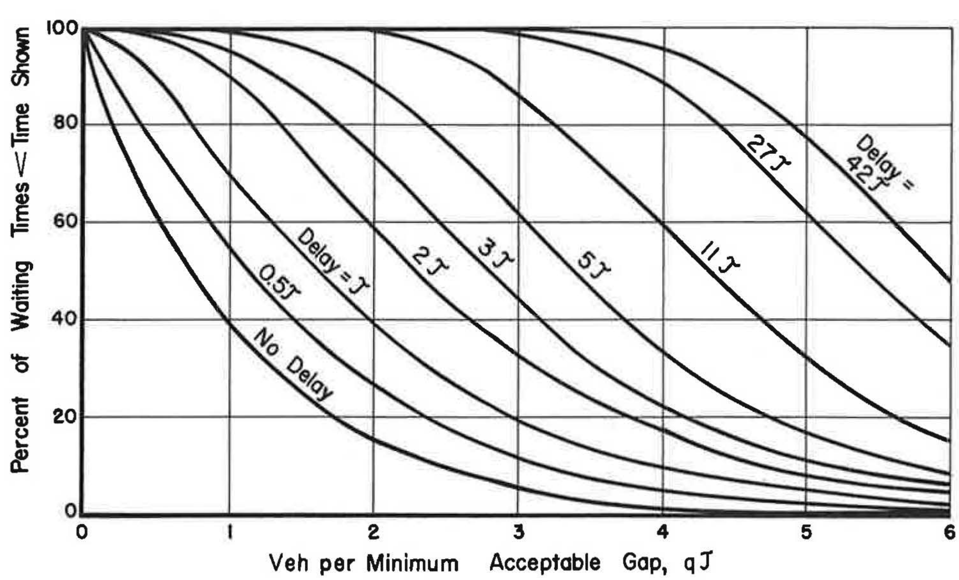


Figure 3.17. Cumulative pedestrian delay.

which main-street flow and delay are expressed in terms of the critical lag τ.

Oliver (*60*) extended earlier work on crossing opportunities. He considered more general arrival distributions than Poisson and derived several important relations involving blocks, antiblocks, delays, and waiting times for these distributions. His paper provides a unified notation and compares this notation with those used by other recent theorists.

In 1951, Tanner (*67*) published the results of a comprehensive study of pedestrian crossing delays. He assumed random arrivals of both main-street vehicles and crossing pedestrians and presented three proofs of Garwood's (*16*) crossing delay distribution. Tanner considered varying values of gap acceptance for different pedestrians and gave some attention to the problem of groups of pedestrians crossing the street.

Tanner derived five relationships for pedestrian arrivals. Two of these are the distribution of size of pedestrian groups crossing together and the distribution of the number of pedestrians waiting.

The average size of a group crossing together is

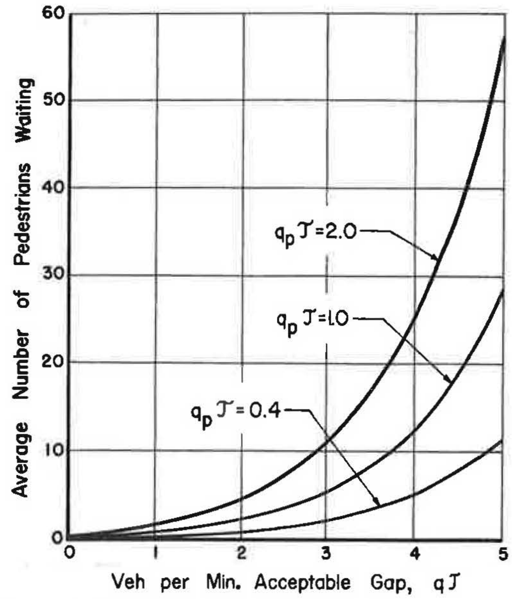


Figure 3.19. Average number of pedestrians waiting to cross street.

$$E(n_c) = \frac{q_p\, e^{-q} + q_p\, e^{p}}{e^{q_p - q}\, (q_p + q)} \qquad (3.52)$$

in which pedestrian flow, q_p, and vehicular flow, q, are expressed in terms of critical lags. Figure 3.18 shows this relationship.

The average number waiting to cross is

$$E(n_w) = \frac{q_p}{q}\,(e^q - q - 1) \qquad (3.53)$$

which is plotted in Figure 3.19.

Tanner also compared the delay to pedestrians crossing the entire roadway at one time with the delay to those stopping in the middle at a refuge island when necessary. His field studies indicated that pedestrians crossing the street without stopping look for a gap of at least the critical lag in both directions of movement rather than for some combination of near- and far-stream gaps. The average delay, expressed in units of the critical gap required to cross the entire street without stopping, is

$$E(t) = \frac{e^{4q} - 4q - 1}{2q} \qquad (3.54)$$

When the pedestrian can stop in the middle

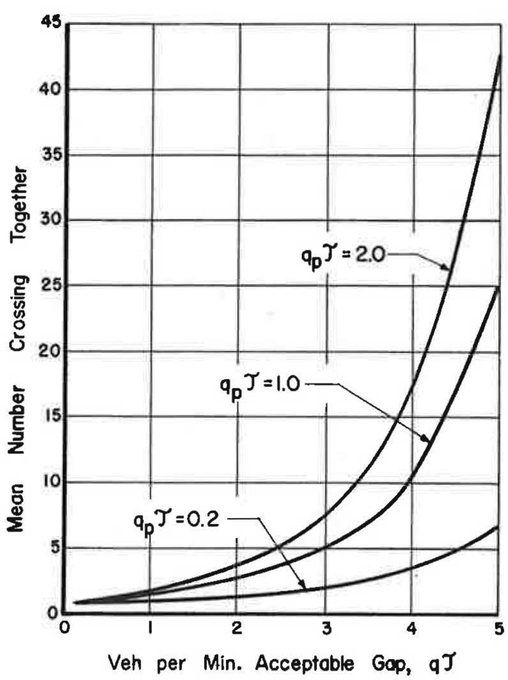


Figure 3.18. Average number of pedestrians crossing together.

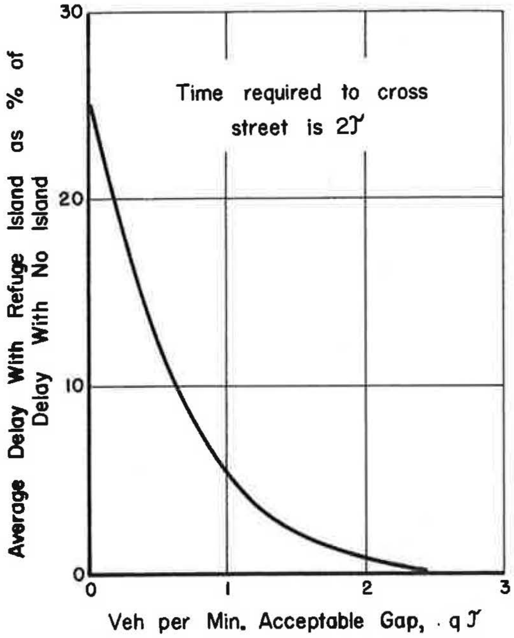


Figure 3.20. Pedestrian delays with and without refuge island.

of the street at a refuge island, the average delay is

$$E(t_s) = \frac{2(e^q - q - 1)}{q} \quad (3.55)$$

These delays are compared in Figure 3.20.

Moskowitz (*51*) applied Garwood's (*16*) waiting time relationship to California traffic with very satisfactory results, as shown in Figure 3.21. Moskowitz also prepared numerous graphs of Garwood's relationship.

Jensen (*35*) postulated that the acceptable gap follows a normal distribution, and developed relationships for the probability of no delay and average delay which are analogous to Adams' (*1*) delay formulas.

Mayne (*44*) generalized Tanner's results to include an arbitrary distribution of independent main-street headways. He also considered the effects of introducing refuge islands on a wide crossing. He showed that for the same average delay the pedestrian flow is at least four times as great when an island is present as when there is no island.

Jewell (*37*) obtained the distribution, mean, and variance, of waiting times for arbitrary main-stream headway distributions and for several main-street situations at the time a side-street vehicle presents itself. His relationships were developed for a critical lag τ and extended for other gap acceptance criteria. He obtained results for the number of minor-street vehicles that can be discharged during a fixed time period when only one side-street vehicle can cross during each acceptable main-street gap. He also showed that the mean delay for the side-street vehicle increases in proportion to the second or higher power of the critical gap and at least linearly with increasing flow. The variance of delay increases in proportion to the third or higher power of the critical gap.

In two recent papers, Weiss and Maradudin (*75*) and Herman and Weiss (*30*) further considered the delay problem at unsignalized intersections. Weiss and Maradudin developed several generalizations of the crossing delay problem studied by earlier investigators. The approach is based on renewal theory described by Feller (*15*). A renewal process in time is the occurrence of random spacings from a known gap distribution. With their technique (*75*), it is possible to deal with a general independent distribution of main-street gaps and a general gap acceptance distribution. This makes it possible to consider the "yield-sign" delay problem where the side-street vehicle has a different initial critical lag, depending on whether it is moving or stopped. It is also possible to develop delay functions for the impatient driver whose probability of accepting a given gap in the main street increases with the passage of main-street vehicles.

Weiss and Maradudin expressed delay characteristics for several gap and gap acceptance distributions. Herman and Weiss (see Section 3.2.3) fitted shifted exponential constants experimentally. For Poisson main-street traffic and shifted exponential gap acceptance, the mean delay to side-street traffic is

$$E(t) = \frac{e^{q\tau} - 1}{q} - \tau + \frac{1}{b}\left\{ e^{q\tau} - 1 - q\tau \left[\frac{q}{q+b}\right]^2 \times \left[(1 + q\tau + b\tau)(1 - e^{-q\tau})\right] + e^{-q\tau}\left[\frac{q}{q+b} + q\tau\right]\right\} \quad (3.56)$$

in which τ is the minimum acceptable gap, q is the main-street flow, and b is the parameter of the shifted exponential gap acceptance distribution, $= 1/(\bar{t} - \tau)$. Then,

$$f(t) = \frac{1}{b} \exp \left[-b(t-\tau) \right], \quad t \geq \tau \tag{3.57}$$

The upper curve of Figure 3.22 presents a graph of this relationship for Herman's and Weiss's constants, $\tau = 3.3$ sec and $b = 2.7$ sec^{-1}. The lower curve shows the results of assuming that all drivers have an acceptable gap of 3.3 sec.

The probability of no delay is given by

$$P_0 = \frac{b}{b+q} e^{-q\tau} \tag{3.58}$$

which is plotted in Figure 3.23. The upper curve shows a relationship using Herman's and Weiss's constants, $\tau = 3.3$ sec and $b = 2.7$ sec^{-1}, whereas the lower curve shows the results of assuming that all drivers have an acceptable gap of 3.3 seconds.

Weiss and Maradudin introduced a term called the "transparency" of the street. This is the fraction of time that a side-street driver would consider it safe to cross the street. They developed an explicit relationship for transparency as a function of the main-street gap distribution and the side-street gap acceptance distribution. For the case of random arrivals and a shifted exponential gap acceptance function, the transparency is

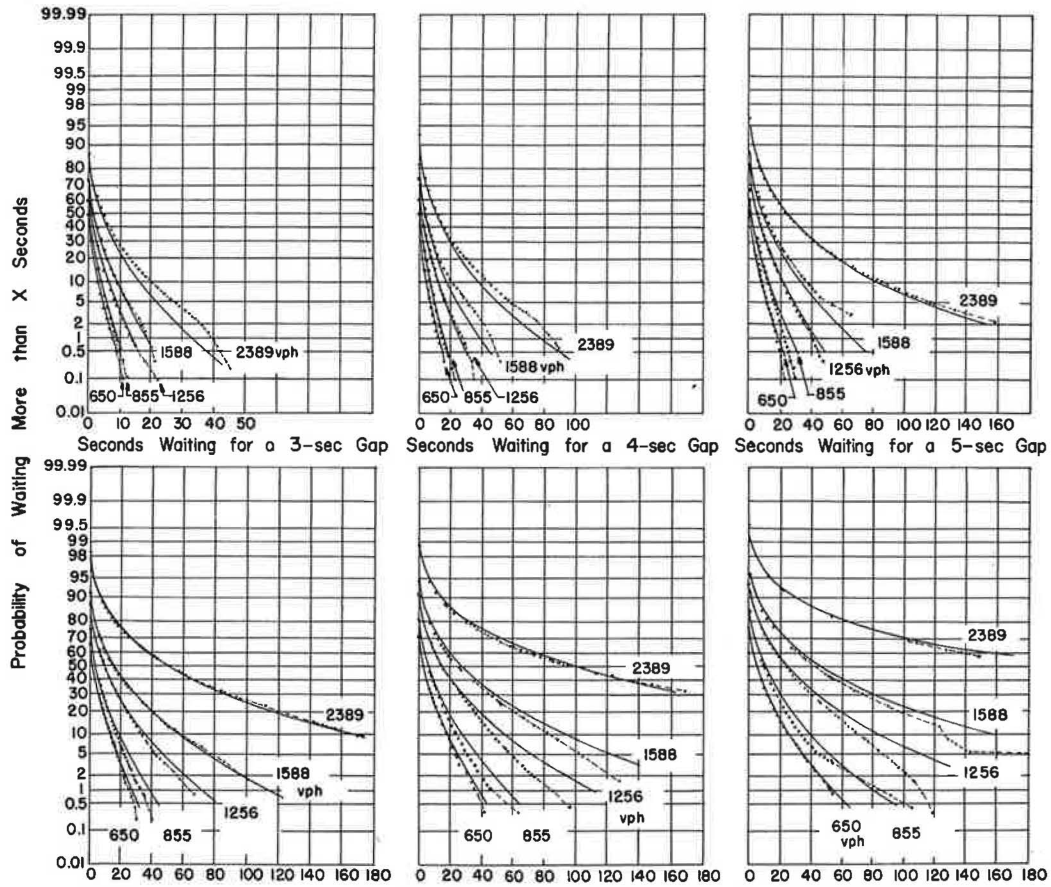


Figure 3.21. Probability of waiting various times for specified gaps at several traffic volume rates.

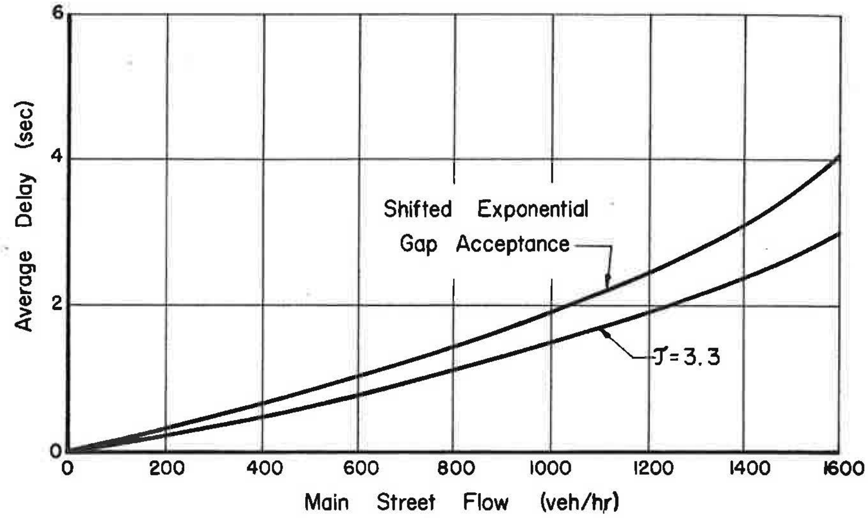


Figure 3.22. Average delay crossing a street.

$$\Phi = \frac{(1 + q/b)}{e^{q\tau}\left(1 + \frac{q}{b}\right)^2 - \frac{2q}{b}\left[1 + \frac{q\tau}{b}(b + q)\right]} \qquad (3.59)$$

which is plotted in Figure 3.24 and compared with the probability of no delay P_0 for the same case.

Weiss and Maradudin also considered the yield-sign problem. If a moving vehicle requires a gap of τ_1, and a stopped vehicle requires a gap of τ_2 ($\tau_1 \leq \tau_2$), the mean delay is

$$E(t) = \frac{e^{q\tau_2}}{q}(1 - e^{-q\tau_1}) + e^{-q\tau_1}(\tau_2 - \tau_1) - \tau_2 \qquad (3.60)$$

As an example, assume that $\tau_2 = 3.3$ sec and $\tau_1 = 2.0$ sec. Substituting these values in Eq. 3.60 yields a plot as shown in Figure 3.25, which shows the average side-street vehicle delay (at a yield sign) compared with that at a stop-sign situation where all side-street drivers are required to stop and wait for a main-street gap of 3.3 sec.

Weiss and Maradudin were able to gen-

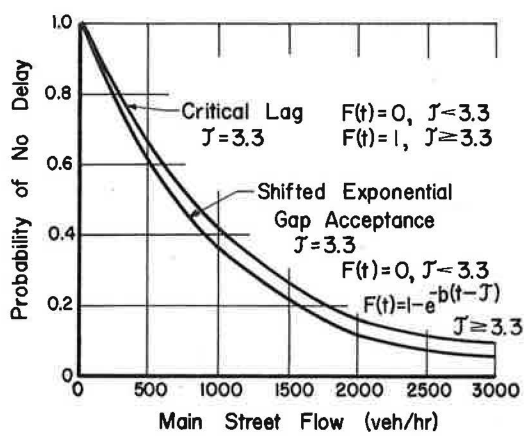


Figure 3.23. Probability of no delay at a signalized intersection.

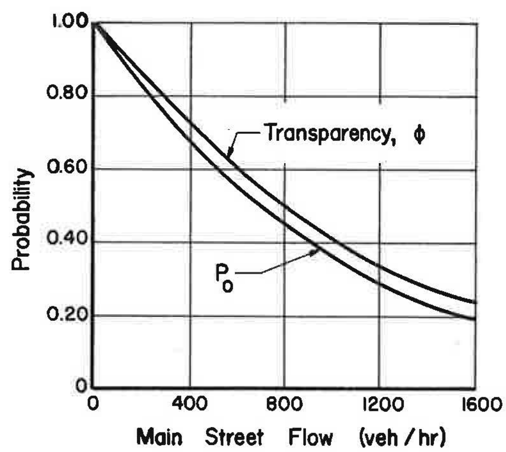


Figure 3.24. Transparency and probability of no delay at a signalized intersection.

eralize the approach to the problem of a pedestrian or vehicle crossing an N-lane highway. Tanner (*67*) considered the problem as mentioned earlier in this section. The flow in N lanes, each with Poisson traffic, yields Poisson traffic with a mean which is the sum of the mean flows for each of the N individual lanes. Weiss and Maradudin also derived an expression for delay when the gap distribution on the main street is dependent on time. Such situations occur during peak traffic flow periods and when the gaps are not independent, such as immediately downstream from a traffic signal.

As described in Section 3.2.2, Miller (*46*) postulated that bunches of vehicles are randomly distributed on a highway. Letting the flow of queues be q and the arrival rate of queues be λ, Miller derived an expression for the mean waiting time

$$E(t) = \lambda\,(\bar{t} + \tau)^2/2 \qquad (3.61)$$

in which $\bar{t}$ is the average time for a queue to pass the crossing point. For example, if it is assumed that a pedestrian needs a time gap of at least 10 sec ($\tau = 10$), that there are 90 queues per hour ($\lambda = 1/40$), and that it takes on the average 10 sec ($\bar{t} = 10$) for a queue to pass, $E(t) = 1/2 \times 1/40\ (10 + 10)^2 = 5$ sec.

The probability that a side-street vehicle can cross immediately is given by

$$P_0 = (1 - q\,\bar{t})\, e^{-\lambda\tau} \qquad (3.62)$$

To solve this relationship one must make use of the relationship $1/q = \bar{t} + 1/\lambda$. Thus,

$$P_0 = (1 - 10/50)\, e^{-0.25} = 0.622 \qquad (3.63)$$

Miller made a limited comparison of the average side-street delay and frequency of undelayed crossings predicted by the random bunches model with those produced by the random vehicle model. He found little difference in average waiting time for crossing vehicles. The random bunches model predicted the opportunities for immediate crossing better than did the random vehicles model. Figure 3.26 gives the observed values for immediate crossing opportunities and the values predicted by the two theoretical models for several levels of main-street traffic flow.

None of the previously described mathematical models fully accounts for what is frequently the most significant cause of delay for vehicles crossing the main street—the additional delay resulting from waiting behind other vehicles in the side-street waiting line. According to Weiss and Maradudin, who considered the case of two vehicles arriving simultaneously from the side street, treatment of the full queueing problem is quite difficult. Even a problem involving only two side-street vehicles is difficult to solve if the vehicles are not assumed to arrive together.

Oliver and Bisbee (*62*) derived the delays for side-street vehicles using several assumptions. They stipulated Poisson arrivals on the minor stream and made the important assumption that only one side-street vehicle can cross for each acceptable main-street gap, an assumption which they show to be reasonable for high main-street flow where few long main-street gaps occur. Their approach is further treated in Section 3.6.1.

Beckmann, McGuire and Winsten (*4*) described a model which takes into account the delay resulting when vehicles on a minor road are delayed by vehicles ahead of them waiting to cross the major stream. They assumed that arrivals and departures take place only at discrete points in time, as if a picture were made at the intersection at equally-spaced time intervals. Only one vehicle can arrive or cross the main stream

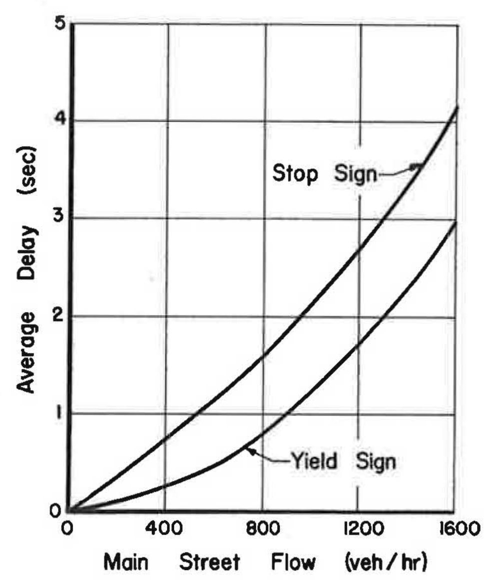


Figure 3.25. Average delay for side-street vehicles.

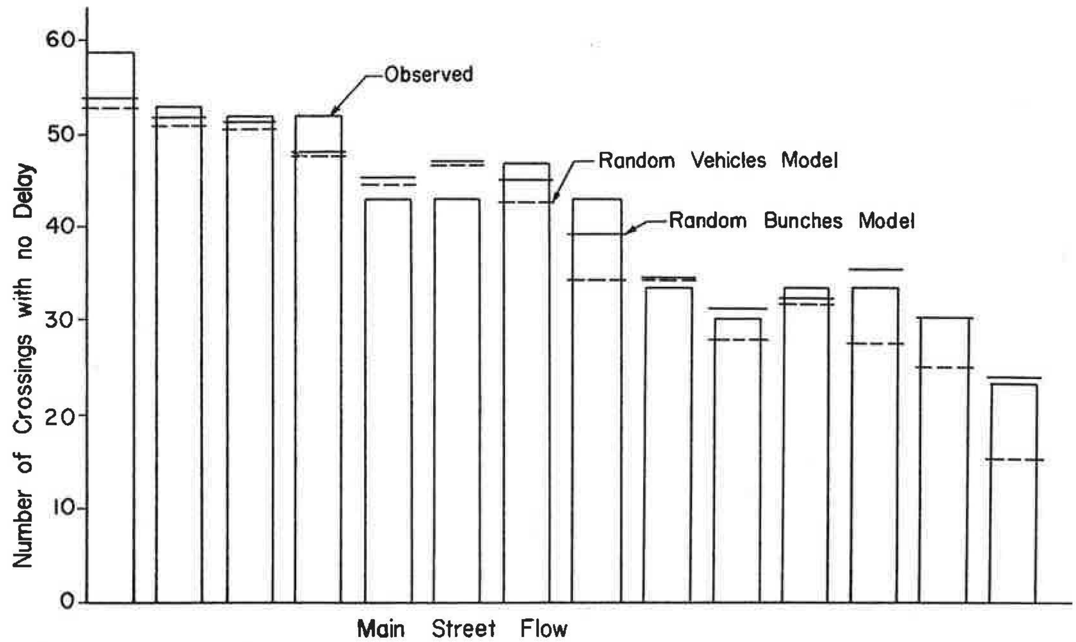


Figure 3.26. Comparison of undelayed crossing opportunities.

at any one time. Each point in time is in a block or antiblock, depending on the crossing opportunity presented by the main stream. A sequence of 10 points might look like this

B B B   B B
' ' ' ' ' ' ' ' ' '

in which B indicates blocked points and the remaining points are in antiblocks. They define a queue sequence as the number of cars held over from one time point to the next. If the side-street arrival distribution, the block-antiblock process, and the number waiting at time zero are known, the queues for later time points can be calculated successively. As an example, consider a situation where no vehicles are waiting to cross at time zero. The following table can be constructed:

Time	0	1	2	3	4	5	6	7	8	9	10
Arrival			A	A			A				A
Blocks				B	B	B	B				B
Queue sequence	0	0	0	1	1	1	2	1	0	0	1
Cars waiting	0		1	1	1	1	2	2	1	0	1
Departures	0		1	0	0	0	0	1	1	0	0

Beckmann, McGuire and Winsten assumed that the side-street arrivals are generated by a binomial process. Figure 3.27 shows their results where k_1 and k_2 are the minor and major road densities, respectively, and the critical lag is τ. The distribution of lengths of antiblocks is assumed to be geometric, or

$$P(x) = p^{x-1}(1-p), \qquad x = 1, 2, \ldots$$

in which p is the probability of a point being in an antiblock, and x is the length of the antiblock. The relative frequency of block lengths is $f(b)$ and the average block length is $E(b)$. They found expected waiting time by developing expected queue lengths at blocked and antiblocked points. The expected queue length is

$$E(q) = \frac{k_1(1-p)[E(b)^2+E(b)]}{2[1+(1-p)E(b)][1-k_1-k_1(1-p)E(b)]} \tag{3.64}$$

When the main-street arrivals are assumed to be binomial approximations to random, Eq. 3.64 can be used. They describe a method of obtaining the necessary moments of the block length distribution. The average waiting time is

$$E(t) = \frac{1 - p\,(1 + \tau k_2)}{k_2\,(p - k_1)} \tag{3.65}$$

in which $p = (1 - k_2)^\tau$.

In a recent paper, Tanner (*71*) considered the minor-street delay problem using random arrivals for both traffic streams. He derived the steady-state mean delay experienced by side-street vehicles under several conditions, including multi-lane traffic on the major road. He further calculated average delay for combinations of representative values of minimum main- and side-street headways and starting sluggishness for side-street vehicles.

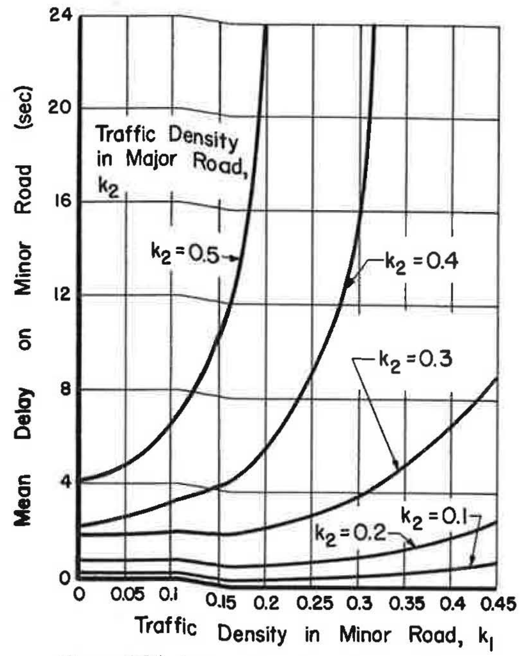


Figure 3.27. Intersection delay for $\tau = 2$ sec.

3.4.2 Signalized Control

There are several probabilistic models which may be used in the investigation of delays at signalized intersections. Several approaches developed by traffic theorists are considered in the following three sections.

Section 3.4.2.1 gives treatment of delays at pretimed signals, Section 3.4.2.2 discusses timing of traffic signals, and Section 3.4.2.3 gives a rationalization of delays at traffic-actuated signals.

3.4.2.1 Delays at Pretimed Signals. Beckmann, McGuire, and Winsten (*4*) extended their model for unsignalized intersections (Section 3.4.1) to include delays at signalized intersections using the "discrete time period" and "block-antiblock" concepts. The blocks are the red phases of the signal and the antiblocks are the green phases.

They developed the following expression for mean delay for all vehicles queued during the red interval:

$$E(t_R) = R\left[E(N_R) + \frac{q}{2}(R + 1)\right] \tag{3.66}$$

in which q is the arrival rate, R is the length of the red interval, and $E(N_R)$ is the mean queue length at the start of the red interval and $E(N_G)$ at the start of the green interval. The mean delay during the green interval is

$$E(t_G) = \frac{1}{2(1 - q)} \times E\left[N_G^2 - N_R^2 + (2q - 1)(N_G - N_R)\right] \tag{3.67}$$

The mean delay per cycle is

$$E(t_c) = \frac{R}{1 - q}\left[E(N_R) + \frac{q}{2}(R + 1)\right] \tag{3.68}$$

The mean delay per vehicle is expressed as

$$E(t) = \frac{R}{(1 - q)(R + G)}\left[\frac{E(N_R)}{q} + \frac{R + 1}{2}\right] \tag{3.69}$$

The only value which must be determined to find mean delay per vehicle is $E(N_R)$. Beckmann *et al.* (*4*) described how the distribution of N_R may be generated by use of the Markov chain technique. Haight's overflow model, discussed later in this section, may also be used to generate the distribution of N_R.

Newell (*53*) derived analytic expressions for the average queue length and average delay under equilibrium conditions for the Beckmann, McGuire and Winsten model. Using the Markov chain approach to obtain the probability of M arrivals per cycle, Newell expressed the average queue length and the average delay in terms of the parameters of the signal (red time, cycle length, etc.) and the probability of an arrival during each time period.

Newell (*54*) considered a simple model for the traffic flow through a loaded intersection controlled by a signal on a narrow two-lane roadway. He described the states of the system as follows:

State 1—Both opposing cars move forward (or turn right) or both turn left and leave the intersection immediately.

State 2—A northbound vehicle wishes to turn left but cannot do so due to interference by opposing forward-moving traffic.

State 3—A southbound vehicle wishes to turn left but cannot do so due to interference by opposing forward-moving traffic.

From these three possible states of the system, Newell developed the probabilities of transition from any of the three states at any time t to any of the states at time $t + \Delta t$. The average number of vehicles able to clear the intersection per signal cycle is expressed in mathematical terms. The resulting general equation is not computationally practical; however, some cases with specific conditions are of interest.

Assume that p is the probability that a vehicle in one direction desires to turn left and p' the probability that an opposing vehicle wishes to turn left.

In the special case where $p = p'$, the capacity of each approach with left turns is

$$q_m = \frac{N(2-p)}{3-2p} + \frac{(1-p)\,[1-(1-p)^N(1-2p)^N]}{p(3-2p)^2} \tag{3.70}$$

in which N is the capacity of each approach with no left turns, expressed in vehicles per cycle. Eq. 3.70 is plotted in Figure 3.28 for various values of N and p.

Newell also investigated the possibility of obtaining an optimum signal cycle. The left-turn values, p and p', were considered as fixed for any given traffic situation and

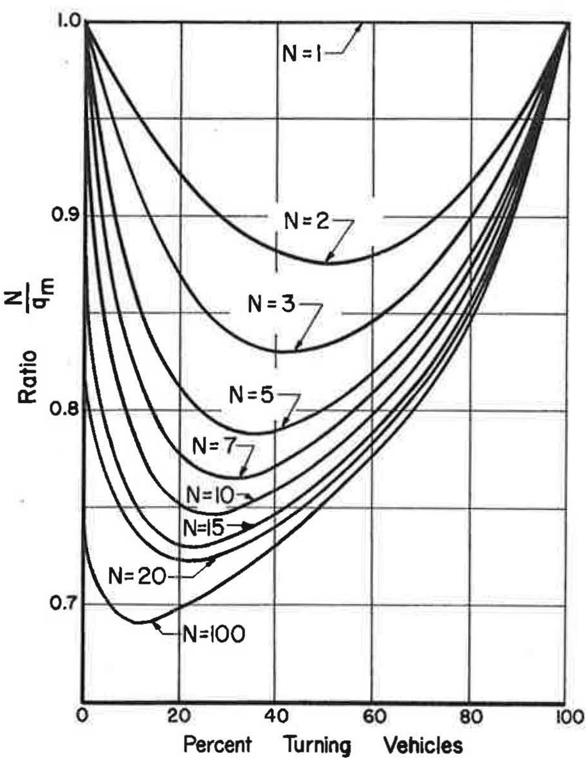


Figure 3.28. The capacity of left-turn movements.

N was varied by changing the total cycle time, C. Newell found that intersection capacity, measured in vehicles per unit time, q_m/c, has a maximum with respect to N if p, the fraction of left turns, is about 1/10 or less.

In the case where $p \neq p'$, the exact formulas for maximum capacities are quite complicated. However, if either p or p' is quite large (Np or $Np' > 1$), an approximation for intersection capacity for one approach may be expressed as

$$q_m \sim \frac{N\{1 + p'[(1-p)/p]\}}{1 + p[(1-p')/p'] + p'[(1-p)/p]} \tag{3.71}$$

Newell (*52*) also considered a two-lane signalized intersection and described the delay to vehicles in terms of the arrival and departure time for each vehicle and the parameters of the signal. Two cases were considered:

Case I—Uniform arrivals.
Case II—Random arrivals.

Using uniform arrivals in Case I, Newell obtained an exact solution for his model. Uniform arrivals, however, are rarely found in the field and to obtain a solution for Case II he made certain simplifying assumptions. The results thus obtained are not exact; however, he indicated that an estimate of the error involved can be obtained.

Haight (*26*) treated the signalized intersection problem by predicting the probability of an overflow queue. He computed the probability that there will be N_R vehicles waiting to cross the intersection at the beginning of a red phase if there were N_G vehicles waiting at the beginning of the preceding green phase, N_R being defined as the overflow into the red phase. Using this approach he derived the probabilities that the queue waiting at the traffic signal would be of various lengths at the beginning of the red and green phases. These probabilities were based on:

(a) The flow on the approach.
(b) The length of red and green phases.
(c) The constant departure headways of the vehicles during the green phase.

Haight's basic assumption was that vehicles arrive at a signalized intersection in such a manner that the probability of their arrival is Poisson distributed. All vehicles move at a speed u through the intersection unless stopped. The vehicle departure headways, h, from the queue are constant. If R is the length of the red phase, the average number of arrivals during a red phase is qR, where q is the flow on the approach being considered. Once the queue waiting at the beginning of the green phase is dissipated, all later arrivals during that green phase continue through the intersection without delay. If the queue cannot be dissipated during the green phase, only N vehicles (an integer) can be discharged; that is,

$$N = \frac{G}{h} - \text{(fraction less than one)}$$

where G is the length of the green phase. Any vehicle arriving at the intersection while a queue exists is assumed to join the queue.

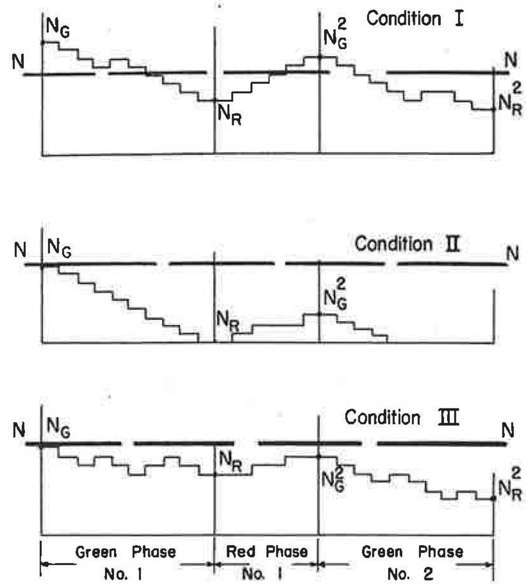


Figure 3.29. Overflow conditions at a signalized intersection.

Three overflow conditions considered are illustrated in Figure 3.29. In Condition I, the number waiting at the beginning of the green phase, N_G, exceeds N. There is an overflow, N_R, of

$$N_R \geq N_G - N \tag{3.72}$$

Because N vehicles will depart from the initial queue of N_G vehicles and arrivals at the intersection during the green phase will be added to the queue, the probability of an overflow, N_R, is

$$P(N_R \mid N_G > N) = P(N_R - N_G + N \mid q\,G) \tag{3.73}$$

in which the right-hand expression is the Poisson probability of an overflow of $(N_R - N_G + N)$, with a flow rate of q and a green phase of length G.

As an illustration, consider a signal which can accommodate 10 vehicles ($N = 10$) during each green phase. With an arrival rate of 3 vehicles per green phase ($q\,G = 3$), Table 3.2 gives the probability that queues greater than 10 vehicles at the start of the green phase will have various overflows at the end of the green phase.

In Condition II, the number of vehicles in the queue at the beginning of the green phase, N_G, is equal to or less than the number of vehicles, N, which can be discharged during the green phase; that is, $N_G \leq N$. With this condition, there is no overflow ($N_R = 0$); the queue is dissipated and later arrivals are not delayed. The probabilities for Condition II can be written in terms of the cumulative Borel-Tanner probabilities as

$$P(N_R = 0 \mid N_G \leq N) = \sum_{j = N_G}^{N} R(j \mid N_G) \tag{3.74}$$

in which $R(j \mid N_G)$, the Borel-Tanner probability, is

Table 3.2 Probability That Queues Greater Than 10 Vehicles in Size Will Have an Overflow

	Probability				
N_R	$N_G =$ 11	$N_G =$ 12	$N_G =$ 13	$N_G =$ 14	......
0	0	0	0	0	......
1	0.05	0	0	0	......
2	0.15	0.05	0	0	......
3	0.22	0.15	0.05	0	......
.	.	.	.	.	
.	.	.	.	.	

$$R(j \mid N_G) = \frac{e^{-\rho j}\,\rho^{j-N_G}\,N_G\,j^{(j-N_G-1)}}{(j - N_G)!}$$

$j = N_G, N_G + 1, \ldots.$ and ρ is the ratio of arrival rate to discharge rate.

In an illustration of Condition II, using an arrival rate of 3 vehicles per green phase, the probability matrix takes the form

N_R	---	N_G 7	N_G 8	N_G 9	N_G 10
0	---	0.16	0.10	0.03	0.002

In Condition III the number of vehicles in the queue at the beginning of the green phase is equal to or less than the number of vehicles which can be discharged during the green phase, but the arrival rate is such that an overflow, N_R, exists at the end of the green phase.

Haight gives the derivation of the probability of various overflows for Condition III and extends these results to give the probability that the queue length will change from N_G to the number waiting, $N_G{}^2$, at the start of the next green phase. He also presents relationships required for the calculation of the probability of N_G and N_R in the steady-state situation.

As an outgrowth of a study of site selection for retail stores, Little (*43*) formulated a series of models for predicting the delay to vehicles in performing various maneuvers under a variety of traffic flow conditions. To obtain models of a practical nature, he used certain approximations which make the results primarily applicable to medium and low traffic flows. The resulting formulas, however, take into account the major variables which contribute to delay and reveal the rather large differences in delay that exist in performing various maneuvers.

In developing the relationship for expected length of queues formed at a traffic signal, Little assumed that:

(a) Arriving traffic is Poisson and in a single lane.
(b) Traffic is held up for a time, T, and then released.
(c) Vehicles starting up leave a constant time, h, between them.
(d) Normal road speed is lost instantaneously on joining the queue and re-

gained instantaneously on starting up.

(e) There are no vehicles turning left or right.

The infinite acceleration and deceleration assumed is not as serious as it might appear. If a vehicle proceeds through the intersection without stopping, there is little or no delay. For those vehicles forced to stop, there will be some additional delay due to deceleration and acceleration. But this can be partially eliminated by using an effective red interval which is equal to the actual red plus the average acceleration delay.

For medium and low traffic flows, in which the carry-over of vehicles from one red interval to the next may be ignored, Little developed an equation for predicting the average queue length at a traffic signal. By using the actual red time R, he approximated T and expressed the average queue length (Fig. 3.30) as

$$\mathrm{E}(N) = \frac{q R}{1 - q h} \tag{3.75}$$

The mean square of the queue length is

$$\mathrm{E}(N^2) = [\mathrm{E}(N)]^2 + \mathrm{E}(N)/(1 - q h)^2 \tag{3.76}$$

Because h (the headway between vehicles leaving the intersection) is assumed to be constant, the average time required for the queue to pass may be expressed as

$$\mathrm{E}(t) = [\mathrm{E}(N)]\, h \tag{3.77}$$

and the mean square time for the queue to pass is

$$\mathrm{E}(t^2) = [\mathrm{E}(N^2)]\, h^2 \tag{3.78}$$

Little has extended this relationship to include multiple lanes. Two cases are considered:

Case I—Arriving vehicles will join the shortest queue at the traffic signal.

Case II—Arriving vehicles form separate and independent streams of traffic.

Letting n denote the number of approach lanes, Case I can be approximated by substituting h/n for h in Eq. 3.77.

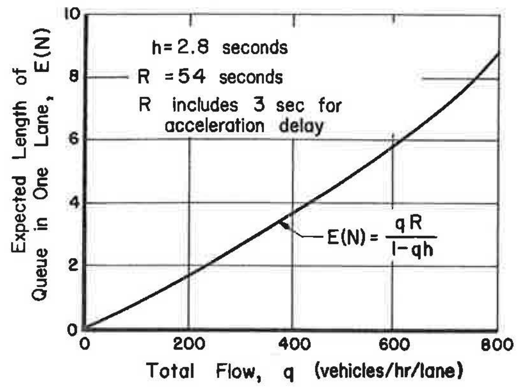


Figure 3.30. Average queue length at a signalized intersection.

For Case II, the separate streams model, the average time for the queue to pass is $M\,h$, where M is the maximum length of queue formed in any lane. Because each of the various lanes is considered as a separate stream of traffic, the average length of queue for each lane can be approximated using Eq. 3.75, in which q for each lane is the total flow divided by the number of approach lanes.

Case II appears to yield a better approximation of the average queue length for most applications when the arrival of traffic is Poisson distributed.

As an example, consider a two-lane approach to a signalized intersection carrying 500 vph. The signal phasing allots 51 sec of red and h is assumed to be 2.8 sec. To adjust for acceleration delay, an effective red time, R_e, is used which is 3 sec longer than the actual red time. This gives a flow of $q = \dfrac{500}{2(3{,}600)} = 0.07$ veh/sec and a mean queue length $\mathrm{E}(N) = \dfrac{0.07\,(54)}{1 - (0.07)(2.8)} = \dfrac{3.75}{0.806} = 4.65$ veh. A plot of this example for various flows, q, is presented in Figure 3.30. This model conforms with the limited amount of field data available.

For a one-lane approach, Little developed the following equation for the average delay to a vehicle passing straight through the intersection:

$$\mathrm{E}(W_1) = \frac{1}{2}\frac{R^2}{C} \times [(1 - q\,h)^{-1} + (q/R)\,q\,h\,(1 - q\,h)^{-2}] \quad (3.79)$$

in which R is the total red time, C is the total cycle time, q is the flow in vehicles per unit of time, and h is the constant starting headway.

The fraction of the vehicles with no delay can be expressed as

$$\mathrm{P}(W_1 = 0) = 1 - \frac{\mathrm{E}(N)}{q\,C} \quad (3.80)$$

in which $\mathrm{E}(N)$ is the average length of queue formed, as expressed by Eq. 3.75.

For multiple-lane approaches, Eq. 3.80 may be used by substituting h/n for h (n is the number of approach lanes). Eq. 3.80 will also yield reasonable results as a right-turn model.

Little's model for the expected delay in making a left turn is based on the following assumptions:

(a) Arriving traffic is Poisson distributed and in a single lane.
(b) The minimum gap required is constant.
(c) If a vehicle arrives during the red phase, it is free to turn in the first acceptable gap that appears in opposing traffic.

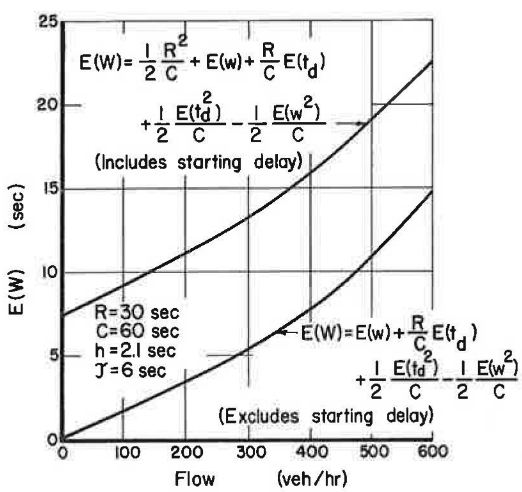


Figure 3.31. Relationship between expected left-turn delay and opposing flow.

(d) If a vehicle is ready to turn during the green phase but cannot before the next red phase, it turns on the red phase.

Assumption (c) neglects the queueing effect because it states that a vehicle will not be delayed by vehicles in its own lane. Little's model is, therefore, restricted to the determination of left-turn delay for a single vehicle. Letting:

τ = the time gap required in the opposing stream for a left turn;
$\mathrm{E}(W)$ = the average wait for an acceptable gap in the opposing stream;
h = the average headway between vehicles starting from a traffic signal;
C = the time for one signal cycle;
t_d = the time required for the opposing queue to dissipate; and
R = the length of the red phase,

Little obtained the following for the average delay to a driver making a left turn at an intersection:

$$\mathrm{E}(W) = \frac{1}{2}\frac{R^2}{C} + \mathrm{E}(w) + \frac{R}{C}\mathrm{E}(t_d) + 1/2\,\frac{\mathrm{E}(t_d^2)}{C} - 1/2\,\frac{\mathrm{E}(w^2)}{C}$$

in which the expected average wait for a gap in the opposing traffic stream having a Poisson arrival rate of q_1 is

$$\mathrm{E}(w) = (1/q_1)(e^{q_1\tau} - 1 - q_1\tau) \quad (3.81)$$

the mean square wait is

$$\mathrm{E}(w^2) = 2\,[\mathrm{E}(w)]^2 + \mathrm{E}(w)(2/q_1) - \tau^2 \quad (3.82)$$

the average time required for the opposing queue to pass is

$$\mathrm{E}(t_d) = \frac{q_1\,R\,h}{(1 - q_1\,h)} \quad (3.83)$$

and the mean square of the average time required for opposing queue to pass is

$$\mathrm{E}(t_d^2) = \mathrm{E}(t_d)^2 + \frac{\mathrm{E}(t_d)\,h}{(1 - q_1\,h)^2} \quad (3.84)$$

Assuming values for each parameter, Figure 3.31 shows the relationship between average left-turn delay and opposing flow. This model also conforms with a limited amount of field data. However, Little concluded that more extensive data are required for definite confirmation of the model.

3.4.2.2 **Timing of Traffic Signals.** In order to minimize delay at an intersection, it is necessary to determine how much of the total time available at the intersection will be apportioned to each traffic movement. Castoldi (7) treated the problem of minimizing delay at signalized intersections by considering the lengths of queues that will develop on all approaches during the respective red phases. He developed equations for establishing appropriate phase lengths for two conditions:

Condition I—The crossing of two vehicular streams.

Condition II—The crossing of two vehicular streams and two pedestrian streams.

Castoldi made the following assumptions:

(a) As the queue of traffic on one approach is dissipated, the opposing stream begins to move through the intersection.
(b) Each vehicle accelerates at the same rate until it reaches the mean speed of its traffic stream.
(c) Waiting times for both streams of traffic, as dictated by the respective red phases, are equal to or larger than the time necessary to dissipate the normal queue buildup in the opposing stream.
(d) The two streams of traffic are moving in direction i and j.

The appropriate lengths of red time for the two traffic streams under condition I are obtained by simultaneous solution of

$$R_i = K_i R_j + \sqrt{a_i R_j + b_i} \qquad (3.85)$$

$$R_j = K_j R_i + \sqrt{a_j R_i + b_j} \qquad (3.86)$$

in which

R_i = length of red phase for direction i;
$\bar{u}_i$ = mean speed of traffic stream i, ft/sec;
$\bar{U}_i$ = mean speed at which traffic stream i moves out from a stopped position at a traffic signal, ft/sec;
x_i = length of traffic queue along the ith approach per unit red time assigned to the i traffic stream, ft/sec;
$K_i = x_i\left(\dfrac{1}{\bar{u}_i} + \dfrac{1}{\bar{U}_i}\right)$, sec;
d_i = intersection width which traffic stream i must cross, ft;
$t_i = d_i/\bar{U}_i$ = time, in sec, to cross intersection of width d_i;
a = acceleration of vehicles proceeding from stopped position up to the mean speed of the traffic stream;
$\alpha_i = 2x_i/a$; and
$\beta_i = 2d_i/a$.

The normal length of queues, N, that will develop when red signal phases of length R_i and R_j are utilized, are given by

$$N_i = x_i R_i \qquad (3.87)$$

Selection of proper time phasing for condition II, the crossing of two vehicular streams and two pedestrian streams, may be obtained through simultaneous solution of

$$R_i = (K_j + 1/2) R_j + t_j \qquad (3.88)$$

$$R_j = 2 K_i R_i + 2 \alpha_i R_i + \beta_i \qquad (3.89)$$

For examples of the use of Eqs. 3.85, 3.86, and 3.87, consider an intersection with the following characteristics:

$d_1 = 40$ ft	$x_2 = 25$ ft/sec
$x_1 = 12$ ft/sec	$\bar{U}_2 = 20$ ft/sec
$\bar{U}_1 = 24$ ft/sec	$\bar{u}_2 = 25$ ft/sec
$\bar{u}_1 = 30$ ft/sec	$a = 5$ ft/sec
$d_2 = 30$ ft	

Then,

$K_1 = 12\,(1/30 + 1/24) = 5/6$	$K_2 = 5(1/25 + 1/20) = 9/20$
$h_1 = 40/36 = 1.11$ sec	$h_2 = 30/25 = 1.20$ sec
$\alpha_1 = 24/5 = 4.8$ sec	$\alpha_2 = 15/5 = 3.0$ sec
$\beta_1 = 80/5 = 16$ sec^2	$\beta_2 = 60/5 = 12$ sec^2

Solving Eqs. 3.85 and 3.86 simultaneously, R_1 and R_2 are found to be 22.0 sec and 29.4 sec, respectively. Substituting these values in Eq. 3.87, the respective normal queue lengths are 264 and 735 ft. Utilizing the same data for condition II as in condition I gives $K_2\,(2K_1 + 1) > 1$, which means that the queues are increasing without limit, causing the system to become more and more congested. The same situation could apply for condition I if $K_1 K_2 \geq 1$.

In a second approach to apportioning

time between phases of a traffic signal, Uematu (*73*) suggested that the time apportionment be determined by the lengths of the waiting lines on the two approaches. A "random walk" concept was utilized to describe the lengths of queues X_n and Y_n of the north-south and the east-west flows, respectively, for the nth cycle, C_n. The fundamental equations are:

$$X_n = (X_{n-1} + x_n - q\,G_X) + x'_n \quad (n = 1, 2, \ldots) \tag{3.90}$$

$$Y_n = (Y_{n-1} + y_n - G_Y) + y'_n \quad (n = 1, 2, \ldots) \tag{3.91}$$

and

$$C_n = G_X + G_Y \tag{3.92}$$

in which

C_n = nth cycle;
X_n, Y_n = length of queue (number of vehicles in cycle C_n);
X_{n-1}, Y_{n-1} = length of queue (number of vehicles remaining from previous cycle, C_{n-1});
x_n, y_n = number of arrivals during green phase;
x'_n, y'_n = number of arrivals during red phase;
q = rate of departure, in veh/sec; and
G_X, G_Y = length of green phase, in sec.

If A and B represent the maximum queue lengths desired on the north-south and east-west highway approaches, respectively, the phase lengths G_X and G_Y must be chosen so as to minimize the probabilities of the waiting lines exceeding A and B. Uematu formulated the transition probabilities which describe the system for any cycle length and derived solutions for some special cases.

3.4.2.3 Delays at Traffic-Actuated Signals. In the study of actuated signals, Haight (*26*) expanded his overflow theory to include semi-actuated signals. His assumptions were predicated on the use of a semi-actuated signal controller, which provides for minimum green and clearance intervals for the main-street traffic and initial, vehicle, maximum green, and clearance intervals for the side-street traffic.

Assuming random arrival rates on all approaches, Haight stated that after obtaining explicit formulas for the distributions of the main-street red phases and the number of vehicles which can proceed through the intersection during the main-street green phases, the overflow probabilities can be calculated by proper substitution of the various parameter values in the equations for the fixed-time overflow conditions.

At an intersection controlled by a fully-actuated signal, the vehicles delayed on any approach must wait until either a specified gap in the opposing traffic or the end of the maximum green interval causes the signal to change.

Utilizing this principle, Garwood (*16*) applied the Poisson distribution to the operation of fully actuated signals. He recognized, however, that the conditions for a Poisson series are not strictly satisfied. He found no significant differences between the theoretical and observed values for several delay characteristics, including frequency of waiting periods and percentage of waiting periods which reach the maximum time allowed. Starting with the assumption that traffic is flowing in the north-south direction and in the east-west direction, Garwood showed that the probability that the first east-west vehicle crossing one of the east-west detectors and having to wait for a time period, t, greater than time, T, is

$$P(t > T) = \sum_{N=1}^{\infty} \frac{e^{-qT}(qT)^N}{N!}\, P\left(\frac{I}{T}\right) \tag{3.93}$$

in which

q = north-south flow;
N = number of north-south vehicles arriving during the north-south maximum period after the arrival of the initial east-west vehicle;
I = north-south vehicle interval; and
$P(I/T)$ = probability that the headways of all north-south vehicles arriving during the north-south maximum period, after the arrival of an east-west vehicle, are all less than I/T.

The problem encountered in making use of Eq. 3.93 is the need to define the probability, $P(I/T)$. Solution of this probability involves finding the number of different ways of arranging the N number of vehicles into $(N - 1)$ groups. Expressed mathematically, this probability is

$$P(I/T) = 1 - C_1^{N+1}(1 - I/T)^N + C_2^{N+1}(1 - 2I/T)^N - \dots + (-)\,C_X^{N+1}(1 - XI/T)^N \quad (3.94)$$

in which X is the integral part of $1/(I/T)$ or T/I, and C_1^{N+1} is the combination of $(N + 1)$ items taken one at a time, and the other terms are as defined earlier. Solution of Eq. 3.94 for varying values of T/I and the expected number of north-south vehicles, qT, is shown in Figure 3.32. Figure 3.33 gives the probability that the waiting period will reach various maximums as determined by changes in vehicle intervals and the intensity of the main stream traffic.

Garwood also showed that the expected waiting time of an initial arriving east-west vehicle is

$$E(t) = q^t - \frac{qT}{2!}\left(1 - \frac{I}{T}\right)(qT - qI + 2) \times e^{-qI} + \frac{qT^2}{3!}\left(1 - \frac{2I}{T}\right)^2 \times (qT - 2qI + 3)\,e^{-2qI} + \dots + (-)\,\frac{qT^X}{(X+1)!}\left(1 - \frac{XI}{T}\right)^X \times (qT - XqI + X + 1)\,e^{-XqI} \quad (3.95)$$

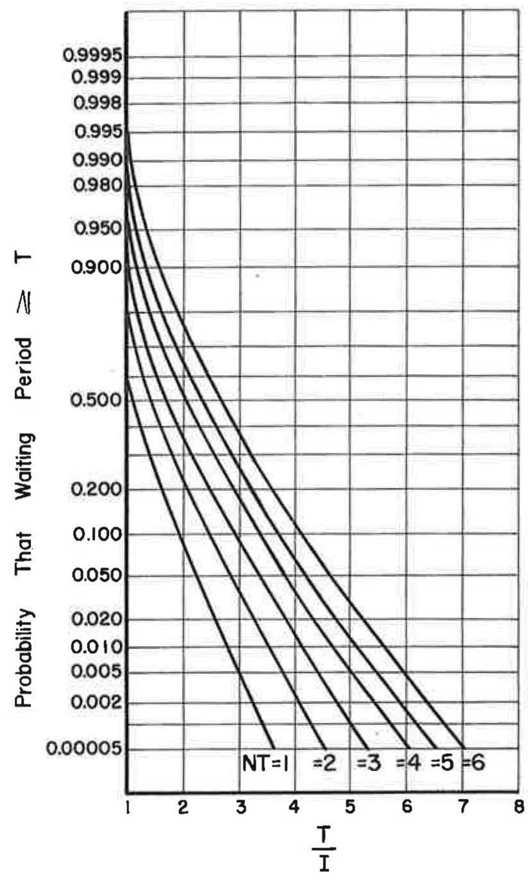


Figure 3.32. Probability of waiting period lasting longer than time T for various vehicle intervals and numbers of vehicles per unit time in opposing stream.

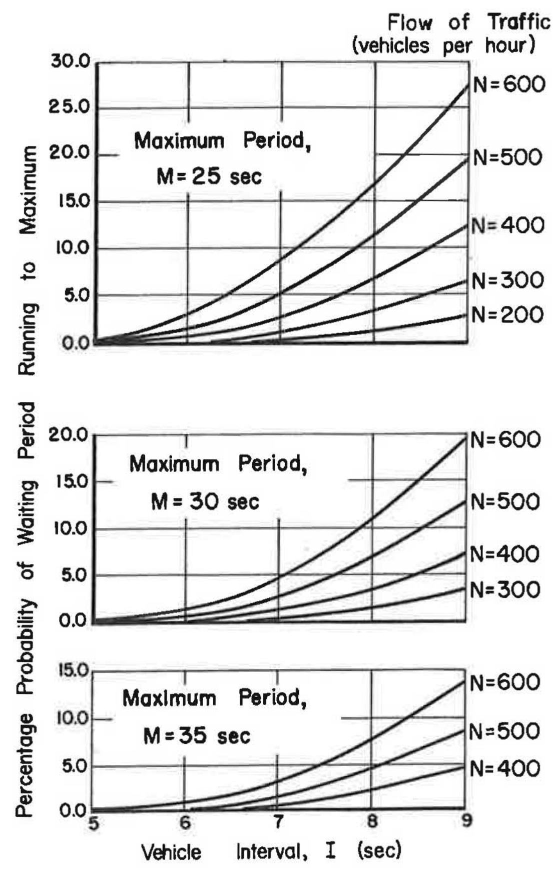


Figure 3.33. Probability of waiting period running to maximum for various values of the maximum period, the vehicle interval, and the main-stream traffic intensity.

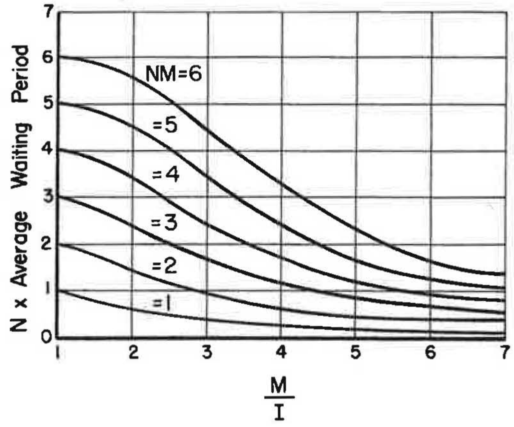


Figure 3.34. Average waiting period for maximum period, M, vehicle interval, I, and number of vehicles per unit time in opposing stream, N.

in which all terms are as defined earlier. Figure 3.34 illustrates solutions of Eq. 3.95 for varying values of M/I (M = maximum period) and the expected number of north-south vehicles, qT.

3.5 DELAYS ON TWO-LANE ROADS

The delay experienced by vehicles while traveling on two-lane roads in accordance with postulated rules has been of particular interest to the traffic flow theorist. Each vehicle, if not interrupted, will travel at its own desired speed. When a slower vehicle or group of vehicles is overtaken, passing without delay will occur if there is an acceptable gap in the opposing stream of vehicles. If an acceptable gap for passing is not available in the opposing stream, the faster vehicle will be required to assume the speed of the slower vehicle or queue of vehicles and follow until an opportunity to pass occurs.

The opportunities for passing were studied in detail by Greenshields (*20*) early in 1935 and more recently by Tanner (*69*), Miller (*46*), Kometani (*41*), and Newell (*56*). If a vehicle with velocity u is to pass a vehicle with velocity u_1, the passing maneuver requires a time of

$$t = \frac{A_1}{u - u_1} \tag{3.96}$$

and a distance of

$$x = \frac{A_1 u}{u - u_1} \tag{3.97}$$

in which A_1 is a parameter describing the distance required for the passing vehicle relative to the vehicle being passed.

This section describes the various hypotheses that have been applied to the probability model for a two-lane road.

3.5.1 Tanner's Model

Tanner's model (*69, 70*) deals with vehicles traveling in both directions along a two-lane road and can be extended to one-way flow on a two-lane facility. Referring to Figure 3.35, the flow in one direction is q_1 vehicles per unit of time. All vehicles travel at the same constant speed u_1 except the one vehicle under study, which travels at some greater desired speed u, or at speed u_1 if it is unable to pass. The minimum spacing of vehicles in this stream is S_1. In the opposite direction the flow is q_2, with all vehicles traveling at the same constant speed u_2 and with no spacing less than S_2. The Borel-Tanner distribution is assumed for the number of vehicles n in the "bunches." The distribution of gaps is a modification of random arrivals which requires vehicles to be moved backward in the stream such that no spacing is less than the minimum.

The delays experienced by the single vehicle traveling at speed u in the q_1 flow direction is the problem for which Tanner offered a model. For the solution of this problem the vehicle is assumed to act in accordance with the following rules:

(a) A group of n vehicles in the q_1 stream traveling at their minimum separation S_1 is overtaken in a single maneuver. The overtaking vehicle can only re-enter the q_1 lane between

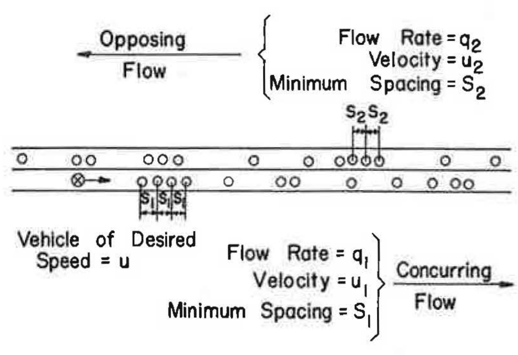


Figure 3.35. Two-lane roadway, showing the corresponding assumed terms in the Tanner model.

two groups and cannot break into any one group or approach the rear vehicle of any group by a distance less than S_1.

(b) When the overtaking vehicle reaches the tail of any group of n vehicles and there is a distance of at least

$$d_n = d + n\,S_1\,(u + u_2)^2/(u - u_1)$$

in the q_2 stream, the vehicle passes without slowing down. (d is defined as the least acceptable clear distance between the u vehicle and the opposing traffic as the u vehicle clears the bunch in passing. It can be expressed by $d = A_1 u/(u - u_1)$, with A_1 being some distance between 50 and 100 ft.)

(c) If the required distance d_n is not available, the vehicle decelerates instantaneously to speed u_1, follows as closely as possible behind the vehicle ahead, waits for a clear distance of at least $D_n = d_n + t\,(u_1 + u_2)$ in the q_2 stream, waits a further time t, accelerates instantly to speed u, and passes. t, defined as the additional time required for the overtaking vehicle to remain in the q_1 stream because of having slowed down, is used to compensate for the assumed instantaneous acceleration and could be expressed as

$$t = \frac{A_2\,(u - u_1)}{a}$$

in which a is the constant acceleration of the overtaking vehicle and A_2 is approximately one.

Tanner's major objective was to determine the average speed $E(u)$ of a single vehicle which desires to travel at a velocity u over an infinitely long trip. He was able to express the average speed $E(u)$ in terms of the average waiting time behind all vehicles, $E(t_w)$, which included zero waits. The expression

$$E(u) = \frac{u\,u_1^2 + q_1\,(u - u_1)\,(u_1 - s_1\,q_1)\,u_1\,E(t_w)}{u_1^2 + q_1\,(u - u_1)\,(u_1 - s_1\,q_1)\,E(t_w)} \tag{3.98}$$

was developed, thus the problem was resolved into that of solving for $E(t_w)$. Algebra involved in the computation of $E(t_w)$ is formidable. The expression for $E(t_w)$ is

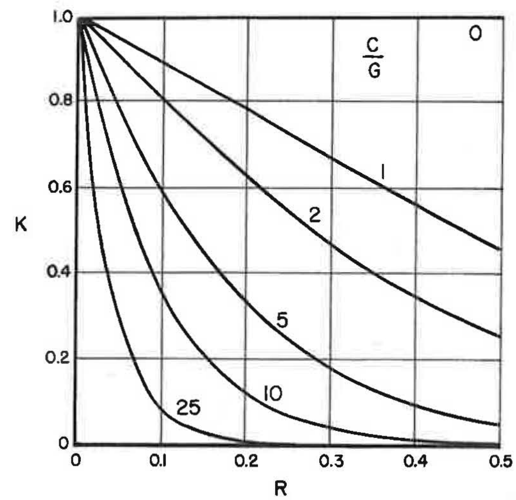


Figure 3.36. Relationship between K and the parameters R and C/G.

$$E(t_w) = \left(\frac{1}{(u_1 + u_2)\,(1 - R)}\right) \left[\left(\frac{K\,e^{-cd}}{c + G - G\,K}\right)\left(1 - \frac{c \exp\,[G(u_1 + u_2)t]}{c + G}\right) + (\exp\,[Gt\,(u_1 + u_2) + Gd)] \left(\frac{N}{G} - \frac{c}{G(c + G)}\right) - \frac{1 - R}{c(1 - n)}\right] \tag{3.99}$$

in which $g = q_1/u_1$, $G = q_2/u_2$, $r = S_1\,g$, $R = S_2\,G$, $c = q_1\,(u - u_1)/u_1\,(u + u_2)$, K = root between 0 and 1 of $K = \exp\,[R(K - 1 - c/G)]$, and N = the smaller real root of $N = \exp\,[n(N - 1 + G/c)]$ (which exists only when $r \exp\,(1 - n + n\,G/c) \leq 1$).

Limited solutions for K and N have been included in Figures 3.36 and 3.37, respectively. It is apparent that $E(t_w)$ is a function of q_1, q_2, u_1, u_2, S_1, S_2, d, t, a and u.

Substitution of the values of $E(t_w)$ in Eq. 3.98 gives an expression for the average speed $E(u)$ in terms of the desired speed u, the velocity of the q_1 stream u_1, and the flow rate q_1 of the stream. Limited solutions of this equation were made by Tanner using specific values of the various parameters. Figure 3.38 shows the effect of traffic flow when $q_1 = q_2$ for various values of u, and u_1 = 30 mph. This model indicates that, for a total flow of more than 800 veh/hr, a vehicle will have to assume very nearly the velocity of the q_1 stream, regardless of its own desired velocity.

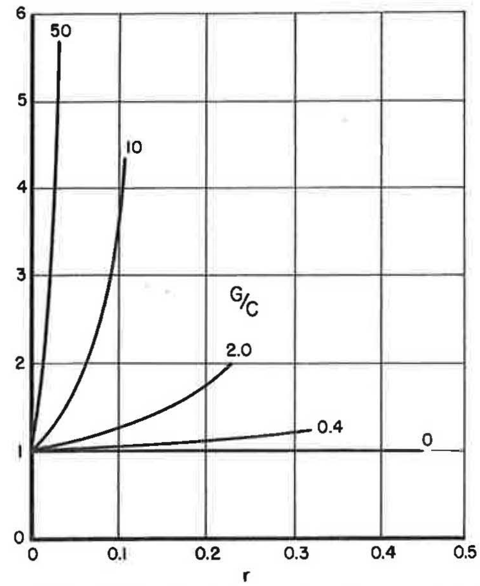


Figure 3.37. Relationship between N and the parameters r and G/C.

The effect of varying proportions of q_1 and q_2 on the average speed $E(u)$ is shown in Figure 3.39, which shows that, in this model, the average speed $E(u)$ is least when one-half to three-fourths of the total traffic is traveling in the opposite direction of flow q_2 with one-half being applicable to low volumes and three-quarters applying to higher volumes.

It is worthwhile to point out that the delay implied by $E(u)$ is the only delay involved and all other vehicles are, by the assumptions, not delayed. The u vehicle would pass all q_1 vehicles ultimately and no passing would occur among q_1 or q_2 vehicles.

3.5.2 Miller's Model

In a recent article by Miller (*47*), a model was developed to estimate the delays to vehicles on a rural two-lane roadway. Miller used a random distribution of "bunches" or queues and a one-parameter distribution of the number of vehicles in the queues, as discussed in Section 3.2.2. An empirical relationship between φ, the number of overtakings per unit of time, and the opposing flow q_2 was developed. In fact, he noted that there was a linear relationship between log φ and log q_2 for the field data he used from a study in Sweden.

This model assumes that there is an intensity of ρ bunches per mile and the average road space occupied by queued vehicles is $\bar{S}$. It is expected that drivers wishing to travel at a speed, u, will be compelled to travel for a portion of the time at a slower speed, u_1, the speed of the queue which they have overtaken. The concentration, or density, in vehicles per mile is k. The ratio, ρ_1/ρ, is a measure of free vehicles on the roadway. It is the fraction of bunches which contain only one vehicle. The standard deviation of the distribution of the velocities of bunches is denoted by σ.

The expected delay to vehicles per mile of road per unit of time is expressed as

$$E(t_d) = k\left(\frac{u_1}{u} - 1\right) \log_e (\rho_1/\rho) \tag{3.100}$$

In this expression, ρ_1/ρ may be determined by

$$\frac{\rho_1}{\rho} = 1 - \frac{0.56\,\sigma\,k\,[1 + \log_e(\rho_1/\rho)]}{(1 - k\bar{S})} \tag{3.101}$$

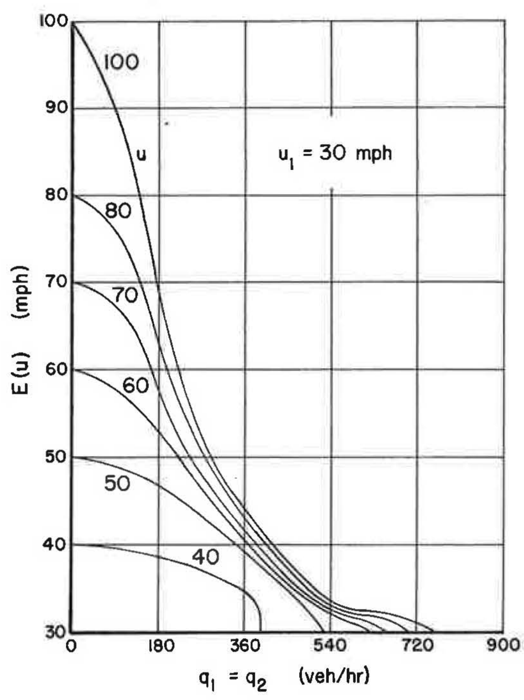


Figure 3.38. Effect of traffic when equally divided between directions for various values of u (two-way traffic).

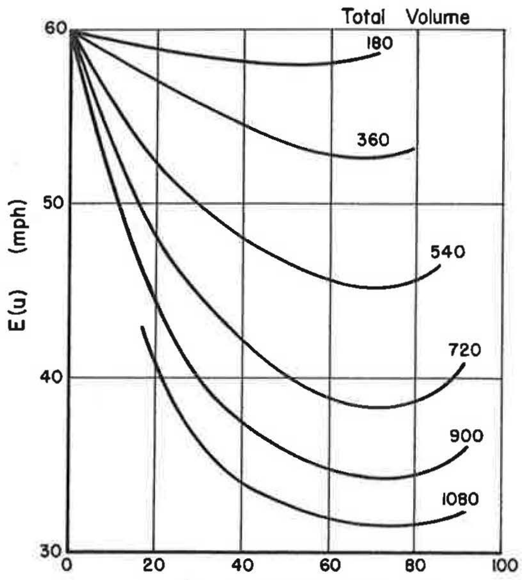


Figure 3.39. Effect of varying proportion of opposing traffic for various levels of total traffic.

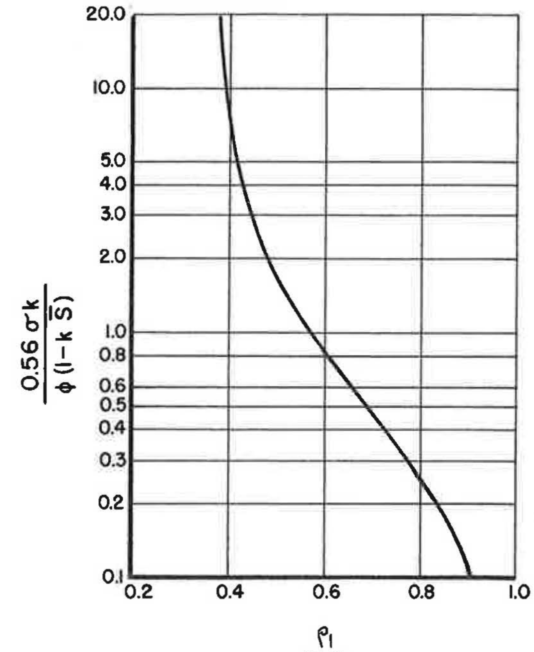


Figure 3.40. Solutions of Eq. 3.101.

Figure 3.40 shows the solutions of Eq. 3.101 for various values of $\frac{0.56\,\sigma\,k}{\varphi\,(1-k\,\bar{S})}$. For values of ρ_1/ρ near one, the ratio may be approximated by

$$\rho_1/\rho \sim \frac{\varphi\,(1-k\,\bar{S})}{1+0.56\,\sigma\,k} \qquad (3.102)$$

Miller used the following constants for a solution of his model: $\bar{S} = 158.4$ ft (0.03 mi), $\sigma = 10$ mph, and $u_1/u = 2/3$. These solutions are plotted in Figure 3.41 as the relationship between the density, k, and the rate of delay per mile of road. The two curves are for opposing flows of 650 vph, which corresponds to 50 passings per hour for the Swedish drivers, and 200 vph, which corresponds to 100 passings per hour.

3.5.3 Kometani's Model

Kometani (*41*) allowed for a finite number of different speed vehicles in the northbound lane $q_1, q_2, \ldots, q_r$ and a finite number of different speed vehicles in the southbound lane $q'_1, q'_2, \ldots, q'_r$ in developing his two-lane model. It is then apparent that

$$V = q_1 + q_2 + \ldots + q_r$$

and

$$V' = q'_1 + q'_2 + \ldots + q'_r$$

in which V and V' are the northbound and southbound traffic volumes. Letting U denote the average speed of the overtaking vehicle, u the average speed of the overtaken vehicles, S the least space headway in which the passing vehicle can follow a slower vehicle without decelerating, s the least space headway in which the low-speed vehicle can follow the vehicle overtaking it, l the least space headway between low-speed vehicles, and n the number of slower-moving

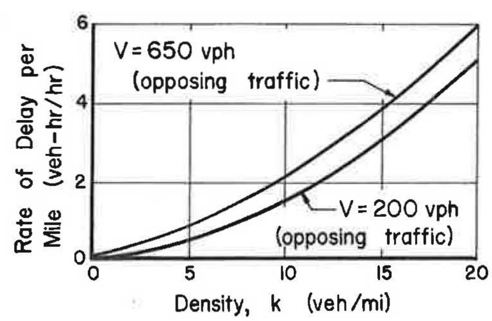


Figure 3.41. Rate of delay versus density.

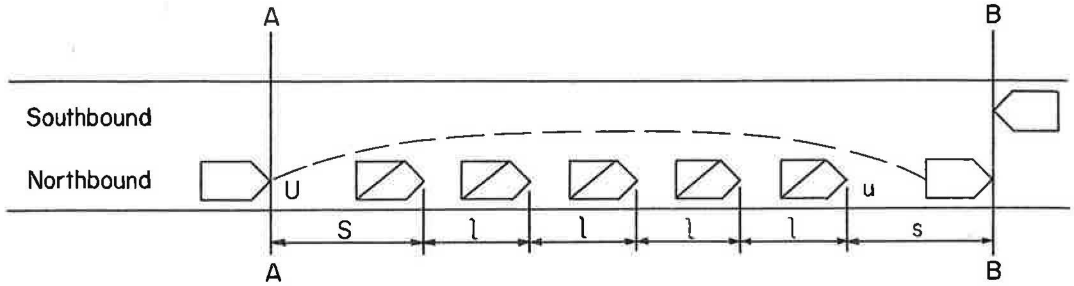


Figure 3.42. Passing n continuous slower-moving vehicles on a two-lane highway.

vehicles in the queue being overtaken, Kometani derived the probability of finding a time gap τ_n in N successive independent trials at intervals of τ'_n as follows:

$$P_{N\tau'_n} = \left[\frac{V_1}{V} \sum_{n=1}^{\infty} \frac{(Vt)^n e^{-Vt}}{n!} \right] \times \left[1 - (1 - e^{-V'\tau_n})^{(N+1)} \right] \quad (3.103)$$

in which

$$\tau'_n = \frac{S + s + (n-1)\,l}{U - u} \quad (3.104)$$

$$\tau_n = \frac{S + s + (n-1)\,l}{U - u} \times \left[2 + \left(\frac{U}{u} - 1 \right) \frac{V'_1}{V'} \right] \quad (3.105)$$

$$t = \frac{S + s}{u} \quad (3.106)$$

V'_1 = volume of low-speed vehicles in opposing lane; and

V_1 = volume of low-speed vehicles moving in direction of passing vehicles.

Assuming a Poisson distribution, the passing phenomenon expressed in Eq. 3.103 can only occur when at least one northbound vehicle belonging to V arrives during the interval t (first term of Eq. 3.103) and no southbound vehicle belonging to V' appears during time τ_n (second term of Eq. 3.103). The value τ'_n in Eq. 3.104 is the time required to pass n slower-moving vehicles in a queue (see Fig. 3.42). The value chosen for τ_n is the sum of the time required to pass, τ'_n, plus the time for the southbound car being met to travel from B to A (Figure 3.42). Figure 3.43 presents a graph of Eq. 3.103 for mean speeds of 20 and 40 mph for low-speed and high-speed vehicles, respectively, an equal directional distribution of traffic volume ($V = V'$), and several values of N.

Using the same parameters and assumptions as for his two-lane model, Kometani derived the probability of being able to pass

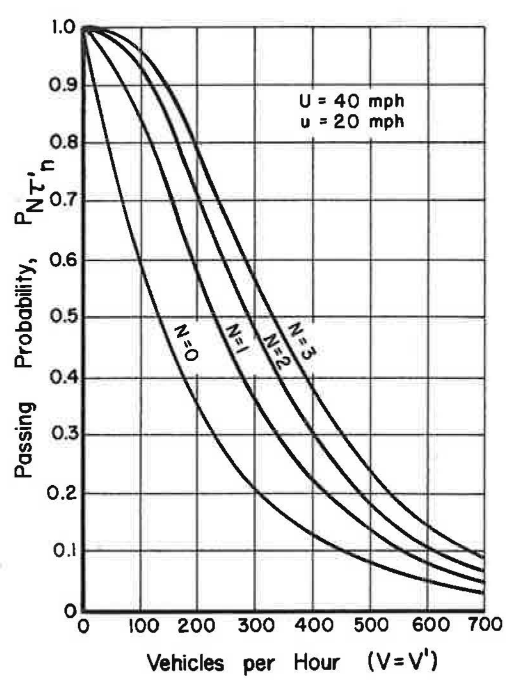


Figure 3.43. Passing probability on a two-lane highway.

on three-lane roads. The computed and measured values of passing probabilities for the two-lane and the three-lane roads do not differ greatly.

3.6 SPECIAL DELAY TOPICS

There are a number of other delay situations for which probabilistic approaches have been used. Among those discussed in this section are merging, one-lane bottlenecks, peak flow, multiple queues, and parking.

3.6.1 Merging Delays

Merging may be defined as the absorption of one stream of traffic by another. This traffic phenomenon occurs when a vehicle joins a through traffic stream with the appearance of a minimum acceptable gap. Although a complete mathematical model for the merging condition has not been formulated, several variations of the merging problem have been studied in an effort to obtain a better understanding of the complexity of the merging situation.

Oliver and Bisbee (*62*), in their study of the merging problem, postulated that the minor stream queue lengths are a function of the major stream flow rates. Assuming that:

(a) A gap of at least τ is required for entry into the major stream;
(b) Only one entry is permitted per acceptable gap;
(c) Entries occur just after the passing of the vehicle that signals the beginning of a gap of acceptable size;
(d) Appearance of gaps in the major stream is not affected by the queue in the minor stream; and
(e) Arrivals into the minor stream queue are Poisson;

they found the average number of vehicles in the minor stream queue to be

$$E(n) = \frac{(q_a/q_b)^2\,(1 - q_b\,\tau e^{-\tau q_b})}{[e^{-\tau q_b} - (q_a/q_b)]\,e^{-\tau q_b}} \tag{3.107}$$

in which q_a is the minor stream flow, q_b is the major stream flow, and τ is the minimum acceptable gap. Figure 3.44 shows a family of curves relating the average length of the minor stream queue, E(n), and the major stream flow rate, q_b, and a value of $\tau = 5$ sec for several minor stream flow rates. This model works particularly well for situations in which the major stream flow rate is high and the vehicles in the minor stream queue are served on a first-

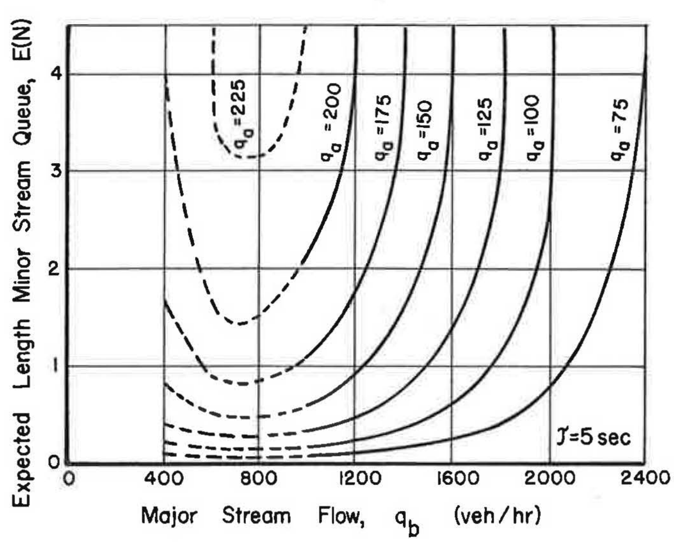


Figure 3.44. Relationship of the average minor stream queue length and the major stream flow rate. Dashed curves represent range in which this model cannot be expected to yield a reasonable approximation of the average queue length.

come first-served basis with the appearance of a minimum acceptable gap length of τ.

Haight, Bisbee and Wojcik (*24*) discussed certain mathematical aspects of the merging problem and gave approaches for dealing with a limited number of special cases. Using vehicle performance characteristics, a method for determining the safe gap to merging was developed. A relationship for the probability of success for a vehicle to merge within a certain distance while moving at a constant velocity was also presented.

Ho (*32*) formulated a model to predict the amount of time required to clear two joining traffic streams through a merging point. This model assumes that:

(a) Merging is permitted only at the merging point; that is, the point just prior to which the merging vehicle will be forced to stop due to obstruction on a blocked lane of a multilane roadway, etc.

(b) Each vehicle entering the merging stream must join those waiting before or at the instant when the car in front begins to merge (first-come first-served condition).

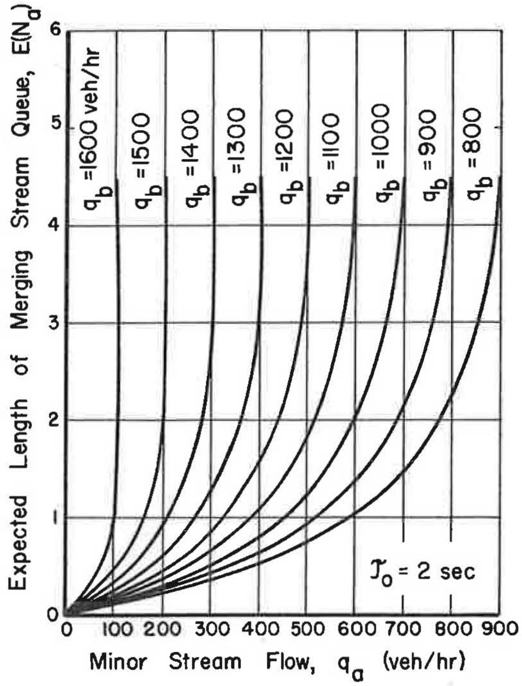


Figure 3.45. Expected length of queue in the merging stream, in terms of the average flow of the major and minor streams.

The total time required for n_1 and n_2 vehicles to pass through the merging point is

$$T = \sum_{i=1}^{n_1 - 1} (h_i - t_0 \alpha) + n_2 t_0 \quad (3.108)$$

in which

h_i = ith time gap between vehicles on the through-traffic road measured at the merging point;

t_0 = time required for a vehicle to merge into the through-traffic stream (assuming all merging vehicles require the same amount of time);

α = number of vehicles of the merging stream which merge into the ith gap of the through-traffic streams at the merging point;

n_1 = number of vehicles in the through-traffic stream; and

n_2 = number of vehicles waiting to merge into a stream of n_1 vehicles.

Ho stated that "the information obtained is a direct measure of the efficiency of the physical system under consideration and may be of some usefulness to highway construction planning and emergency evacuation planning."

Oliver (*59*) has formulated a model for the merging of two streams of high-speed traffic. A typical example of this case is the freeway on-ramp which has an acceleration lane. Because the merging stream will be operating at a speed very near the speed of the major stream, the required size of gap for merging is considerably smaller than that required if the merging vehicles were required to stop prior to merging with the major stream.

Considering the minimum headway between vehicles to be τ_o, Oliver has solved for the expected length of queue in the merging stream, $E(N_a)$, and the expected average delay to a vehicle in the minor stream, $E(W_a)$, in terms of the average flow of the major and minor streams, q_b and q_a, respectively. The expression for the expected length of queue in the minor or merging stream is

$$E(N_a) = \frac{q_a \tau_o (1 - q_a \tau_o)}{1 - q_b \tau_o - q_a \tau_o} \quad (3.109)$$

This relationship is plotted in Figure 3.45 for a value of $\tau_o = 2$ sec. The expected length of time a minor stream vehicle will be delayed is

$$E(W_a) = \frac{\tau_o}{2} + \frac{\tau_o(1 - q_a\tau_o)}{1 - (q_a + q_b)\tau_o} - \frac{\tau_o\,[1 - (q_a + q_b)\tau_o]}{1 - q_b\tau_o} \tag{3.110}$$

In any case where either $q_a\,\tau_o \approx 1$ or $(q_a + q_b)\,\tau_o \approx 1$, the average delay will be very large. A plot of this equation is shown in Figure 3.46.

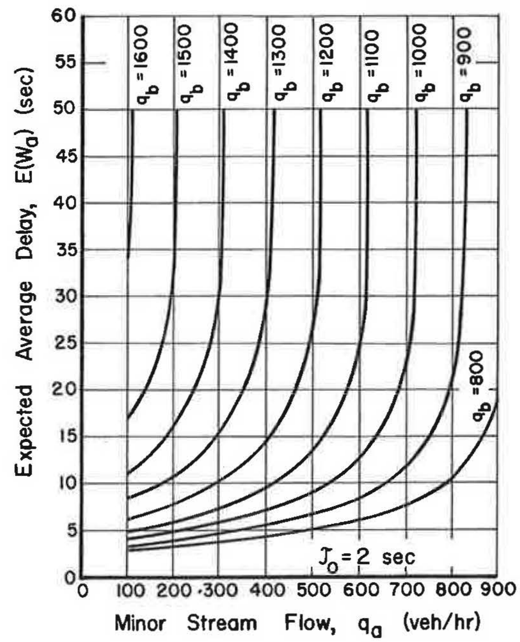


Figure 3.46. Expected average delay to a vehicle in the merging stream, in terms of the average flow of the major and minor streams.

3.6.2 One-Lane Operation

Tanner (*68*) has considered the problem of the delays that occur when opposing streams of traffic on a two-lane road must pass through a one-lane section. This type of operation is encountered when maintenance crews work on one of the two lanes. In Tanner's model, traffic is permitted to control itself. A vehicle upon reaching the beginning of the one-lane section proceeds ahead if there are no opposing vehicles occupying the one-lane section, or if a vehicle moving in the same direction is within the one-lane section. Tanner derived the mean waiting time of a vehicle in terms of the mean and variance of the distribution of length of period when one direction of movement controls the one-lane section. If one considers the two streams as moving in opposite directions i and j, the mean waiting time for i-bound traffic is

$$E(t) = \frac{1}{2\,(1 - \rho_i)} \left(\frac{\rho_i}{\mu_i} + \frac{d_i\, m_j\, (1 - \rho_i - \rho_j)}{\alpha + B_i\, d_j + B_j\, d_i} \right) \tag{3.111}$$

in which

λ_i = the flow in direction i;
μ_i = the capacity in direction i when direction i has control of the one-lane section;
$\rho_i = \lambda_i/\mu_i$;
$\rho_i + \rho_j < 1$;
t_i = the length of time flow is controlled by direction i;
$M_i = E(e^{-\lambda_j t_i})$;
$m_i = E(t_i^2)$;
$\alpha = M_i + M_j - M_i\,M_j$;
$d_i = \lambda_i + \lambda_j - \lambda_i\,M_i$;
T_i = the time required for a vehicle moving in direction i to pass through the one-lane section;

$$B_i = \frac{1}{\lambda_i}\left[\exp\,[\lambda_i\,(T_i - 1/\mu_i)] - 1\right]$$

Tanner provided several explicit solutions for one-lane operation. For example, he showed that if the minimum headway between vehicles moving in the same direction is assumed to be zero, then μ_i and μ_j equal infinity, and the mean waiting period of a vehicle for varying opposing traffic flows is as shown in Figure 3.47. (Normally, $T_i = T_j$, as the time required for a vehicle to pass through the one-lane section would be the same for both directions of travel. When the times are equal, flow is expressed as vehicles per unit time; when not equal, traffic flow is expressed as vehicles per unit time per bottleneck travel time.)

3.6.3 Peak Flows

So far, only applications have been considered in which the traffic intensity, ρ, is less than unity; that is, the mean service

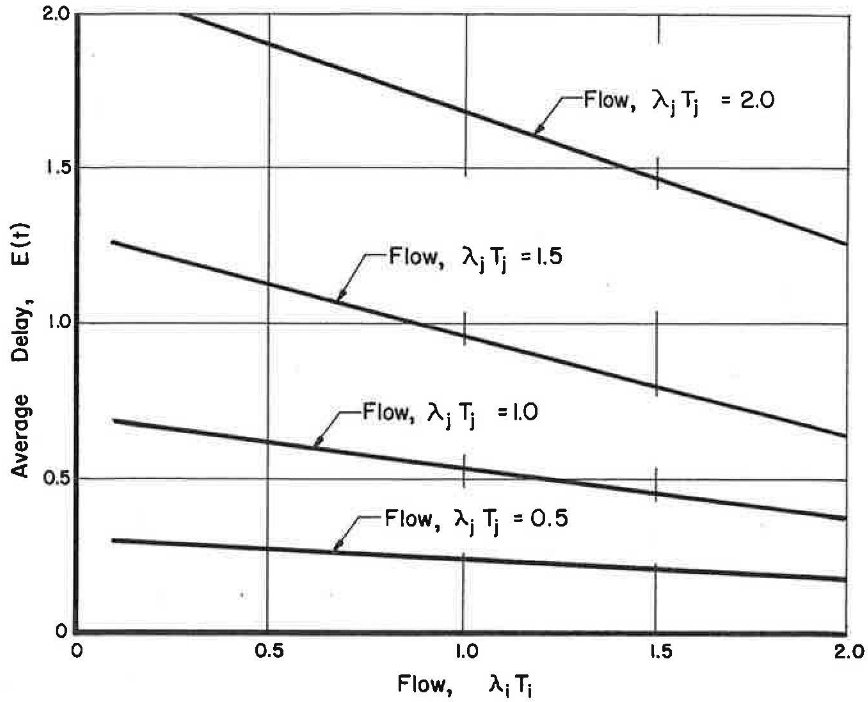


Figure 3.47. Average delays at a one-lane bottleneck for varying traffic flows and passage time through the bottleneck (see note in text).

rate μ exceeds the mean rate of arrivals λ. Such a system is in statistical equilibrium; that is, there is no continuing buildup of traffic. Solutions for this type of problem are assumed to be independent of time, but what happens when the traffic intensity exceeds unity? This question, which is of great significance in traffic engineering in describing those peak periods in which traffic demand exceeds capacity, can only be answered by considering the transient state, in which there is a continuing buildup of traffic. If the mean rate of arrivals, λ, exceeds the mean service rate, μ, the number n waiting in the system at time t, expressed by $E[n(t)]$, will grow indefinitely as t increases.

Suppose, for example, that before the peak traffic hour the highway system is in equilibrium with an initial traffic intensity ρ_0 and an expected number of vehicles in the system of $E(n)$. Now suppose that immediately the traffic intensity increases to ρ_1, such that $\rho_1 > 1$. If arrivals are Poisson distributed, the variance of the arrivals equals the mean of the arrivals. When considering that departures are also Poisson distributed, the mean number in the system at some time t after the beginning of the peak hour may be approximated by adding the expected number of arrivals and subtracting the expected number of departures to the initial number in the system as follows:

$$E[n(t)] \simeq E(n) + \lambda t - \mu t = E(n) + \mu\,(\rho_1 - 1)\,t \qquad (3.112)$$

Similar treatment of variances yields

$$\mathrm{Var}\,[n(t)] \simeq \mathrm{Var}(n) + \mathrm{Var}\,[\lambda(t)] + \mathrm{Var}\,[\mu(t)] \qquad (3.113)$$

Inasmuch as $E(n)$ and $\mathrm{Var}(n)$ are parameters of a standard distribution called the geometric distribution, and because $\mathrm{Var}\,[\lambda(t)]$ and $\mathrm{Var}\,[\mu(t)]$ are variances of the number arriving and being served in time t, which, respectively, equal λt and μt for Poisson arrivals and negative exponential service times, the number waiting in the system at time t may be expressed by

$$\mathrm{E}[n(t)] \simeq \frac{\rho_0}{1-\rho_0} + \mu\,(\rho_1 - 1)\,t \tag{3.114}$$

and their variance by

$$\mathrm{Var}\,[n(t)] \simeq \frac{\rho_0}{(1-\rho_0)^2} + \mu\,(\rho_1 + 1)\,t \tag{3.115}$$

in which ρ_0 is the initial traffic intensity ($\rho_0 = \lambda_0/\mu$) and λ_0 is the initial arrival rate.

If service times are constant rather than exponentially distributed throughout both the normal and peak periods, E(n) becomes the limiting case of the Erlang distribution and Eq. 3.114 becomes

$$\mathrm{E}\,[n(t)] \simeq 1/2\,\frac{\lambda_0^2}{\mu\,(\mu-\lambda_0)} + \frac{\lambda_0}{\mu} + \mu(\rho_1 - 1) \tag{3.116}$$

Eqs. 3.114 and 3.115 are illustrated by numerical examples. Consider a simple queue with a random arrival rate of one vehicle per minute and a mean service time of 45 sec (exponentially distributed) so that $\rho_0 = 3/4$. Now suppose that the arrival rate suddenly doubles, so that $\rho_1 = 3/2$ and that this peak period rate of traffic flow is maintained for one hour, the arrivals still being random and service times unaltered. By Eqs. 3.114 and 3.115, the mean and variance of the number in the system at the end of the hour are $\mathrm{E}\,[n(60)] = \frac{3/4}{1-3/4} + \frac{4}{3}\left(\frac{3}{2} - 1\right) 60 = 43$; $\mathrm{Var}\,[n(60)] = \frac{3/4}{(1-3/4)^2} + \frac{4}{3}\left(\frac{3}{2} + 1\right) 60 = 212.$

If the service rate μ were constant, then from Eq. 3.116 the expected number in the system becomes $\mathrm{E}\,[n(60)] = \frac{1}{2}\left(\frac{1}{4/3\,(4/3-1)}\right) + \frac{1}{4/3} + \frac{4}{3}\left(\frac{3}{2} - 1\right) 60 = 41.87 = 42.$

These equations show that the rate of growth of a queue during periods of peak demand is influenced relatively little by the assumption concerning service time distributions for traffic intensities encountered in the traffic engineering field. However, the magnitude of the variance, as compared to the mean, suggests that discretion be exercised in the application of transient state equations.

Consider now the problem of determining how long the peak hour queue takes to dissipate. Cox (*11*) made the following assumptions to solve this problem:

(a) Service time is constant.
(b) When the traffic starts to dissipate there are a large number of vehicles in the queue and the traffic intensity ρ_1 has decreased to less than one.
(c) The queue is dissipated when the queueing time of a newly arrived vehicle to the system is just equal to the average queueing time of vehicles when the system is in statistical equilibrium.

The equations developed by Cox (*11*) are

$$\mathrm{E}(T) = \left(\frac{\mathrm{E}\,[n(t)]}{\mu} - \frac{\rho_0}{2(1-\rho_0)}\right) \Big/ (1-\rho_0) \tag{3.117}$$

$$\mathrm{Var}(T) = \left(\frac{\rho_0}{\mu}\right)\left(\frac{\mathrm{E}\,[n(t)]}{\mu} - \frac{\rho_0}{2(1-\rho_0)}\right)(1-\rho_0)^{-3} + \frac{1}{\mu^2}\left([\mathrm{Var}\,n(t)]\,[1-\rho_0]^{-2}\right) \tag{3.118}$$

Referring back to the example, the mean and variance of the time it takes the queue to dissipate, as given by Eqs. 3.117 and 3.118 are $\mathrm{E}(T) = \left[\frac{43}{4/3} - \frac{3/4}{2(1-3/4)}\right] \Big/ (1-3/4) = 123$ min; $\mathrm{Var}\,(T) = \left(\frac{3/4}{4/3}\right)\left(\frac{43}{4/3} - \frac{3/4}{2(1-3/4)}\right)(1-3/4)^{-3} + \frac{1}{(4/3)^2}(212)\,(1-3/4)^{-2} = 3{,}069$ min. Thus, the effects of the rush hour last for 123 min, with a standard deviation of 55 min ($\sqrt{3{,}069}$).

3.6.4 Multiple Queues

It is possible to describe some simple traffic networks by *series* or *parallel* channels, or combinations of both. Toll booths and supermarket check-out operations are examples of parallel systems. Series arrangements include car washes and a sequence of intersections through which a vehicle must pass. Much attention has been

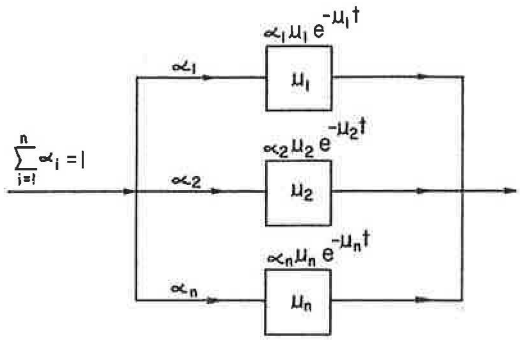


Figure 3.48. Exponential service facilities in parallel (hyper-exponential).

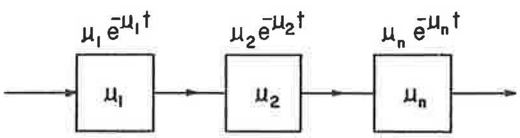


Figure 3.49. Exponential service facilities in series (Erlang, if $\mu_1 = \mu_2 = \ldots = \mu_n$).

given to networks of waiting lines because of the many industrial flow activities which can be meaningfully studied using this queueing approach.

There are two building blocks frequently used in the development of descriptions of waiting-line networks, and they are based on combinations of exponential service facilities. Parallel exponential service facilities in the general case can be represented by a hyper-exponential distribution of service times μ_i acting on some fraction of the total number of arrivals, α_i (Fig. 3.48). Multiple channels in parallel with identical exponential holding times ($\mu_i = \mu_1 = \mu_2 = \ldots = \mu_n$), first-come first-served, is a special case of the hyper-exponential.

For combinations of exponential service facilities in series (Figure 3.49), in which $\mu_i = \mu_1 = \mu_2 = \ldots = \mu_n$, a general probability distribution called the Erlang distribution, of which the negative exponential and uniform service time distributions become special cases, is a useful model for describing the queueing phenomenon. The probability density function of service times t is

$$f(t \mid \mu, K) = C_K t^{K-1} e^{-K\mu t} \quad (3.119)$$

in which

$$C_K = \frac{(\mu K)^K}{(K-1)!}$$

and K is the number of service facilities in series. It can be seen that when $K = 1$,

$$f(t \mid \mu, 1) = e^{-\mu t} \quad (3.120)$$

which is the negative exponential distribution. On the other hand, when $K = \infty$ it can be shown that the variance is zero, therefore the service time μ is constant.

The processing of vehicles at a toll station may be compared to a number of servicing channels (individual booths) arranged in parallel. The state of the system can be described in terms of the number of vehicles present, N, and the number of toll booths, M. When $N < M$ there is no queueing problem. On the other hand, when $N > M$ there is a queue of $N - M$ vehicles. The results of queueing theory approaches can be used to schedule toll station operations in order to minimize delays to customers under varying traffic flow conditions at minimum cost to the operating agency.

A comprehensive study of traffic delays at toll booths was made in New York by Edie (*13*) in 1954. The general objectives of the study were to evaluate the operating conditions existing at toll plazas and to establish methods for optimizing operations.

Edie recorded data on traffic arrivals at toll plazas, the extent of queueing in each toll lane, and the toll transaction count. From these data, the following factors involving vehicle delay were calculated:

(a) Average time required for vehicle to pass through the toll station.
(b) Average booth servicing time.
(c) Average delay or waiting time.
(d) "Delay ratio," or average delay divided by average booth servicing time.

These factors were used to establish empirical measures of delay, which when compared with appropriate queueing theories showed that the average booth servicing times at a given volume were more nearly constant than exponential in distribution. However, average servicing times showed a decrease as traffic volumes increased.

Traffic arrivals were found to be randomly distributed. For volumes below 600 vph the Poisson distribution gave a better fit to the actual data, whereas for flows above 600 vph the normal distribution gave a better approximation.

Edie was able to develop curves relating average delays to traffic volumes. From these curves (Fig. 3.50) it can be seen that the traffic carrying capacity of different toll booths for a given delay is not constant, but varies greatly between different combinations of booths.

Edie also investigated the magnitude and occurrence of maximum queues for a given set of conditions. For all traffic volumes, the distribution of queued vehicles showed a better fit to the Poisson than to the normal distribution. A relationship was established between traffic volumes and mean size of queue. An example plot of this relationship is shown in Figure 3.51. Edie showed several other similar curves for various toll booth combinations.

This study indicated that right-hand toll booths (those opposite the driver's side) were inferior to left-hand toll booths. Consequently, four Port of New York Authority toll plazas were reconstructed to eliminate the right-hand booths. A later study (*14*) indicated that this change reduced delays.

Utilizing the data on average delay per vehicle and probable maximum backup, which can be predicted for a given volume, a method for determining optimum level of service based on the principle of diminishing returns was established. Edie stated:

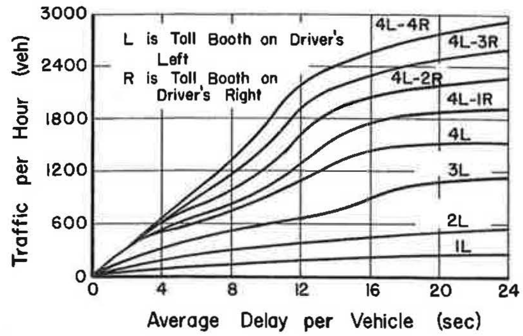


Figure 3.50. Average delay for various volumes and toll booth combinations.

"The cost is characterized by delay and the return by traffic. The point where return starts diminishing in relation to the cost is that of minimum curvature of the curves. Above this point the increases in traffic volume attained for each increment of increase of delay becomes smaller and smaller, approaching zero as the delay approaches infinity." The average delay chosen for the Port Authority's facilities was 11 sec. This average waiting delay will naturally vary with each facility.

After establishing the two criteria of average delay and maximum backup (both can be predicted when the traffic volume is known), optimum scheduling of toll station

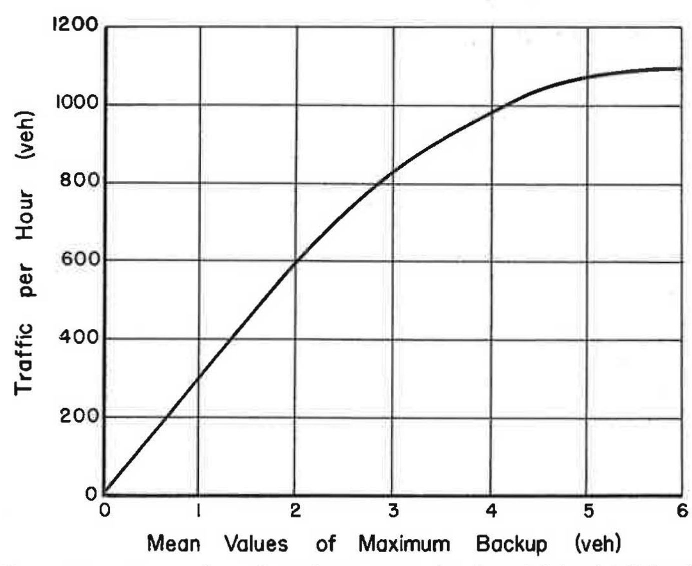


Figure 3.51. Mean values of maximum queue for three left-hand toll booths.

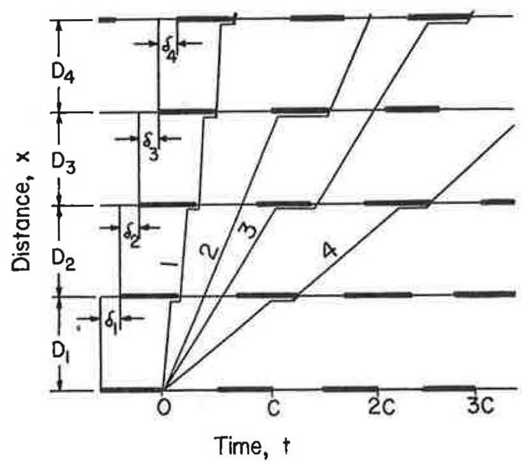


Figure 3.52. Trajectories of cars passing through a series of traffic lights (heavy bars represent red phases).

operation was accomplished. This method of scheduling by the Port of New York Authority proved to be highly satisfactory.

Newell (*55, 52*) studied the flow of highway traffic through a series of synchronized traffic signals, limiting his attention to the case where the motion of individual vehicles is independent of that of other vehicles, a situation prevailing only at low densities on wide roads. It was assumed that each vehicle has a desired speed that is maintained at all times except in the vicinity of a traffic signal. He simplified the trajectory of the vehicles and introduced a constant effective red period longer than the actual red period to account for necessary decelerations and accelerations in the vicinity of the intersection. It was indicated that these assumptions might be satisfactory for a flow as large as three vehicles stopped per cycle.

Figure 3.52 shows the trajectories of several vehicles passing through a sequence of traffic signals at various velocities. Newell limited his consideration to equally-spaced signals and a constant signal offset. With these assumptions vehicles traveling at different velocities will be stopped every block, second block, etc., depending on their speed. The most important relationship is the fraction of a cycle by which the arrival time of a vehicle with velocity u changes from one signal to the next, measured in relation to the signal offset δ, and is

$$Z = \frac{1}{C}\left[\frac{D}{u} - \delta\right] \qquad (3.121)$$

in which

Z = the fraction of the cycle change;
C = the cycle length;
D = the distance between signals; and
δ = the signal offset.

If the distance between signals is small (on the order of a city block) and if vehicle speeds are normally distributed, Newell concluded that the offset should be selected so that there is a probability p that the velocity of a vehicle is less than the usual value D/δ, where p is the fraction of the cycle which is green.

3.6.5 Parking

Queueing theory analysis can also be applied to a limited number of parking problems if the arrival, departure and queue discipline processes can be described mathematically. Those characteristics of queueing analysis dealing with length of queue and waiting times are not too meaningful because potential parkers usually leave and seek another location rather than wait in line when the parking facility (lot or curbside) is full. However, the fraction of time that the facility is full can be meaningful and useful in the planning of parking facilities and their operation.

Kometani and Kato (*42*) and Haight and Jacobson (*25*) indicated that for several curb and off-street facilities for shoppers which they studied, an assumption of random arrivals and departures fitted the observed parking behavior. For this same type of parking, Feller (*15*) has shown that the probability of K vehicles being parked in an infinitely large facility is given by the Poisson distribution:

$$\mathrm{P}(K \mid \rho) = e^{-K\rho}\frac{(K\rho)^K}{K!} \qquad (3.122)$$

However, if the lot has only N spaces, the fraction of parkers who will form a queue or be turned away is determined by the fraction of time that the parking facility is full.

If ρ is the average occupancy of the facility expressed as a fraction and N is the number of parking spaces in the facility, the probability of being full (or the fraction of time that the facility is full) is given by

$$P_L(N \mid \rho) = \frac{P(N \mid \rho)}{\sum_{j=0}^{N} P(j \mid \rho)}$$

$$= \frac{(N\rho)^N/N!}{1 + N\rho + \frac{(N\rho)^2}{2!} + \frac{(N\rho)^3}{3!} + \ldots + \frac{(N\rho)^n}{N!}} \quad (3.123)$$

The fraction of occupancy, ρ, is also equal to $\frac{\lambda E(t)}{N}$, λ being the number of vehicles arriving per hour and $E(t)$ the average parking duration.

$P_L(N \mid \rho)$, the fraction of parkers lost, is called "Erlang's loss formula," in honor of the Danish queueing pioneer, and has been used in the telephone industry for many years.

Figure 3.53 shows the fraction of parkers turned away from parking facilities with space capacities of 10, 20, 60, and 100 stalls for various fractions of occupancy, ρ.

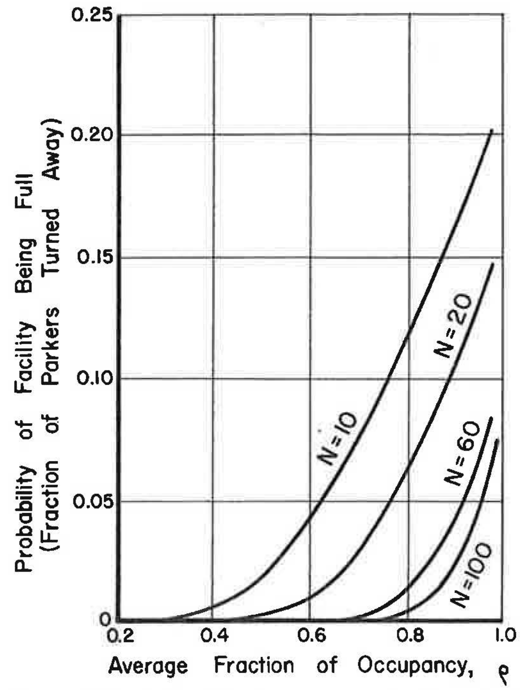


Figure 3.53. Probability of a parking facility with N spaces being full.

REFERENCES

1. Adams, W. F., "Road Traffic Considered as a Random Series," *Jour.*, Inst. of Civil Eng., 4: 121-130 (1936).
2. Bailey, N. T. J., "On Queueing Processes with Bulk Service." *Jour.*, Royal Statistical Soc., Series B, 16; 80-87 (1954).
3. Bartlett, M. S., "Some Problems Associated with Random Velocities." *Publ. de l'Inst. de Statistique de l'Univ. de Paris,* 6: 4, 261-270 (1957).
4. Beckmann, M., McGuire, C. B., and Winsten, C. B., *Studies in the Economics of Transportation.* Yale Univ. Press (1956).
5. Borel, E., "Sur l'Emploi du Theoreme de Bernoulli pour Faciliter le Calcul d'un Infinite de Coefficients." *Comptes Rendus,* Acad. des Sciences, Paris, 214: 452-456 (1942).
6. Carleson, L., "En Matematisk Modell fór Landsvágstrafik." *Nordisk Matematisk Tidskrift,* 5: 176-180 (with English summary) (1957).
7. Castoldi, L., "Queue Alternance and Traffic Flow at a Crossword." *Bollettino del Centro per la Ricerca Operativa,* No. 5-6, pp. 1-13 (1957).
8. Clarke, A. B., "A Waiting Line Process in Markov Type." *Annals of Math. Statistics,* 27: 452-459 (1956).
9. Clayton, A. J. H., "Road Traffic Calculations." *Jour.*, Inst. of Civil Eng., 16: 247 (1941).
10. Cohen, J., Dearnaley, E. J., and Hansel, C. E. M., "The Risk Taken in Crossing a Road." *Oper. Res. Quart.*, 6: 3, 120-128 (Sept. 1955).
11. Cox, D. R., "The Statistical Analysis of Congestion." *Jour.*, Royal Statistical Soc., Series A, 118: 324-335 (1955).
12. Crawford, A., "The Overtaking Driver." British Road Res. Lab., Research Note RN/3771/AC (June 1960).
13. Edie, L. C., "Traffic Delays at Toll Booths." *Oper. Res.,* 2: 2, 107-138 (1954).
14. Edie, L. C., "Review of Port of New York Authority Study." *Oper. Res.,* 8: 2, 263 (Mar. 1960).
15. Feller, W., *An Introduction to Probability Theory and its Applications.* Vol. 1, Wiley and Sons, New York (1950).
16. Garwood, F., "The Application of the

Theory of Probability to the Operation of Vehicular Controlled Traffic Signals." *Jour.*, Royal Statistical Soc., Supplement, 7: 65-77 (1940).
17. Gerlough, D. L., "Use of the Poisson Distribution in Highway Traffic." *Poisson and Traffic,* The Eno Foundation for Highway Traffic Control, pp. 1-58 (1955).
18. Glickstein, A., Findley, L. D., and Levy, S. L., "Application of Computer Simulation Techniques to Interchange Design Problems." HRB Bull. 291, pp. 139-162 (1962).
19. Goodman, L. A., "On the Poisson-Gamma Distribution Problem." *Annals,* Inst. of Statistical Math., Tokyo, 3: 123-125 (1952).
20. Greenshields, B. D., "Distance and Time Required for Passing Vehicles." *HRB Proc.,* 15: 332-342 (1935).
21. Haight, F. A., and Breuer, M. A., "The Borel-Tanner Distribution." *Biometrika,* 47: Pts. 1 and 2, 143-150 (1960).
22. Haight, F. A., "The Generalized Poisson Distribution." *Annals,* Inst. of Statistical Math., Tokyo, 11: 2, 101-105 (1959).
23. Haight, F. A., and Mosher, W. W., Jr., "A Practical Method for Improving the Accuracy of Vehicular Speed Distribution Measurements." HRB Bull. 341, pp. 92-96 (1962).
24. Haight, F. A., Bisbee, E. F., and Wojcik, C., "Some Mathematical Aspects of the Problem of Merging." HRB Bull. 356, pp. 1-14 (1962).
25. Haight, F. A., Jacobson, A. S., "Some Mathematical Aspects of the Parking Problem." *HRB Proc.,* 41: 363-374 (1962).
26. Haight, F. A., "Overflow at a Traffic Light." *Biometrika,* 46: Pts. 3 and 4, 420-424 (1959).
27. Haight, F. A., *Mathematical Models for Traffic Flow.* Academic Press, New York (1963).
28. Haight, F. A., "Towards a Unified Theory of Road Traffic." *Oper. Res.,* 6: 6, 813-826 (1958).
29. Haight, F. A., Whisler, B. F., and Mosher, W. W., Jr., "New Statistical Method for Describing Highway Distribution of Cars." *HRB Proc.,* 40: 557-564 (1961).
30. Herman, R., and Weiss, G., "Comments on the Highway-Crossing Problem." *Oper. Res.,* 9: 6, 828-840 (Nov.-Dec. 1961).
31. "Highway Capacity Manual." Bureau of Public Roads, U. S. Government Printing Office, Washington, D. C. (1950).
32. Ho, E. C., "Statistical Analysis of Congested Merging Traffic." In *Analysis and Simulation of Vehicular Traffic Flow,* Res. Rpt. No. 20, Inst. of Transp. and Traffic Eng., Univ. of California (1954).
33. Hunt, G. C., "Sequential Arrays of Waiting Lines." *Oper. Res.,* 4: 6, 674-683 (Dec. 1956).
34. Jackson, J. R., "Networks of Waiting Lines." *Oper. Res.,* 5: 4, 518-521 (Aug. 1957).
35. Jensen, A., "Traffic Theory as an Aid in the Planning and Operation of the Road Grid." *Ingeniøren,* 2: 1, 48-61 (Oct. 1957).
36. Jewell, W. S., "The Properties of Recurrent-Event Processes." *Oper. Res.,* 8: 4, 446-472 (July 1960).
37. Jewell, W. S., *Waiting for a Gap in Traffic.* Res. Rpt. No. 6, Operations Res. Center, Univ. of California (June 1961).
38. Kell, J. H., "A Theory of Traffic Flow on Urban Streets." Proc. of 13th Ann. Meeting, Western Sect., Inst. of Traffic Eng. (1960).
39. Kell, J. H., "Analyzing Vehicular Delay at Intersections Through Simulation." HRB Bull. 356, pp. 28-39 (1962).
40. Kendall, D. G., "Some Problems in the Theory of Queues." *Jour.,* Royal Statistical Soc., Series B, 13: 151-185 (1951).
41. Kometani, E., "On the Theoretical Solution of Highway Traffic Under Mixed Traffic." *Memoirs, Faculty of Engineering, Kyoto Univ.,* 17: 79-88 (1955).
42. Kometani, E., and Kato, A., "On the Theoretical Capacity of an Off-Street Parking Space." *Memoirs, Faculty of Engineering, Kyoto Univ.,* 18: 4, 315-328 (1956).

43. Little, J. D. C., "Approximate Expected Delays for Several Maneuvers by a Driver in Poisson Traffic." *Oper. Res.*, 9: 1, 39-52 (1961).
44. Mayne, A. J., "Some Further Results in the Theory of Pedestrians and Road Traffic." *Biometrika*, 41: 375-389 (1954).
45. Miller, A. J., "Traffic Flow Treated as a Stochastic Process." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 165-174 (1961).
46. Miller, A. J., "A Queueing Model for Road Traffic Flow." *Jour.*, Royal Statistical Soc., Series B, No. 23, pp. 64-75 (1961).
47. Miller, A. J., "Analysis of Bunching in Rural Two-Lane Traffic." *Oper. Res.*, 11: 2, 236-247 (Mar.-Apr. 1963).
48. Mori, M., "Traffic Characteristics of Roads Under Mixed Traffic Conditions." *Traffic Eng.*, 30: 1, 23-28 (Oct. 1959).
49. Mori, M., "On the Road Constant Indicating the Passing Characteristics of Roads Under Mixed Traffic Conditions." *Memoirs, Faculty of Engineering, Osaka Univ.*, 2: 9-18 (1960).
50. Morse, P. M., *Queues, Inventories and Maintenance*, Wiley and Sons, New York (1958).
51. Moskowitz, K., "Waiting for a Gap in a Traffic Stream." *HRB Proc.*, 33: 385-395 (1954).
52. Newell, G. F., "The Flow of Highway Traffic Through a Sequence of Synchronized Traffic Signals." *Oper. Res.*, 8: 3, 390-405 (1960).
53. Newell, G. F., "Queues for a Fixed Cycle Traffic Light." *Annals of Math. Statistics*, 31: 3, 589-597 (1960).
54. Newell, G. F., "The Effect of Left Turns on the Capacity of a Traffic Intersection." *Quart. Appl. Math.*, 17: 67-76 (1959).
55. Newell, G. F., "Statistical Analysis of the Flow of Highway Traffic Through a Signalized Intersection." *Quart. Appl. Math.*, 13: 353-369 (1956).
56. Newell, G. F., "Mathematical Models for Freely-Flowing Highway Traffic." *Oper. Res.*, 3: 2, 176-186 (May 1955).
57. Oliver, R. M., *Zero-One Bunching and a Passing Rule*. Res. Rpt. 16, Oper. Res. Center, Univ. of California, Berkeley (July 1961).
58. Oliver, R. M., "A Traffic Counting Distribution." *Oper. Res.*, 9: 6, 802-810 (Nov.-Dec. 1961).
59. Oliver, R. M., "On High-Speed Two-Lane Traffic Merges." Unpublished.
60. Oliver, R. M., "Distribution of Gaps and Blocks in a Traffic Stream." *Oper. Res.*, 10: No. 2 (Mar.-Apr. 1962).
61. Oliver, R. M., and Thibault, B., "A High Flow Traffic Counting Distribution." HRB Bull. 356, pp. 15-27 (1962).
62. Oliver, R. M., and Bisbee, E. F., "Queueing for Gaps in High Flow Traffic." *Oper. Res.*, 10: 1, 105-114 (Jan.-Feb. 1962).
63. Oliver, R. M., and Jewell, W. S., *The Distribution of Spread*. Res. Rpt. No. 20, Oper. Res. Center, Univ. of California, Berkeley (Dec. 1961).
64. Raff, M. S., "The Distribution of Blocks in an Uncongested Stream of Traffic." *Jour.*, Amer. Statistical Assoc., 46: 114-123 (1951).
65. Saaty, T. L., "Resume of Useful Formulas in Queueing Theory." *Oper. Res.*, 5: 2, 161-200 (1957).
66. Schuhl, A., "The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highway." *Poisson and Traffic*, The Eno Foundation for Highway Traffic Control (1955).
67. Tanner, J. C., "The Delay to Pedestrians Crossing a Road." *Biometrika*, 38: 383-392 (1951).
68. Tanner, J. C., "A Problem of Interference Between Two Queues." *Biometrika*, 40: 58-69 (1953).
69. Tanner, J. C., "A Simplified Model for Delays in Overtaking on a Two-Lane Road." *Jour.*, Royal Statistical Soc., Series B, 20: 2, 408-414 (1958).
70. Tanner, J. C., "Delays on a Two-Lane Road." *Jour.*, Royal Statistical Soc., Series B, 23: 1, 38-63 (1961).
71. Tanner, J. C., "A Theoretical Analysis of Delays at an Uncontrolled Intersection." *Biometrika*, 49: 1 and 2, 163-170 (June 1952).
72. Tanner, J. C., "A Derivation of the Borel Distribution." *Biometrika*, 48: 1 and 2, 222-224 (June 1961).
73. Uematu, T., "On Traffic Control at an

Intersection Controlled by the Repeated Fixed Cycle Traffic Light." *Ann.*, Inst. Statistical Math., Tokyo, 9: 87-107 (1958).

74. Weiss, G. H., "The Pedestrian Queueing Problem." Proc. 33rd Session, Internat. Statistical Inst., Paris (1961).

75. Weiss, G. H., and Maradudin, A. A., "Some Problems in Traffic Delay." *Oper. Res.*, 10: 1, 74-104 (Jan.-Feb. 1962).

76. Whittlesey, J. R. B., and Haight, F. A., "Counting Distributions for Erlang Processes." *Ann.*, Inst. Statistical Math., Tokyo, 13: 2, 91-103 (1961).