HIGHWAY RESEARCH BOARD
Special Report 79

# AN INTRODUCTION TO
# Traffic Flow Theory

Edited by
D. L. Gerlough
and
D. G. Capelle

HIGHWAY RESEARCH BOARD
of the
Division of Engineering and Industrial Research
National Academy of Sciences—
National Research Council
Washington, D. C.
1964

NATIONAL ACADEMY OF SCIENCES—NATIONAL RESEARCH COUNCIL

## Special Committee on Publication of Selected Information on Theory of Traffic Flow

*Representing the Highway Research Board, An Agency of the Division of Engineering and Industrial Research*

D. E. Cleveland
R. S. Foote
D. L. Gerlough, *Chairman*
C. C. Robinson

*Representing the Division of Mathematics*

H. P. Galliher
E. W. Montroll
G. F. Newell
R. W. Rothery

This publication was prepared by the Special Committee on Publication of Selected Information on Theory of Traffic Flow. The Special Committee is composed of representatives of the Division of Mathematics and the Highway Research Board, which is an agency of the Division of Engineering and Industrial Research. Both divisions operate under the auspices of the National Academy of Sciences—National Research Council.

The National Academy of Sciences—National Research Council gratefully acknowledge the work of the Special Committee and express appreciation for the cooperation of the authors who either wrote original manuscripts or granted permission for the reprinting of their papers. Appreciation is also extended to the Automotive Safety Foundation for financial assistance in printing this publication.

# Preface

What is "traffic flow theory" and why is it of interest to highway and traffic engineers?

A theory is a set of scientifically acceptable principles that explain phenomena—in this case, the phenomena of vehicular traffic flow. Traffic flow theory—like all theories—has developed as practitioners and theorists contributed their findings to the general fund of knowledge.

The earliest contributors were practicing engineers who measured the performance of motor vehicles on the highways and used these basic field data in their search for an understanding of the characteristics of traffic flow. These researchers, many of them still pre-eminent in their profession, supplied the vast resource of data and knowledge that has enabled traffic engineering to keep pace with rapid advances in motor vehicle transportation.

In recent years, other investigators from widely varying disciplines have contributed immeasurably to the understanding of traffic flow. These investigators are scientists—predominantly physicists, mathematicians and psychologists—who are totally removed from the day-to-day problems of the traffic engineer. Their orientation, born of an academic interest in traffic engineering, is directed toward the understanding of relatively narrow problems in order that comparisons can be made between the experimental approach of the traffic engineer and the theoretical approach of the scientist. Contributions of scientists, however, are frequently published in journals which do not circulate among highway and traffic engineers: *Biometrika, Operational Research Quarterly,* and *Quarterly of Applied Mathematics,* for example.

In these publications authors frequently use terminology and symbols which are neither familiar to the engineer nor consistent among various descriptions of the same phenomenon. Thus, the diligent reader seeking an understanding of these theories is faced with a formidable task of reconciling differences in basic vocabularies. This publication, therefore, is an attempt to consolidate recent contributions to traffic flow theory and present them in a related manner with a consistent set of notations.

It is hoped that the publication will encourage further testing and validation of the theories presented. The theorist is frequently unable to measure traffic flow or analyze the mass of data necessary to validate, repudiate or refine his theories. Although tested, many of the theoretical descriptions presented have not been completely validated. Additional verification and refinement are required before the theories can become useful analytic tools.

This cannot be accomplished in the laboratory. It is possible only through the effort and interest of highway officials who have access to the final proving grounds for all traffic flow theory—the operating highway facility.

*Carlton Robinson,*
Automotive Safety Foundation

# Contents

# DEFINITIONS AND NOTATIONS

Symbols and terminology familiar to the highway profession are used throughout this publication wherever possible. The Special Committee found it necessary, however, to adopt some new symbols and definitions in order to standardize terminology in use among highway and traffic engineers. The following terms and symbols are considered the most practical from the standpoint of clarity and general acceptance:

| *Symbol* | *Term* | *Definition* |
|---|---|---|
| $a$ | Acceleration | The time rate of change of speed, $d^2x/dt^2$. |
| $a_n$ | | The acceleration of the $n$th vehicle. |
| $c$ | Wave speed | The speed at which a wave of differing concentrations is propagated in the traffic stream. |
| $D$ | Density | See "concentration" ($k$). |
| $h$ | Time headway | The time interval between passages of consecutive vehicles moving in the same lane (measured between corresponding points on the vehicles). |
| $h_n$ | | Headway between the ($n$–1)st vehicle and the $n$th vehicle. |
| $H_n$ | Total headway time | The time interval between passages of the first and the $n$th vehicle moving in the same lane. |
| $k$ | Concentration | The number of vehicles occupying a unit length of a lane at a given instant; often referred to as "density" when expressed in vehicles per mile. |
| $k_j$ | Jam concentration | The maximum concentration of vehicles when jammed at a stop. |
| $k_m$ | Optimum concentration | The concentration when flow is at a maximum rate. |
| $L$ | Car length | The length of a vehicle. |
| $n$ | | The vehicle number. |
| $N$ | | Total number of vehicles. |
| $\mathrm{p}(x)$, $\mathrm{P}(x)$ | Probability | The likelihood of occurrence of an event. |
| $q$ | Flow | The number of vehicles passing a point during a specified period of time; often referred to as "volume" when expressed in vehicles per hour measured over an hour. |

| *Symbol* | *Term* | *Definition* |
|---|---|---|
| $q_m$ | Maximum flow | The maximum attainable flow. |
| $r$ | Correlation coefficient | A statistical measure of the association between data and a regression line. |
| $s$ | Spacing | The distance between consecutive vehicles moving in the same lane (measured between corresponding points on the vehicles). |
| $s_n$ | Spacing | The spacing between the $(n-1)$st vehicle and the $n$th vehicle. |
| $t$ | Time | An interval or index of time. |
| $T$ | Time | Total time. |
| $u$ | Speed | The time rate of change of distance, $dx/dt$. |
| $u_m$ | Optimum speed | The speed when flow is at a maximum rate. |
| $\bar{u}_s$ | Space-mean speed | The arithmetic mean of the speeds of the vehicles occupying a given length of lane at a given instant. |
| $\bar{u}_t$ | Time-mean speed | The arithmetic mean of the speeds of vehicles passing a point during a given interval of time. |
| $V$ | Volume | See "flow" ($q$). |
| $x, y, z, X, Y, Z$ | Position | An index of position. |
| $\Delta$ | Increment | |
| $\eta$ | Normalized concentration | The ratio $k/k_j$. |
| $\sigma$ | Standard deviation | A statistical measure of the dispersion of data from the mean. |
| $\dot{x}$ | First derivative (speed) | The differentiation of $x$ with respect to some independent variable; *i.e.*, $dx/dt$. |
| $\ddot{x}$ | Second derivative (acceleration) | The second differentiation of $x$ with respect to some independent variable; *i.e.*, $d^2x/dt^2$. |
| Var $(\cdot)$ | | Statistical variance. |
| E $(\cdot)$ | | Expected or mean value. |
| exp $(x - y)$ | | $e^{x-y}$. |
| p$(x \mid a, b)$ | | Probability of $x$ given conditions a and b. |

# Chapter 1

# HYDRODYNAMIC APPROACHES

Part I

L. A. Pipes
University of California, Los Angeles

Part II

M. J. Lighthill
Royal Aircraft Establishment

and

G. B. Whitham
California Institute of Technology

## Contents

# Chapter 1

# HYDRODYNAMIC APPROACHES

## PART I

### 1.1 INTRODUCTION

In recent years, numerous mathematical theories of traffic flow applicable to long crowded roads have been developed. Although many of these theories involve a statistical approach, several are described in terms of fluid or hydrodynamic flows. The latter regard traffic as a compressible fluid having a certain density or concentration and a certain fluid velocity. Their analyses are based on a partial differential equation expressing the conservation of matter and an assumed empirical relation between the flow and the concentration. These analyses can be adjusted to fit flow-concentration curves of particular highways. The solution of the equation indicates that discontinuities in traffic flow are propagated in a manner similar to "shock waves" in the theory of compressible fluids. It is, therefore, the purpose of this chapter to discuss the application of fluid flow principles to the traffic stream.

### 1.2 FUNDAMENTAL CONCEPTS

Lighthill and Whitham prepared an outstanding paper on the theory of traffic flow in which they discussed the behavior of shock waves in the traffic stream and developed a theory of the propagation of changes in traffic distribution. Part I is an elementary approach to this theory.

Consider the movement of two distinct concentrations of traffic $k_1$ and $k_2$ along a straight highway (Fig. 1.1). The two concentrations $k_1$ and $k_2$ are separated by the vertical line S, which has a velocity of $c$. This velocity is considered positive if the line moves in the direction of positive $x$ as depicted by the arrow. With the following notations:

$u_1$ = Mean speed of vehicles in region A;
$u_2$ = Mean speed of vehicles in region B;
$U_{r1} = (u_1 - c)$ = Relative speed of vehicles in region A to the moving line S; and
$U_{r2} = (u_2 - c)$ = Relative speed of vehicles in region B to the moving line S,

it is evident that in time $t$ the number of vehicles $N$ crossing the dividing line S is

$$N = U_{r1}\, k_1\, t = U_{r2}\, k_2\, t \qquad (1.1)$$

or

$$(u_1 - c)\, k_1 = (u_2 - c)\, k_2 \qquad (1.2)$$

This equation is a statement of the conservation of matter applied to the vehicles that cross the line S and may be written in the form

$$u_1 k_1 - u_2 k_2 = c\,(k_1 - k_2) \qquad (1.3)$$

If the rate of traffic flow in region A is $q_1$, and the rate of traffic flow in region B is $q_2$,

$$q_1 = k_1 u_1 \qquad (1.4)$$

and

$$q_2 = k_2 u_2 \qquad (1.5)$$

These relations follow from the definition of the quantities involved. In terms of the rates of flow $q_1$ and $q_2$, Eq. 1.3 becomes

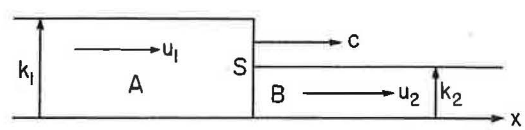



Figure 1.1. Movement of two concentrations.

$$c = (q_2 - q_1)/(k_2 - k_1) \qquad (1.6)$$

If the rates of flow and the concentrations are nearly equal,

$$(q - q_1) = \Delta q\ ,\ (k - k_1) = \Delta k \qquad (1.7)$$

and Eq. 1.6 becomes

$$c = \Delta q/\Delta k = dq/dk \qquad (1.8)$$

which is the equation for the velocity $c$ with which small *disturbances* in the traffic stream are propagated.

In the general case in which the differences in the concentrations on the two sides of the moving line S are not infinitesimally small, Eq. 1.3 may be written in the form

$$c = (u_1\,k_1 - u_2\,k_2)/(k_1 - k_2) \qquad (1.9)$$

So far, the elementary analysis has not considered any relation between the mean velocities $u_1$ and $u_2$ and the concentrations $k_1$ and $k_2$. Greenshields (*1*) found in his study of traffic capacity that

$$u_1 = \bar{u}_s(1-\eta_1) \text{ and } u_2 = \bar{u}_s(1-\eta_2) \qquad (1.10)$$

in which $\bar{u}_s$ is the space-mean speed of the traffic stream, and $\eta_1$ and $\eta_2$ are the normalized concentrations on both sides of the boundary line S. Substituting these values in Eq. 1.9 gives a wave speed of

$$c = \frac{[k_1\,\bar{u}_s\,(1-\eta_1) - k_2\,\bar{u}\,(1-\eta_2)]}{(k_1 - k_2)} \qquad (1.11)$$

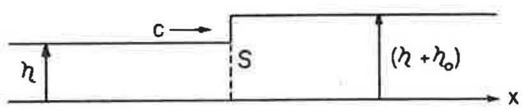



Figure 1.2. Small discontinuity in concentration.

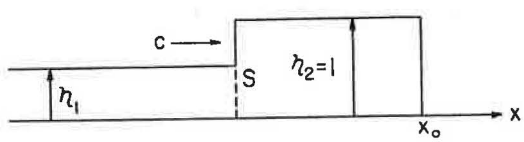



Figure 1.3. Shock wave caused by stopping.

The normalized concentrations $\eta_1$ and $\eta_2$ are given by

$$\eta_1 = k_1/k_j,\ \eta_2 = k_2/k_j \qquad (1.12)$$

in which $k_j$, the jam concentration, is the maximum concentration of vehicles when jammed at a stop. Both $k_1$ and $k_2$ may be eliminated from Eq. 1.11, the resulting wave speed being

$$c = \bar{u}_s\,[1 - (\eta_1 + \eta_2)] \qquad (1.13)$$

which gives the velocity of the line S in terms of the normalized concentrations on either side of the moving discontinuity.

### 1.2.1 The Case of Nearly Equal Concentrations

If the normalized concentrations $\eta_1$ and $\eta_2$ on both sides of the boundary line S are nearly equal, the situation shown in Figure 1.2 exists. The normalized concentration to the left of S is $\eta$, whereas the normalized concentration to the right of S is $(\eta + \eta_0)$, where $\eta + \eta_0 \leq 1$. In this case, let

$$\eta_1 = \eta\ ,\ \eta_2 = [\eta + \eta_0] \qquad (1.14)$$

and

$$[1 - (\eta_1 + \eta_2)] = [1 - (2\eta + \eta_0)] = [1 - 2\eta] \qquad (1.15)$$

in which $\eta_0$ is neglected. If Eq. 1.13 is substituted in Eq. 1.15, the wave of discontinuity is propagated with a velocity of

$$c = \bar{u}_s[1 - 2\eta] \qquad (1.16)$$

This is the equation for the propagation of shock waves obtained by Lighthill and Whitham by a more elaborate analysis.

### 1.2.2 Waves of Stopping

Consider a line of traffic moving with a normalized concentration $\eta_1$ and a mean vehicle velocity of

$$u_1 = \bar{u}_s[1 - \eta_1] \qquad (1.17)$$

At a position $x = x_0$ on the highway, a traffic signal causes the traffic to halt, and the stream immediately assumes a saturated normalized concentration of $\eta_2 = 1$, as shown in Figure 1.3. To the left of the line S, the traffic is still moving with a mean velocity given by Eq. 1.17 at the original concentration of $\eta_1$. Under these conditions, the shock wave velocity is given by substituting $\eta_1=\eta_1$ and $\eta_2=1$ in Eq. 1.13 to give

$$c=\bar{u}_s[1-(\eta_1+1)]=-\bar{u}_s\eta_1 \qquad (1.18)$$

which indicates that the shock wave of stopping travels backward with a velocity of $\bar{u}_s\eta_1$. If the signal at $x=x_0$ turns red at $t=0$, then in time $t$ later, a line of cars of length $\bar{u}_s\,\eta_1 t$ will be stopped behind $x_0$.

### 1.2.3 Waves of Starting

In order to discuss the nature of the shock wave produced by the starting of a line of vehicles, assume that at $t=0$ a line of vehicles has accumulated behind a signal located at $x=x_0$. Because this line of vehicles is standing still, it has a saturated concentration of $\eta_1=1$, as shown in Figure 1.4. Assume that at $t=0$ the signal at $x=x_0$ turns green and permits vehicles to move forward with a velocity of $u_2$. Because $u_2=\bar{u}_s[1-\eta_2]$ there exists a concentration of

$$\eta_2=[1-(u_2/\bar{u}_s)] \qquad (1.19)$$

Therefore, a shock wave of starting forms as soon as the line of vehicles begins to move. The velocity of this shock wave is obtained by substituting $\eta_1=1$ and $\eta_2=\eta_2$ in Eq. 1.13, thus

$$c=\bar{u}_s[1-(1+\eta_2)]=-\bar{u}_s\,\eta_2=-(\bar{u}_s-u_2) \qquad (1.20)$$

Therefore, the shock wave of starting travels backward from $x_0$ with a velocity of $(\bar{u}_s-u_2)$. Because the starting velocity is small, it is seen that the shock wave of starting travels backward with a velocity essentially equal to $-\bar{u}_s$.

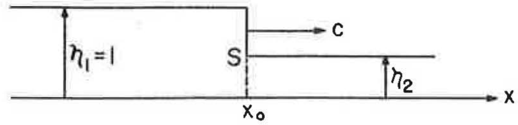



Figure 1.4. Shock wave caused by starting.

## 1.3 COMPARISON OF LIGHTHILL-WHITHAM AND RICHARDS THEORIES

Richards (*2*) prepared a paper on the theory of traffic shock waves, covering the same material as Lighthill and Whitham, at about the same time and without knowledge of their work.

Essentially these two theories are identical. Lighthill and Whitham center their attention on the discontinuities in the rate of flow $q$, whereas Richards centers his attention on the discontinuities in the concentration $k$, which he calls the density function $D$. In both theories the fundamental equation is the one that expresses the conservation of matter. However, because Lighthill and Whitham do not restrict themselves to any definite flow-concentration curve, their analysis is somewhat more general than that of Richards.

Richards incorporates in his basic equations the straightline relation $u=\bar{u}_s\,(1-\eta)$ for the mean velocity of the vehicles. Therefore, the conclusions reached by Richards are limited to situations in which this law of velocity and concentration hold. If this hypothesis is incorporated into the Lighthill-Whitham theory, their theory is identical with that of Richards. The difference between the two theories is then seen to be only one of notation and graphical interpretation.

## PART II

Part II (opposite page) is reprinted with permission of the Royal Society and the authors. The original paper appeared in *Proceedings of the Royal Society*, A229, No. 1178, 1955, 317-345. The style and notation of the original paper does not conform with that established for this publication. At the authors' insistence, however, the notation and the figure and equation designations have been left exactly as originally printed. Therefore, the following list of comparative notations is provided:

| *Lighthill & Whitham* | *Established Standard* |
|---|---|
| $v = q/k$ = Mean speed of traffic. | $\bar{u}_s$ = Space mean speed. |
| $n$ = Total number of vehicles. | $N$ = Total number of vehicles. |
| $U$ = Uniform speed. | $u$ = Time rate of change of distance. |
| $v_f$ = Free mean speed. | $\bar{u}_f$ = Free mean speed. |

All other notations are as used elsewhere throughout the book and given on pages *vii* and *viii*.

PART II

# On kinematic waves:

# II. A theory of traffic flow on long crowded roads

By M. J. Lighthill, F.R.S. and G. B. Whitham

(*Department of Mathematics, University of Manchester*)

(*Received* 15 *November* 1954—*Read* 17 *March* 1955)

This paper uses the method of kinematic waves, developed in part I, but may be read independently. A functional relationship between flow and concentration for traffic on crowded arterial roads has been postulated for some time, and has experimental backing (§2). From this a theory of the propagation of changes in traffic distribution along these roads may be deduced (§§2, 3). The theory is applied (§4) to the problem of estimating how a 'hump', or region of increased concentration, will move along a crowded main road. It is suggested that it will move slightly slower than the mean vehicle speed, and that vehicles passing through it will have to reduce speed rather suddenly (at a 'shock wave') on entering it, but can increase speed again only very gradually as they leave it. The hump gradually spreads out along the road, and the time scale of this process is estimated. The behaviour of such a hump on entering a bottleneck, which is too narrow to admit the increased flow, is studied (§5), and methods are obtained for estimating the extent and duration of the resulting hold-up.

The theory is applicable principally to traffic behaviour over a long stretch of road, but the paper concludes (§6) with a discussion of its relevance to problems of flow near junctions, including a discussion of the starting flow at a controlled junction.

In the introductory sections 1 and 2, we have included some elementary material on the quantitative study of traffic flow for the benefit of scientific readers unfamiliar with the subject.

## 1. Introduction

A new problem, which has arisen in the twentieth century, is how to organize road traffic so that the full benefits of our increased mobility can be enjoyed at the lowest cost in human life and capital. The problem has many sides—constructional, legal, educational, administrative. The early lines of attack were largely intuitive. But, more recently, there has been an increasing tendency to adopt scientific methods, and try to assess the relative merits of different lines of attack by means of controlled experiments. This has been done both by the various authorities responsible for road lay-out, administration and propaganda, and also, more comprehensively, by organizations like the Road Research Laboratory in Great Britain, and the Bureau of Public Roads (formerly the Public Roads Administration) in the U.S.A. (Glanville 1953; Smeed 1952).

An important branch of the subject, with repercussions on all the other branches, is the quantitative study of traffic flow. An account of the experimental methods employed in this field has been given by the head of the traffic-flow section at the Road Research Laboratory (Charlesworth 1950). They include methods for measuring the means and standard deviations of vehicle speed at a point or journey time over a stretch of road, and for measuring the flow (number of vehicles passing a given point per unit of time). Attempts to correlate these variables for roads of particular mean width, mean curvature, etc., are made. Also, traffic performance is studied before and after some change in road conditions, and statistical technique is used to find out whether the change significantly reduces journey times or accidents. Extensive researches on similar lines are carried out in the U.S.A., notably by the Division of Highway Transport Research, and by certain university departments such as the Post-graduate School of Highway Engineering at Yale.

In contrast to the well-developed character of traffic flow as an experimental science, theoretical approaches to the subject are in their infancy. Wardrop (1952) has given a valuable account of such theoretical investigations as have been made. He emphasizes the need for theoretical ideas to be used in conjunction with experimental data and the experience of individuals. It is well known, of course, in all branches of science and technology, that judicious use of theoretical ideas can save a lot of time by suggesting how experimental results obtained under one set of conditions can be extrapolated to another set of conditions. For example, theory may suggest in what form a set of results should be graphed, to give a curve likely to vary as little as possible with change of conditions. It may also suggest what things can most usefully be measured.

The theories which Wardrop (1952) describes are, as might be expected, statistical. First, the kinds of mean values which can be taken are discussed—for example, a 'space mean' over a length of road, or a 'time mean' over an interval of time at a fixed point. The space-mean speed (which we use in this paper) is the length of road divided by the average journey time of vehicles traversing it. It is also the ratio of the flow (vehicles per hour) to the concentration (vehicles per mile). The time-mean speed is somewhat greater because fast vehicles pass a fixed point more frequently (relative to their distribution in space) than slow vehicles.

Wardrop discusses the effect of increase of flow on overtaking. The number of 'desired overtakings' might be expected to increase as the square of the flow, so evidently, beyond a certain value of the flow, the proportion of desired overtakings which are possible must decrease. (For detailed observations on this point, see Norman, 1942.) This would clearly cause a reduction of mean speed with increase of flow, which is observed. He discusses also how traffic with uniform origin and destination may be expected to distribute itself over alternative routes, and he gives useful applications of the 'theory of queues' to the problem of delay at traffic lights (see also Tanner 1953).

In this paper we introduce a quite different method, suggested by theories of the flow about supersonic projectiles and of flood movement in rivers. It is the method of kinematic waves, introduced in part I (Lighthill & Whitham 1955); however, it is not essential to have read part I to understand the account which follows.

Now, a theoretical approach to road-traffic problems using methods from fluid dynamics is limited in advance to a restricted range of problems. Other ranges undoubtedly require statistical treatment of the kind described above, based on the theory of queues or the general theory of 'stochastic processes' (random time series). The 'continuous-flow' approach represents the limiting behaviour of a stochastic process for a large 'population' (total number of vehicles), and is therefore applicable to large-scale problems only—principally to the distribution of traffic along long, crowded roads.

This 'arterial road' problem is an important one, however, which would be almost impossible to treat by purely statistical methods (though it may later be found desirable to use the present approach only as a first approximation, passing to higher approximations by means of a suitable blend with statistical ideas). To illustrate the theory, we use it to predict (§4) the progress of a traffic 'hump' in a long main road (due to a period of increased inflow at the main feed point), and (§5) the extent of the hold-up which results when such a hump passes through a bottleneck, which is too narrow to admit the increased flow. We also apply the method (§6) to junctions, especially controlled junctions, on long main roads.

The fundamental hypothesis of the theory is that at any point of the road the flow $q$ (vehicles per hour) is a function of the concentration $k$ (vehicles per mile). The evidence for this is discussed at length in §2. The hypothesis implies, as was shown in part I, that slight changes in flow are propagated back through the stream of vehicles along 'kinematic waves', whose velocity relative to the road is the slope of the graph of flow against concentration. A driver experiences such a wave whenever he adjusts his speed in accordance with the behaviour of the car or cars in front of him—for example, on observing a brake light, or an opportunity to overtake. It was seen also in part I that kinematic waves can run together to form 'kinematic shock waves', at which fairly large reductions in velocity occur very quickly. These too are very common on roads, notably at the rear of a traffic 'hump', and behind a bottleneck.

The properties of kinematic shock waves, and of continuous kinematic waves, will be derived again, by purely descriptive arguments, in §2. The more mathematical derivation, which some readers may prefer, will be found in §1 of part I.

The later sections are devoted to examples of the kinds already mentioned. The predictions are found to agree qualitatively with experience, but the extent of quantitative agreement is not yet known. Experiments to determine this are being planned.

It should be mentioned that essentially the same methods and results apply to pedestrian traffic of a congested character. The bottleneck theory (§5) is particularly relevant to the movement of crowds through passages. However, the following exposition is confined to the more serious problem of vehicular traffic flow.

## 2. The flow-concentration curve

Although the flow $q$ and the concentration $k$ have no significance except as means, the purpose of the theory is to ask how they vary in space and time. However, on a long crowded road this is reasonable, since the means can be taken over relatively short distances or time intervals, and we are interested in variations over much greater distances and times.

The precise definitions of $q$ and $k$, at a given point $x$ on the road and a given time $t$, are included in the following instructions for measuring them. Draw two lines across the road, a short distance $dx$ apart, to form a slice of road with the point $x$ in the middle. Take averages over a time interval of moderate length $\tau$, with the time $t$ in the middle. The interval $\tau$ must be long enough for many vehicles to pass. Then the flow $q$ is

$$q = n/\tau, \tag{1}$$

where $n$ is the number of vehicles crossing the slice in time $\tau$. The concentration $k$ is

$$k = \frac{\Sigma\, \mathrm{d}t}{\tau\, \mathrm{d}x}, \tag{2}$$

where $\Sigma\, \mathrm{d}t$ means the sum of the times taken by each vehicle to cross the slice. Thus $k$ is the average number of vehicles $(\Sigma\, \mathrm{d}t/\tau)$ on the slice of road, divided by the length $\mathrm{d}x$ of the slice; in other words, $k$ is the number of vehicles per unit length of road.

A third important quantity is

$$v = \frac{q}{k} = \frac{\mathrm{d}x}{\frac{1}{n}\Sigma\, \mathrm{d}t}. \tag{3}$$

This is the 'space-mean speed' of Wardrop (1952), being both the ratio of flow to concentration and the ratio of length of slice to average crossing time. Thus it is an average of vehicle speeds weighted according to the time they remain on the slice of road. (If conditions were uniform, on the average, over a much longer stretch of road, $v$ would also be the average speed of all vehicles while they remain on that stretch; the further averaging with respect to time would then be unnecessary, since the fluctuations with time would become small for a long stretch of road. This explains the name 'space-mean speed'.) The time-mean speed, which we shall not use, is the unweighted average speed of vehicles crossing the slice, namely $n^{-1}\Sigma\, (\mathrm{d}x/\mathrm{d}t)$. This exceeds $v$. If speeds at a point are measured directly (as by a Radar speedmeter), instead of in terms of times, one can still derive the space-mean speed (Wardrop 1952) by taking the 'harmonic' mean of the observed speeds, namely,

$$\left(\frac{1}{n}\Sigma\frac{1}{\mathrm{d}x/\mathrm{d}t}\right)^{-1} = \frac{\mathrm{d}x}{\frac{1}{n}\Sigma\, \mathrm{d}t} = v. \tag{4}$$

Most road-traffic observers have concentrated on measuring $q$ and $v$, as being the quantities of greatest practical interest. The concentration $k$ must be obtained from such measurements by division. Sometimes, however, $k$ is observed directly by taking photographs of the road from above. Such results are sometimes quoted in terms of mean 'headway' (distance between the fronts of successive vehicles in the same lane of traffic). The mean headway is $N/k$, where $N$ is the number of lanes travelling in the direction considered.

Vehicle counts are sometimes made by moving observers, especially (Charlesworth 1950; Wardrop & Charlesworth 1954) by observers in cars filtered into the traffic. If an observer moving at uniform speed $U$ records the number of vehicles which pass him, minus the number which he passes, and divides the difference by the total time of observation (say $\tau$), the result is

$$q - kU. \tag{5}$$

(A number $q\tau$ of vehicles would pass him if he were stationary, but this is reduced by $k(U\tau)$, namely, the average number of vehicles in the distance $U$ which he travels.) By measuring expression (5) successively for two values of $U$ (in practice, values with opposite signs), $q$ and $k$ may be separately deduced.

This experimental method is closely linked to the basic theoretical idea of this paper. Consider two observers moving with uniform speed $U$, the second starting, and remaining, a time $\tau$ behind the first.* Suppose now that the flow and concentration are changing with time, but that nevertheless the observers adjust their speed $U$ so that the number of vehicles which pass them, minus the number which they pass, is, on the average, the same for each. Then by (5), $q-kU$ is the same for each, and so

$$U = \frac{\Delta q}{\Delta k}, \tag{6}$$

where $\Delta q$ and $\Delta k$ are the change in flow and concentration after time $\tau$.

Now, in the circumstances mentioned, the number of vehicles between the observers must remain the same. But the number of vehicles passing any point between the times at which the observers pass it is $q\tau$. Since $\tau$ is fixed, it follows that the flow $q$ *remains unchanged along the path of observers travelling with the speed* (6).

In other words, when changes of flow are occurring, the waves which carry such changes through the stream of vehicles travel at a velocity given by equation (6). This velocity, relative to the road, may, as we shall see, be positive or negative. However, it never exceeds $+v$, the space-mean speed; hence the waves are always transmitted backwards relative to the vehicles on the road.

Now, it has been conjectured by many authors that, on any uniform stretch of a road, the flow $q$ is a function of the concentration $k$. If this is true, equation (6) becomes especially valuable, since it shows that *small* changes of flow are propagated at the speed

$$c = \frac{\mathrm{d}q}{\mathrm{d}k}, \tag{7}$$

which is known if $k$ (or $q$) is known.

The relationship between flow and concentration has usually been stated in rather different forms. At low values of the concentration, the mean speed $v = q/k$ has been regarded as a function of the flow $q$ (Normann 1942; Normann & Walker 1949; Glanville 1949, 1951). It falls off as $q$ increases, with a slope which is steep for narrow roads but more gradual for wide roads. Wardrop (1952) ascribes the effects of increased flow, in the main, to increased interference with overtaking, which tends to reduce the mean speed to nearer the speed of the slowest vehicles on the road. Doubtless, a general sense of the greater possibility of accidents also contributes to the reduction in mean speed.

At high values of the concentration, however, most writers have regarded the 'mean headway' $N/k$ as a function of the mean speed $v = q/k$. At $v = 0$, the mean

* Imagine them to be cyclists on an adjacent cycle track, so that they can maintain their uniform speed $U$ unimpeded, and in turn will not influence the observed traffic flow (we are not suggesting this as a practical method of observation, but as a convenient way of thinking about the flow).

headway takes a value (around 17 ft. in Great Britain) only just greater than the average vehicle length. As $v$ increases, the mean headway increases almost linearly (by about 1·2 ft. for each 1 mile/h increase in speed). Many authors (see, for example, Normann & Taragin 1942) have interpreted such results by saying that a driver allows just enough headway so that no collision will result if the vehicle in front brakes suddenly, and he himself brakes after a certain 'reaction time'. Glanville (1949) points out that the observed rate of increase of headway with speed would correspond to a uniform braking force, equal for both vehicles, and a reaction time of 0·8 s. The reader may easily verify this. Attempts have been made to apply such considerations also at low values of the concentration, but then the greater freedom to overtake alters the situation completely.

Different experimental methods are appropriate for determining these two kinds of relationship. Our contention, however, is that the information obtained from these two sources should be combined into a single curve, and that the curve which sums up all the properties of a stretch of road which are relevant to its ability to handle the flow of congested traffic is a graph of the two fundamental quantities, flow against concentration.

The form of such a curve must be as in figure 1. As the concentration $k$ tends to zero, the flow $q$ must also become zero. Again, in the limiting case of high concentration $k = k_j$ ($j$ for jam) the vehicles travelling in a given direction are packed tight on the part of the road where they are permitted to be; the flow $q$ is then again zero. For some value of the concentration between these two extremes, the flow $q$ must have a maximum $q_m$, which may be called the capacity of the road.

The deduction in the last paragraph (which a mathematician would call an application of 'Rolle's theorem'!) does not seem to have been clearly made in the traffic-flow literature, except perhaps by Greenshields (1935). Considerable effort has been put into finding a suitable definition of road capacity, but it has not been noticed that the very simple and relevant one 'maximum flow of which the road is capable' is available.*

Experimentally, this was because flow at the particular concentration $k_m$ corresponding to this maximum flow is not often observed, for reasons which will appear later. Flow at smaller concentrations is commonly observed, and described by a speed-flow relation. (A description in such terms is inconvenient for the complete range of speeds, since there are two speeds for a given value of the flow.) Flow at concentrations near to $k_j$ is commonly observed, and described by a headway-speed relation. (This description is unsuitable at low concentrations because headway ceases to have significance when overtakings are prominent.)

To complete the curve satisfactorily, an independent measurement of $q_m$ and $k_m$ (flow and concentration for maximum flow) is desirable, since interpolation between the two measured parts of the curve is very difficult without knowledge

* Normann (1942) introduced a 'theoretical maximum capacity', obtained by assuming that the flow at all concentrations was governed by the theoretical speed-headway curve, but he points out that observed flows are hardly ever more than about half of this 'theoretical maximum'. The maximum here discussed, on the other hand, is the real, experimentally determined, maximum. Again, it should not be confused with a statistical 'extreme value', since the flow-concentration curve represents the average relationship between the quantities.

of some intermediate point. Fortunately, the theory of this paper provides a special method of measuring these two quantities, as follows.

If a stream of vehicles is stopped, as at a traffic light, and then started again after a considerable delay, as when the lights go green, a system of waves is emitted.* Each carries a particular value of the flow $q$ and concentration $k$, and hence also a particular value of the wave velocity $c$, and propagates with this uniform velocity, some forwards and some backwards (see §6 below). One wave alone remains stationary at the original stopping-point. Now this wave has $c = 0$, so by (7) it corresponds to a value of $k$ for which $\mathrm{d}q/\mathrm{d}k = 0$, namely, to $k = k_m$, for which $q$ is a maximum. This shows that the mean flow and concentration measured

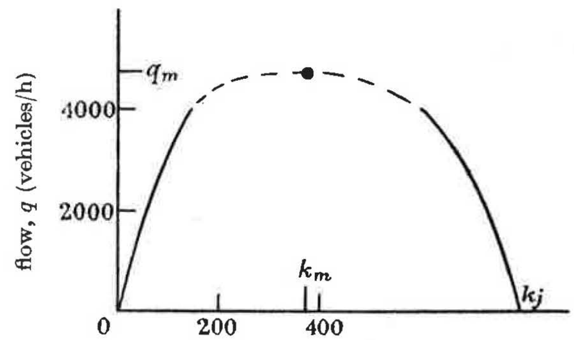



FIGURE 1. A flow-concentration curve.

*at the stopping-point itself* (after the stream of vehicles has started up, and before all those slowed down by the original stoppage have passed through—the need for these restrictions will become clear in §6) are the required quantities $q_m$ and $k_m$.

A typical flow-concentration curve constructed in the manner indicated is shown in figure 1. The full line on the left is derived from speed-flow data, that on the right from headway-speed data, and the central point $(q_m, k_m)$ from measurements at the stopping-point after a long line of traffic had been stopped and then allowed to flow forward freely again. The curve refers to a certain one-way three-lane section of the Great West Road, and the speed-flow data were obtained during the period of peak evening traffic between 5 and 7 o'clock. The authors are grateful to the Director of the Traffic and Safety Division, Road Research Laboratory, for permission to use the unpublished results displayed on this curve.†

* A really long lane of vehicles must be stopped if the theory is to be applicable, as will appear later (§6).

† Mr Wardrop has recently indicated to the authors that he would now consider a rather lower value (say 3200) more typical of the flow $q_m$ past a stopping point on this particular stretch of road than the earlier value (round 4700) supplied to the authors and quoted in figure 1. However, R.R.L. measurements for single-lane traffic yield values of $q_m$ of 1500 v/h., so that values of around three times this would be expected for three-lane traffic. If they were not observed, the cars were probably not filling the three available lanes when stopped. The flow $q_m$ will be achieved only if all available lanes are fully used.

Another method of deriving the curve was used by Greenshields (1935), who plotted $v = q/k$ against $k$ for one-lane traffic, as in figure 2, and drew a straight line through his points. This involved a rather drastic interpolation since there is a large intermediate range where there are no points, and where in fact the true curve probably lies below the straight line. However, the method gives a simple and probably not too inaccurate result, which led to the predicted existence of a maximum flow on any road much earlier than had been inferred elsewhere, as mentioned above. Greenshields introduced a 'kink' at the top of his graph, to

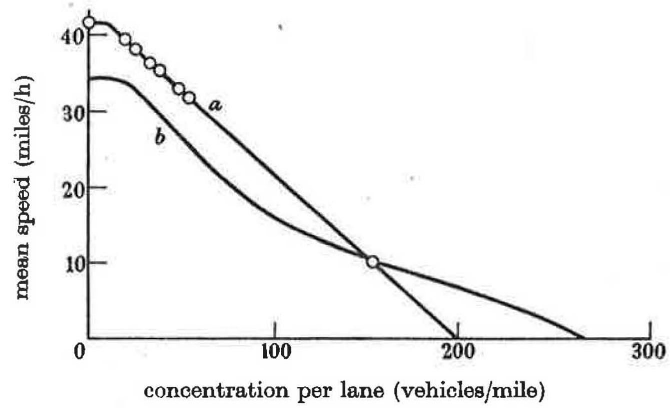



FIGURE 2. Two examples of a speed concentration curve. $a$, Greenshields; $b$, Road Research Laboratory (see figure 1).

make the speed flatten out at the independently determined 'free speed' for the road. A flat portion like this must be expected on any speed-concentration curve, since the mean speed will be unaffected by concentration below a certain limiting value. On a wide road like that of figure 1 this limit may be as much as 50 vehicles per mile.

One may use the word 'crowded' to describe road conditions on which the concentration exceeds this limit. Then a road is crowded if any increase in concentration will lead to a reduction in mean speed. The theory of this paper is applicable only to long, 'crowded' roads.

For comparison with Greenshields's result the curve corresponding to figure 1 is also shown in figure 2, with the densities divided by 3 to allow for the greater number of lanes. In comparing the two curves, one must bear in mind the differences between English and American driving habits and vehicle lengths.

The two curves are shown also in figure 3, as flow-concentration curves per lane of traffic. That of Greenshields is the arc of a parabola with vertex upwards. A portion of the arc near the origin is replaced by a chord through the origin. This corresponds to a range of non-'crowded' conditions, in which the mean speed is constant.

To conclude this section it may be noted that the flow-concentration curve for a particular stretch of road may vary from time to time (especially with the day of the week, but also with the time of day), owing to changes in the proportion of commercial vehicles on the road, or in the quantity of traffic travelling in the opposite direction. Some care is therefore needed in specifying the conditions under which a particular determination of the curve has been made. Again, the variations along a given road, due to differences of width, gradient, curvature, population density, etc., between different stretches of the road, may be very great. The velocity of a wave in any one stretch of road, however, will be given by the slope of the flow-concentration curve for that particular stretch of road, as the argument leading to equation (6) makes clear. The use of the theory in such cases is possible, therefore, and will be fully illustrated in §5.

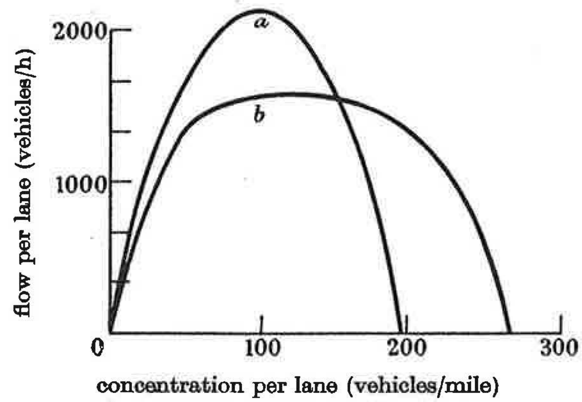



FIGURE 3. Flow-concentration curves per lane of traffic. *a*, Greenshields; *b*, Road Research Laboratory.

### 3. Use of the flow-concentration curve

To make practical use of the flow-concentration curve for a particular stretch of road, a geometrical expression of the results of §2 is often valuable.

First, note that, corresponding to any point on the curve, the space-mean speed $v = q/k$ (under the conditions represented by that point) is the slope of the radius vector from the origin (figure 4). The speed $c = \mathrm{d}q/\mathrm{d}k$ of waves carrying continuous changes of flow through the stream of vehicles is the slope of the tangent to the curve at the point (figure 4). This slope is the smaller,* provided that the mean speed *decreases* with increase of concentration; in other words, if the road is 'crowded'. For we can write

$$c = \frac{\mathrm{d}}{\mathrm{d}k}(kv) = v + k\frac{\mathrm{d}v}{\mathrm{d}k}, \tag{8}$$

which is less than $v$ if $\mathrm{d}v/\mathrm{d}k$ is negative. The velocities $c$ and $v$ are equal only at low concentrations, below the limit (mentioned in §2) at which significant interaction between different vehicles on the road first occurs. At such concentrations, $\mathrm{d}v/\mathrm{d}k = 0$.

To express velocities as slopes in this way is convenient if conditions on a road are to be represented in a space-time diagram. If the road is represented as

* Meaning that waves travel backwards relative to the mean vehicle flow.

stretching up the paper, with time travelling to the right, then a path on this diagram, representing the motion of a wave or of a vehicle, will have a slope $dx/dt$ equal to the velocity. Since lines of equal slope are parallel, it follows that a mean vehicle path on this diagram must be parallel to the radius vector from the origin to the relevant point on the flow-concentration curve, while a wave must be parallel to the tangent to the curve.

A second use of the flow-concentration curve refers to discontinuous waves. These are likely to occur on any stretch of road when the traffic is denser in front, and less dense behind. For waves on which the flow is less dense travel forward

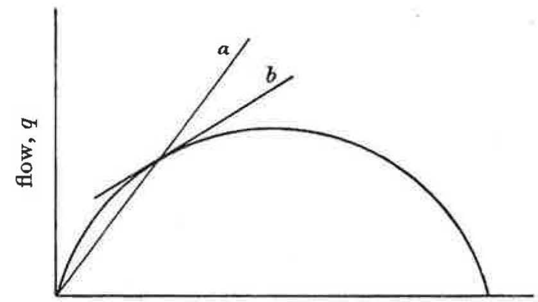



FIGURE 4. Use of the flow-concentration curve. Slope of radius vector ($a$) gives average velocity of vehicles; slope of tangent ($b$) gives wave velocity.

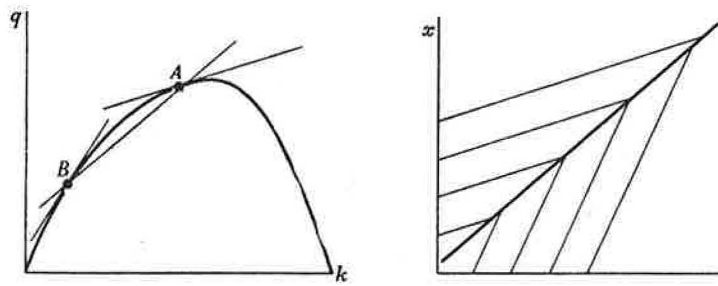



FIGURE 5. Use of flow-concentration curve to predict the local conditions near a shock wave.

faster than, and hence tend to catch up with, those on which the flow is denser. When this happens a bunch of continuous waves can coalesce into a discontinuous wave, or 'shock wave'. When vehicles enter this their mean speed is substantially reduced very quickly. The wave is not totally discontinuous of course, but its duration is not much longer than the braking time that each vehicle needs to make the required reduction of speed.

The speed of a discontinuous wave, or shock wave,* is given by (6) as $\Delta q/\Delta k$, the slope of the *chord* joining the two points of the flow-concentration curve which represent conditions ahead of and behind the shock wave. (Note that the argument

* In future we shall prefer the latter name, suggested by the very strong analogy with shock waves in gases.

leading to (6) is applicable, provided that the time interval $\tau$ between the two observers exceeds the duration of the shock wave. The number of vehicles between two observers with the shock wave between them can remain constant only if they travel at the speed of the shock wave.)

Figure 5 illustrates the use of the flow-concentration curve to predict conditions near a shock wave. The shock wave is shown as a heavy line on the space-time diagram on the right. Ahead of it the flow is denser and the waves (plain lines) are drawn parallel to the tangent to the flow-concentration curve at $A$. Behind it the concentration is less and the waves travel faster; they are drawn parallel to the tangent to the curve at $B$. The shock wave, generated by the running together of these waves, travels at an intermediate speed, and must be drawn parallel to the chord $AB$. The mean vehicle paths (not shown) would be parallel to the radius vectors $OB$ (behind the shock wave) and $OA$ (ahead of it).

## 4. The progress of a traffic hump

As a first illustration of the method we apply it to a problem where the road is uniform, so that all stretches of it have the same flow-concentration curve. In these circumstances, each continuous wave is propagated at a constant velocity $c$, since $q$ is constant along it. In a space-time diagram the wave paths are straight lines, each parallel to the tangent to the flow-concentration curve at the corresponding point.

The source of traffic is taken to be at one end of the road, and we consider the case when the inflow rises to a peak and then falls to its original value, producing a traffic hump.* The rise and fall of inflow can be easily measured by an observer at the feed point. A problem of some importance is then: How can the behaviour of the hump as it passes down the road be predicted in advance? For example, when will it reach a given point? Will it spread out, or become more concentrated, and how fast? How will it affect average journey times?

The wave theory gives convenient answers to these questions. Figure 6 shows the wave pattern in a space-time diagram. The wave path starting from the feed point at any time is parallel to the tangent to the left-hand part of the flow-concentration curve at the point which corresponds to the inflow at that time. The waves travel more slowly inside the hump than outside it. Hence the wave paths in figure 6 'fan out' at the front and become concentrated at the rear, where they must ultimately run together.

It must be emphasized that the lines drawn are 'waves' (lines of constant flow, and hence also, for a uniform road, lines of constant mean speed) and *not* vehicle paths. Vehicles go (on the average) faster than the waves, and most vehicles starting at the rear of the hump will in time get through it. On entering the hump a driver has to slow down fairly rapidly (since the lines of constant speed are

* Traffic humps (regions of increased concentration) generated in *this* way have concentrations remaining solely on the left-hand half of the flow-concentration curve. But humps at the much higher concentrations corresponding to the right-hand half have similar properties; the only important difference is that the waves travel backwards relative to the road. Examples of humps of *this* kind occur below, especially in the theory of bottlenecks (§5).

bunched together on the right of figure 6), but on leaving it he can increase his mean speed only slowly as he traverses the fan of waves on the left.

Figure 6 gives a clear answer to the question of the speed of the front of the hump, which turns out to be the wave velocity associated with conditions in front of it. Note that this may be considerably less than the space-mean speed (which in turn is less than the time-mean speed) of the vehicles in this region. The other questions noted above can be answered only after the path of the shock wave, which results from the running together of the waves at the rear of the hump, has been determined.

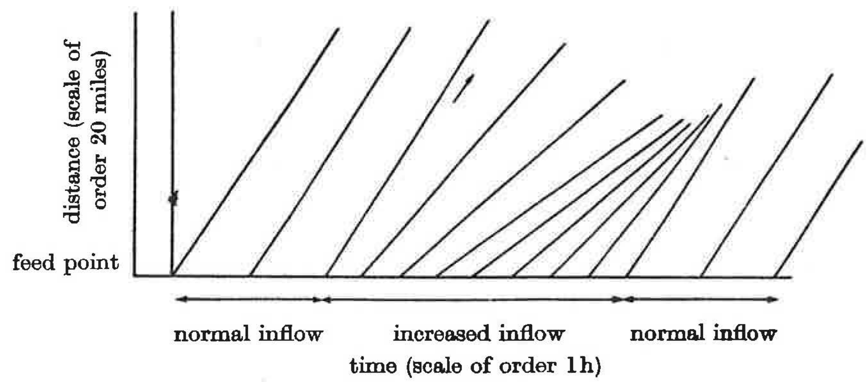



FIGURE 6. Wave forms in traffic hump.

The shock wave starts at the point where two waves first run together, and its progress after that is governed by the simple law stated in §3: at each point of the shock, the two waves which meet there are represented by two points on the flow-concentration curve, and the shock wave path must be drawn parallel to the chord joining those points. This gives a straightforward geometrical step-by-step method for constructing the path of the shock wave.

In practice it is convenient to note that the slope of the chord is approximately the mean of the slopes of the tangents at its end-points, so that the speed of the shock wave is approximately the mean of the speeds of the waves running into it from either side. This approximation is exact for a parabolic arc with vertical axis, such as Greenshields's flow-concentration curve (figure 3). For other smooth curves with nothing approaching a vertical tangent, the approximation is still fairly good, as the known series for the slope of the chord,

$$\frac{q(k_2)-q(k_1)}{k_2-k_1} = \tfrac{1}{2}(q'(k_1)+q'(k_2)) - \frac{(k_2-k_1)^2}{24}(q'''(k_1)+q'''(k_2)) + \ldots, \tag{9}$$

shows. In view of the approximate character of the whole theory, the additional approximation is probably worth making wherever it will make an effective simplification.

It certainly makes the shock wave easier to draw in by eye, as no further reference to the flow-concentration curve is then necessary. One has simply to draw a path on the space-time diagram whose slope at any point is the mean of the slopes of the waves running into it from either side. This process is illustrated in figure 7; it can be mastered with only a little practice.

As an alternative, or as a check, one has the analytical solution for the shock path (Whitham 1952) which again is based on the approximation noted above. This also can be expressed as a geometrical construction, as follows.* Given the

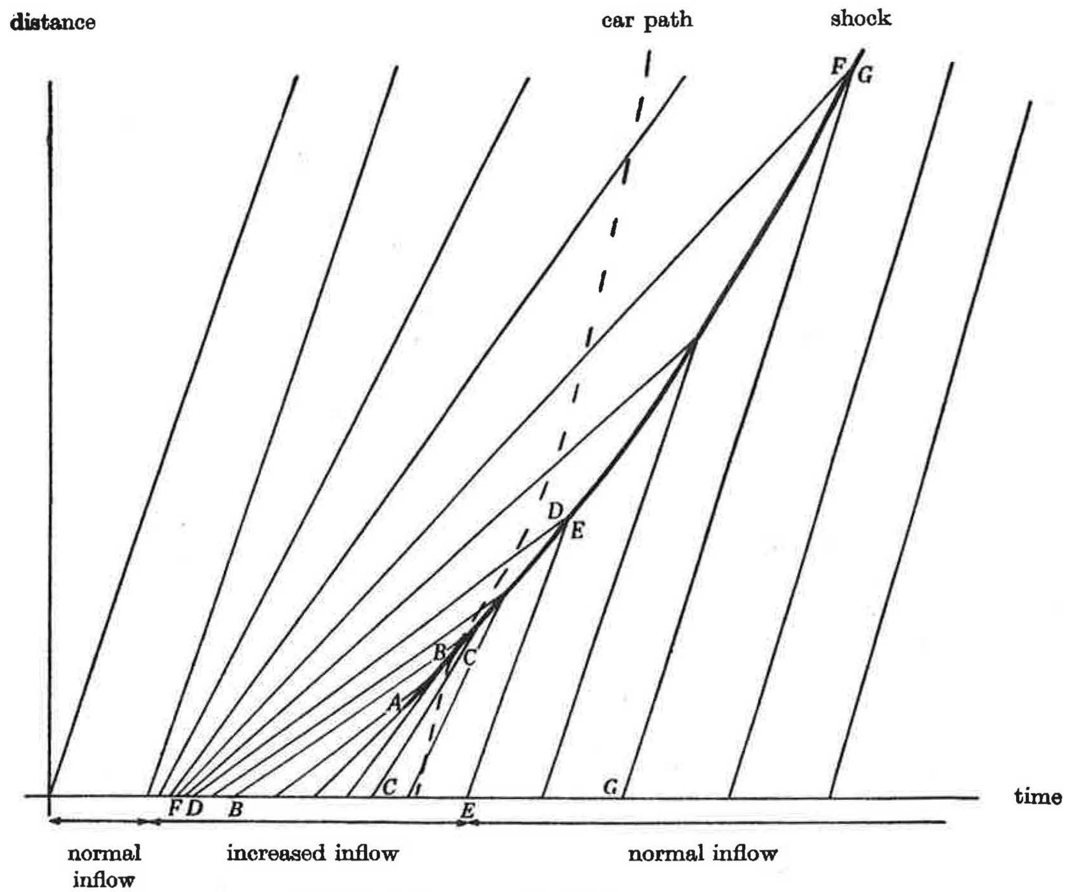



FIGURE 7. Progress of traffic hump with time.

variation with time $t$ of the inflow rate observed at the feed point, plot a graph (figure 8) with the corresponding wave velocity $c$ (the slope of the flow-concentration curve for the observed value of the inflow) as ordinate, and its product with the time, $ct$, as abscissa. Then the time at which the shock wave first appears is

* The present problem is somewhat simpler than that treated by Whitham (1952), in which our ordinate $c$, the rate of change of $x$ with respect to $t$ on a wave, is replaced by $F(y)$, the rate of change of $\alpha r - x$ with respect to $kr^{\frac{1}{2}}$; and in which the abscissa is $y$, the value of $x$ when $r = 0$. The analogous abscissa in our problem is evidently the value of $-x$ when $t = 0$. For the wave which passes $x = 0$ at time $t$, with velocity $c$, this is $ct$. Readers of part I should note that another approach, in which $c^{-1}$ and not $c$ replaced $F(y)$, was found convenient there (§4); however, that approach cannot be used if the flow-concentration curve has a stationary point.

given by the reciprocal of the slope of the tangent to this graph at its right-hand point of inflexion $A$; the value of $c$ (or $t$) at $A$ also determines, through its velocity (or time of origin, respectively), *on which wave* the shock wave first forms. To determine the further progress of the shock wave, draw chords on the graph (e.g. $BC$, $DE$, $FG$) which cut off lobes *of equal area* above and below between them and the curve. Then the slope of any one of these chords is the reciprocal of the time at which the shock wave absorbs the two waves on which $c$ and $t$ have the values corresponding to the end-points of the chord.

It is evident from this construction that the shock wave initially grows in strength, the maximum increase in concentration at the shock wave occurring when one of the end-points of the chord is somewhere near the bottom of the graph (see $BC$ in figure 8, and also in figure 7, where the wave corresponding to each point in figure 8 is marked with the same letter, so that points on the shock in figure 7 are marked exactly like the chords in figure 8 which correspond to them). At this time vehicles entering the hump suffer instantaneously almost the full reduction of speed associated with it. The path of such a vehicle is indicated by the broken line.

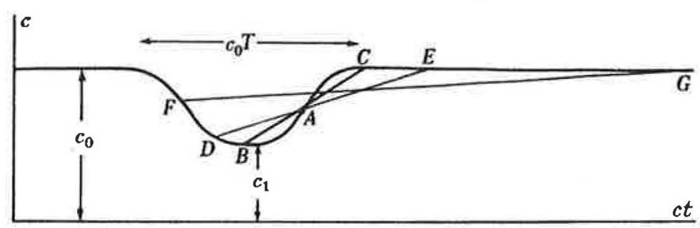



FIGURE 8. Geometrical construction for the shock wave.

As time goes on, however, the left-hand end-point in figure 8 penetrates farther and farther into the front part of the hump, so that the shock wave absorbs, one after another, all the waves on which there is substantially increased density. When this process is completed, the hump has disappeared and what remains of the shock wave is negligibly weak. This happens after a time equal to the reciprocal of the slope of $FG$ (figure 8), where $F$ is a point at which $c$ is sufficiently near to the value it takes on the left of the graph. Note, however, that the section of road satisfying the conditions postulated may in many cases come to an end before the hump is dispersed in this way.

Regarding the hump as a region of increased concentration, it may be asked how the excess of vehicles can effectively vanish in this way. The answer is that the region of increased concentration spreads backwards (relative to the front of the hump, which has a constant mean speed), so that the excess of vehicles is dispersed over a constantly increasing length of road. A quantitative estimate of the process may be obtained if one knows the duration, say $T$, of the increased inflow at the feed point, the wave velocity $c_0$ outside the hump and the lowest value, say $c_1$, of the wave velocity inside the hump. Then the shock wave is at its strongest at a time about

$$\frac{c_0 T}{2(c_0 - c_1)} \tag{10}$$

after the time of maximum inflow. At this time (corresponding to $BC$* in figures 7 and 8) the hump has hardly spread backwards at all; it has simply altered its shape so that the increase of concentration is sudden and the subsequent decrease is spread over the whole length of the hump. Later, the decrease of wave velocity at the shock wave becomes a small quantity $\delta$ after a time†

$$t = \frac{c_0(c_0 - c_1)\,T}{\delta^2}, \tag{11}$$

and the length of the hump is then about

$$l = \{c_0(c_0 - c_1)\,Tt\}^{\frac{1}{2}}, \tag{12}$$

which may be compared with its original length $c_0 T$.

It is interesting to compare this result with the results of ordinary diffusion processes. It corresponds to a diffusion coefficient of the order of $c_0(c_0 - c_1)\,T$, namely, the product of the length of the hump and the maximum reduction in wave velocity within it. By comparison, any diffusion which may be present due to statistical fluctuations with a mean free path, or due to a dependence of mean flow on concentration-gradient as well as on the concentration itself (see part I, and § 6 below), would have a diffusivity independent of the length of the hump. This indicates that diffusion by the wave process described in the present section will at any rate be predominant for sufficiently long humps—in other words that, for *sufficiently* 'long, crowded roads', the present theory is appropriate.

## 5. A theory of bottlenecks

We now consider a typical problem where the capacity of the road varies along it. We suppose that some bottleneck is present, where the maximum possible flow $q_m$ falls to a lower value than on the main part of the road. Then, presumably, the whole flow-concentration curve is reduced in its vertical scale. (It may well be reduced in horizontal scale too (that is, $k_j$ may become less), but figures 9, 11 and 13 illustrating the theory have actually been drawn for the case where this does not happen.) The local minimum value of $q_m$ may be called the *capacity* of the bottleneck.

We consider first a stream of vehicles approaching the bottleneck at a flow rate which remains always *less* than its capacity. Then each vehicle suffers simply a temporary reduction in speed as it passes through. The waves are also reduced in speed while in the bottleneck. For the flow $q$ remains constant on any wave, as was shown in § 2 independently of whether the flow-concentration curve varies with position. Hence (figure 9) conditions on a wave as it passes through the bottleneck are represented by points of flow-concentration curves all at the same horizontal level. Since the tangent to the lower curves at a given horizontal level has a smaller slope, the wave velocity is reduced inside the bottleneck, and the

* The quantities $c_0$, $c_1$ and $T$ are indicated in figure 8, and it is evident that the slope of $BC$ is approximately the reciprocal of (10).

† The area of the hump in figure 8 is about $\frac{1}{2}(c_0 - c_1)\,c_0 T$, and this will be equal to the area above $FG$, namely, $\frac{1}{2}\delta^2 t$, where $t^{-1}$ is the slope of $FG$, if (11) holds. Here $c_0 - \delta$ is the value of $c$ at $F$.

wave paths behave as in figure 10. Under the conditions illustrated in this figure the delay to each vehicle is relatively small.

Next, we consider the more serious hold-up resulting when, as time goes on, the oncoming flow rate increases above the capacity of the bottleneck. Waves then turn back before reaching the centre of the bottleneck and form a shock wave. This passes back down the main road and forces vehicles to pile up behind the bottleneck at a rate given by the *difference* between the oncoming flow and its capacity. In practice, the oncoming flow would exceed the capacity of the bottle-

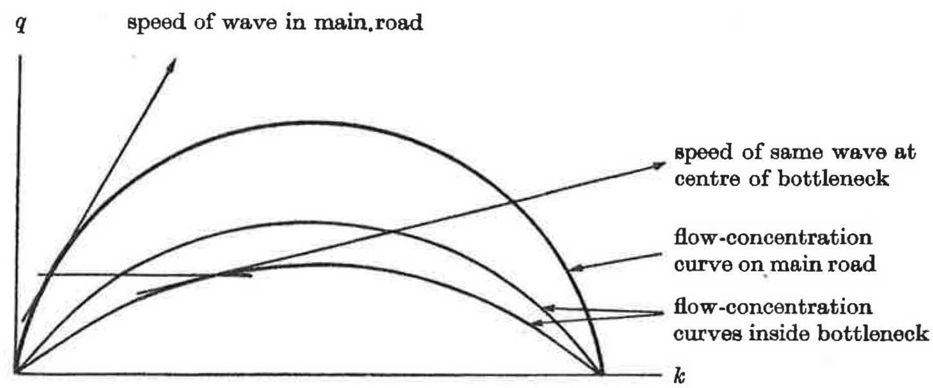



FIGURE 9. Variation of flow-concentration curve in a bottleneck.

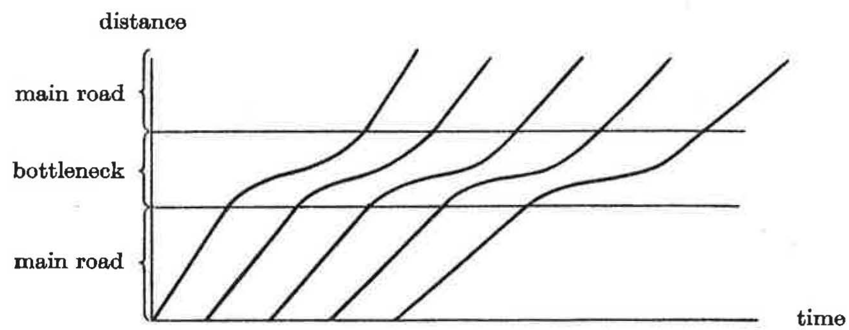



FIGURE 10. Passage of waves through a bottleneck, the capacity of which exceeds the incoming flow rate.

neck only for a finite time, during which the oncoming traffic is in the form of a hump. An important question is the duration of the hold-up resulting from the passage of a given traffic hump through the bottleneck. This will be solved by a detailed study of the shock wave paths.

To understand the formation of the characteristic 'bottleneck shock wave', note that no wave carrying a flow exceeding the capacity of the bottleneck can possibly pass through it, since the flow must remain constant on the wave, and such a large flow is impossible in the centre of the bottleneck. It is not important at which precise point of the bottleneck the wave turns back, but theoretically (if the flow-concentration curve varies continuously through the bottleneck) it should do so at the point where the flow carried by the wave is the maximum possible flow;

for here only is the wave velocity (slope of the tangent to the flow-concentration curve) zero. In figure 11, this point is $B$; the slope of the tangent at $C$ indicates the speed at which the wave will come out of the bottleneck again. Compare the points $A$, $B$, $C$ in figure 12, which shows in a space-time diagram the turning back of such a wave. For short bottlenecks, the details of the predicted flow within the bottleneck could not be relied on. However, the qualitative fact that the wave turns back, and its progress beyond $C$, are predictions on which greater reliance can be placed.

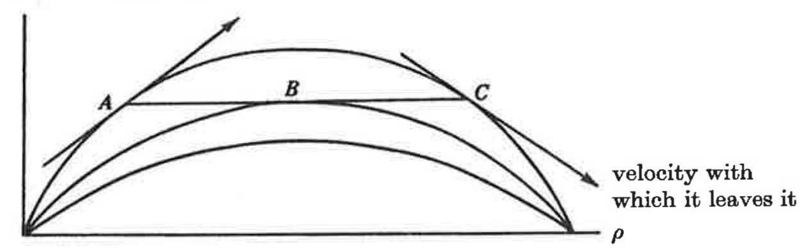



FIGURE 11. Illustrating the 'reflexion' of a wave from a bottleneck.

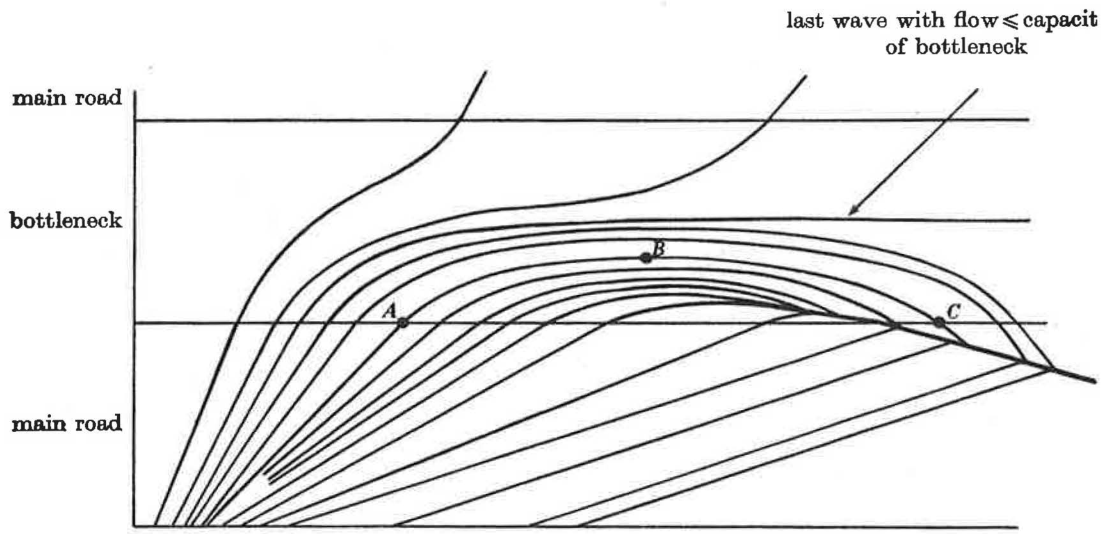



FIGURE 12. Formation of shock wave in the front of a hump as it enters a bottleneck of inadequate capacity.

The need for waves to intersect is at once evident from figure 12, where the beginning of the resulting shock wave is sketched in. This shock wave involves a reduction of flow, so its velocity (the slope of the chord joining points on the flow-concentration curve corresponding to conditions in front and behind) must be backwards relative to the road. As soon as it passes back out of the bottleneck, it must reduce the oncoming flow to almost exactly the capacity of the bottleneck. This is because waves carrying flows less than this have passed through, and waves carrying greater flows have turned back and been absorbed by the shock wave, so that only waves carrying flows approximately equal to the capacity of the bottleneck remain in its neighbourhood. Those just behind it are travelling backwards, corresponding to a point (e.g. $B$ in figure 13) on the *right-hand half* of the flow-concentration curve for the main road, at a flow level corresponding to the capacity of the bottleneck. The speed of vehicles in the slow crawl up to the bottleneck is given by the slope of $OB$. Conversely, the waves just ahead of the bottleneck are travelling forwards, corresponding to a point (e.g. $F$ in figure 13) on the left-hand half of the curve. Thus, vehicles after passing through the bottleneck are able to accelerate up to a mean speed given by the slope of $OF$.

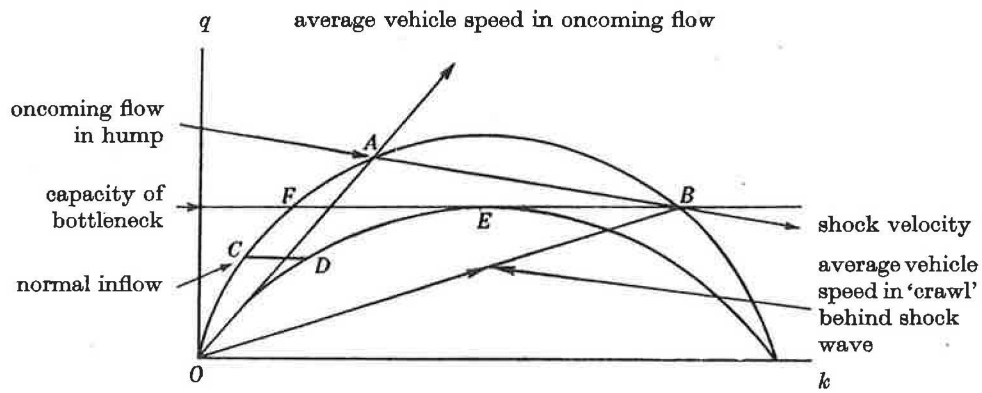



FIGURE 13. Illustrating 'crawl' produced by bottleneck and its final resolution.

The growth of the queue of crawling vehicles behind the bottleneck is easily calculated from the shock-wave path. For example, at a point where the oncoming wave carries a flow specified by the point $A$ in figure 13, the shock-wave velocity is the slope of $AB$.*

How will the deadlock be resolved? Evidently the shock wave will continue to move backwards until the point $A$ falls below the level of $B$, in other words, until the oncoming flow starts being less than the capacity of the bottleneck. If this improved state of affairs continues for long enough, the shock wave will move far enough forward to pass through the bottleneck. On doing so it will greatly increase its speed, for conditions downstream of the bottleneck are respresented by the point $F$ in figure 13, so that the shock-wave speed will be the slope of a chord such as $CF$. Thus, after it has passed back through the bottleneck, the shock wave will be just like the ordinary shock wave in the rear of any traffic hump (§4).

These considerations enable the course of the hold-up, and its approximate duration, to be determined graphically if the approaching hump is known, for example, if the variation of flow with time has been measured at some upstream point. The situation is little changed if there is already a shock wave in the rear of the approaching hump, as is likely in practice to be the case. When this meets the 'bottleneck shock wave', the two shock waves 'unite', a familiar process in

* The fact that increases of concentration from values well below $k_m$ to values well above it are normally made (as here) by means of shock waves, explains why (as noticed in §2) the maximum flow $q_m$ of a road is not often observed.

gas dynamics. No alternative behaviour is possible, as whatever they become has got to change the flow and concentration from their values behind the hump shock wave to the values associated with the bottleneck crawl. This could not be done by means of two shock waves, for example, because the one behind, which has to make the first increase of concentration, would have a greater speed than the one in front, which is responsible for the final increase to the crawl concentration; this relationship between speeds follows inevitably from the fact that the flow-concentration curve is convex upwards, but, on the other hand, is geometrically impossible since both waves must start at the same time.

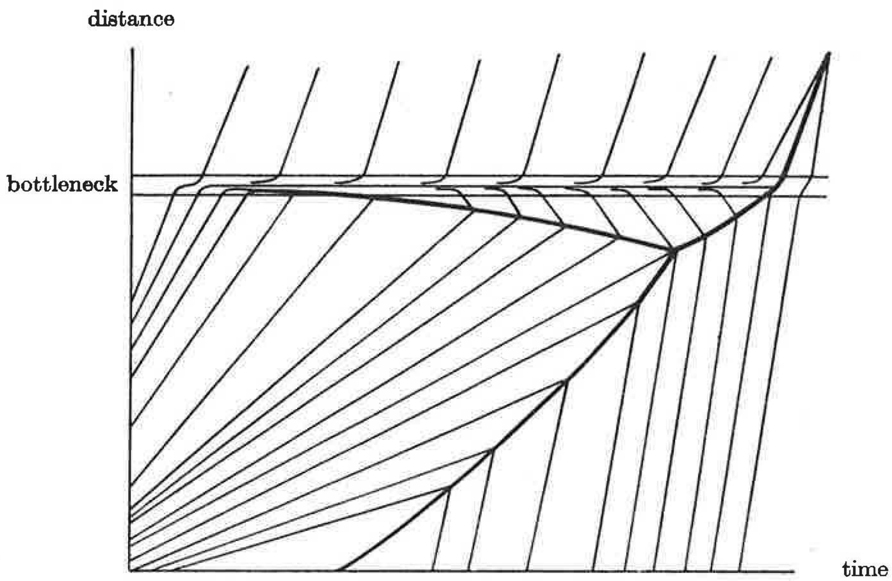



FIGURE 14. Resolution of 'crawl' by arrival of the shock wave in the rear of oncoming traffic hump.

The case when a bottleneck crawl is resolved by the union of the shock wave in the rear of the approaching traffic hump with the 'bottleneck shock wave' is illustrated in figure 14. The path of the shock wave formed by this union is easily traced, since it is still governed by the condition that the flow in front of it is equal to the capacity of the bottleneck—the concentration taking the *greater* of the two values compatible with this flow rate *upstream* of the bottleneck, and the lesser one downstream of it. It is important to notice that the only data required for estimating the course of a bottleneck hold-up in this manner are the flow-concentration curve for the main road, the capacity of the bottleneck, and the variation of inflow with time measured at some upstream point.

As a final theoretical point, it may be noted that the flow near the bottleneck during the crawl is steady. It has often been remarked that the increase of speed on the passage of vehicles (or crowds) through a bottleneck under steady conditions is similar to the effect of a Laval nozzle on the flow of a gas. The above analysis shows how close the similarity is. Upstream of the bottleneck the waves are propagated upstream (as sound waves can be in subsonic flow); downstream of it they are propagated downstream (as sound waves must be in supersonic flow). As the centre of the bottleneck is approached, the mean speed $v$ is increased, and the wave velocity relative to the mean vehicle speed (namely, $v-c$) is decreased, so that both are equal, just as the fluid velocity equals the velocity of sound in the throat of a Laval nozzle. The only essential difference* between the two situations is that the gas is able to transmit disturbances forwards as well as backwards relative to the mean flow. It is this that made the above analysis of the *transients* in the traffic flow problem so much easier than it is in the problem of the Laval nozzle.†

On a road with several bottlenecks in rapid succession, the one with least capacity will define the greatest flow possible under steady conditions. An inflow of vehicles exceeding this capacity can only pile up in a continually increasing 'queue' or 'crawl' in front of the bottleneck system. In the steady part of the flow, the flow $q$ is uniformly equal to the capacity of this narrowest bottleneck, while the concentration $k$ takes the larger value appropriate to this flow upstream of that bottleneck, and the smaller value downstream.

The transients could easily be worked out in this problem. As a hump enters the system of bottlenecks, a shock wave is first formed at the narrowest one (at least if the flow increases slowly enough), and begin to move upstream. If there is a slightly wider bottleneck farther upstream, a shock wave might later form there too, perhaps before the first shock wave had reached it. However, in due course the first shock wave would catch it up, as its speed backwards is greater, and so the two would unite into a single shock wave reducing the oncoming flow to the capacity of the narrowest bottleneck.

## 6. Some notes on traffic flow at junctions

In this section we attempt a preliminary study of how the method of this paper might be used to predict traffic behaviour at road junctions of various kinds. First, we consider junctions which are not 'controlled' (either by police or by traffic lights).

The simplest junctions are those where minor roads introduce new traffic on to, or abstract traffic from, such a long arterial road as has been considered in the preceding sections. This is normally achieved without significant impedance to the traffic on the major road. Vehicles wishing to enter it have to wait until they can do so without causing obstruction. Vehicles leaving the major road have often to

* A less essential, though more spectacular, difference is that in the traffic problem the typical 'unchoked' flow is totally supersonic, instead of totally subsonic. But in both problems both possibilities exist.

† Students of gas dynamics may wonder, on reading this, whether a rough approximation to the calculation of transients in a Laval nozzle might not be made by regarding them as kinematic waves, on the approximation (*accurate* only for steady flow) that the stagnation enthalpy is everywhere constant. This is found to give the wave velocity as $v-a^2/v$ instead of $v-a$ (where $a$ is the local speed of sound), so that its quantitative value would be small, but it might indicate qualitative behaviour reasonably correctly.

slow down, or even stop for a time, before they can leave it, but they usually signal their intention in time to enable vehicles behind to pass them on the appropriate side with little loss of speed.

The effect of such a junction on a wave moving past it along the major road, is then to change the flow carried by the wave by an amount equal to the 'mean net inflow rate' from the minor road. This rate is defined as the difference between inflow and outflow, smoothed (as a function of time) by averaging it over such a time $\tau$ as was considered in § 2. If the road is 'crowded', in the sense defined in § 2, the change in flow will change also the speed $c$ of the wave, as well as the mean vehicle speed $v$. In a space-time diagram, therefore, the waves bend slightly at junctions (backwards where the net inflow is positive, forwards where negative). These rules enable the arterial road theory of §§ 4 and 5 to be corrected for minor inflows and outflows at junctions.

However, there is a limit to the amount of inflow (especially) which is possible under those conditions. Further, this limit becomes more and more reduced as the flow on the major road increases. These are truisms. It might be thought, however, that the limit was just that increase of flow which would be required to raise the flow on the major road to the maximum possible. The real limit, however, is always much less than this. For inflow under most conditions can occur only when gaps in the traffic pass the junction. As the flow increases, such gaps become rarer and rarer, and for large enough flows, but still well below the capacity of the road, the gaps may be too rare to permit any significant inflow at all.

At cross-roads, where some traffic on the minor road seeks to cross the major road, a closely similar limit exists on the *total flow* originating from the minor road. (This is the sum of the inflow and the cross-over flow.) Evidently, if the minor road carries a flow exceeding this limit, the major road may act for the time being as an effective bottleneck, for the flow on the minor road, which could then be treated by the theory of § 5.

It will now be clear why stoppages occurring at junctions under heavy flow conditions can often be resolved by sending a policeman to control the junction. If he stops successively the traffic on the major, and then on the minor, road, the flow originating from each will be approximately the maximum for the road during nearly all the period when the other road is stopped (see § 2 above, and also the discussion which follows). The total flow can therefore be made fairly near to this maximum (or, if the capacities of the roads are different, to a weighted mean of them), and this will be greater, as just explained, than what can be achieved under uncontrolled conditions. To achieve best results, the policeman gives each road a time allocation proportional to the flow originating from it. Where traffic lights are installed, one can allocate times on the basis of a mean ratio of flows over an extended period, or else use a vehicle-actuated system of a type calculated to give a better approximation to the optimum at any instant.

Where major roads meet at the same level, a roundabout is preferable to a simple controlled crossing. For this to remain effective under the heaviest traffic conditions, the circular arcs of road which compose the roundabout should each have a capacity equal to one-quarter of the sum of the capacities of the roads radiating from it. For on the average each vehicle uses half the total number of arcs, so that the average flow in an arc will be half the total inflow, or one-quarter of the total flow (inflow and outflow) on all the radial roads. When there are four of these, the argument indicates a width for each arc equal to that of one of the radial roads. Since excessive width for the arcs reduces safety, it may be that these limits should be closely followed.

To conclude the paper, we describe an attempt to discover whether the theory can be successfully applied to flows on a small scale, by using it to predict the effect on the oncoming flow of the compulsory stops and starts at a controlled junction.

First, consider the effect of a sudden stoppage (as when traffic lights turn red) on a *uniform* oncoming flow. It sends a shock wave back into the oncoming stream, at which the flow is reduced to zero and the concentration increased to approximately $k_j$, the maximum concentration of which the road is capable. (As when the union of two shock waves was discussed in § 2, there is no possibility but a single shock wave in this situation, since if there were more than one wave involved the velocity forwards of the wave making the *first* increase in concentration would have to be *greater* than that of the others, and this is impossible because all originate at the same place and time, and the first wave must be at the rear.) The speed of the shock wave is the slope of the chord on the flow-concentration curve which joins the point representing the oncoming flow to the point $(k_j, 0)$.

A more difficult question, where the limitations of the theory become apparent, is what happens when the traffic is permitted to flow forward again (as when the lights turn green; we ignore, to start with, complications due to some vehicles seeking to turn right or left at the junction). The solution, when the assumptions of the theory are retained without change, will first be given in detail (it was already indicated in § 2) and afterwards criticized.

The front vehicle can accelerate unhindered to a speed characteristic of an unimpeded road, but the theory ignores the time taken for adjustments of speed (consequent on changes of concentration) to be made. Hence, it represents the front of the stream as moving off instantly at a mean velocity equal to the 'free' mean speed $v_F$. The wave velocity is also $v_F$, both being the slope of the flow-concentration curve at the origin. At the same time a wave starts backwards through the stream of waiting vehicles, giving the signal to start. This has a (negative) velocity equal to the slope $c_j$ of the flow-concentration curve at the right-hand limit (corresponding to 'jam' conditions). In between these two extremes there is room for waves of all intermediate velocities, each carrying a corresponding mean vehicle velocity. Since in conditions when the wave velocity is greatest in front there is no tendency for waves to run together and form shock waves, we may suppose that only continuous waves will be present and so that the increase in speed will be achieved through a fan of waves of all possible velocities.

Figure 15 shows the shock wave produced when the lights turn red, and the postulated fan of continuous waves appearing when they turn green, in a space-time diagram. A typical vehicle path is shown as a broken line. The stationary wave which remains at the stopping point is that referred to in § 2; the flow across this point is $q_m$. Figure 15 shows also the 'weakening' of the shock wave when it is caught up by the fan; evidently, its speed must be rapidly reduced when the flow behind it begins to climb up the flow-concentration curve.

If the period of stoppage ('red period') is $T_r$, and the period of permitted flow $T_g$, then on the average the total number of vehicles $q_i(T_r + T_g)$ coming up (at the inflow rate $q_i$) during the complete cycle will pass across the stopping point during the time $T_g$ only if

$$q_i(T_r + T_g) < q_m T_g. \tag{13}$$

This sets an upper limit

$$\frac{T_g}{T_r + T_g} q_m \tag{14}$$

to the inflow (from the road in question) which can be handled by the controlled crossing, without leading to a queue of increasing length. This limit (14) is the 'capacity' of the controlled crossing, when regarded as a bottleneck.

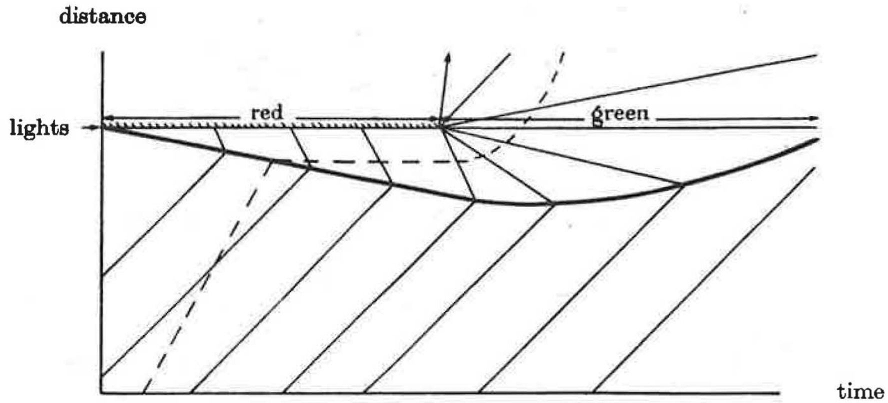



FIGURE 15. Uniform incoming flow stopped for a time and then started again.

If condition (13) is satisfied, that is if the inflow is less than the capacity, then the maximum flow $q_m$ at the stopping point cannot be maintained during the whole period $T_g$, but only for a reduced period of length $T_f$ (during which the crossing is running 'full') given by the equations

$$q_i(T_r + T_f) = q_m T_f, \quad T_f = \frac{q_i T_r}{q_m - q_i}. \tag{15}$$

After a time $T_f$, then, the flow ceases to be that carried by the wave which remains at the stopping point, and this must be because the shock wave in figure 15 has moved forward again and passed through the stopping point. This is illustrated in figure 16. Behind the shock wave the flow is the undisturbed inflow $q_i$. After passing through the stopping point it is just the ordinary shock wave in the rear of any traffic hump (§4).

A simple construction for the path of the shock wave is obtained as follows. The number of vehicles crossing a wave such as $OA$ in figure 16 (on which the flow is $q$ and the concentration $k$, and whose speed is $c$) is $(q - kc)\,t$, by §2, if $t$ is the time difference between $O$ and $A$. This number of vehicles must equal the number

$q_i(t+T_r)$ going up to the stopping point in the time $t+T_r$ since the stream was first stopped, minus the number $k_i ct$ left at time $t$ in the distance $ct$ between the stopping point and $A$.

Thus $$(q-kc)\,t = q_i(t+T_r) - k_i ct, \tag{16}$$

or $$t = \frac{q_i T_r}{q - q_i - c(k-k_i)}. \tag{17}$$

This equation can be used, with $x = ct$, to trace the shock-wave path on the $(x, t)$ diagram, if in both $k$ is varied from 0 to $k_j$, the corresponding values of $q$ and of $c = \mathrm{d}q/\mathrm{d}k$ being deduced from the flow-concentration curve. (Note that equation (15) is a special case of (17), with $t = T_f$, $c = 0$, $q = q_m$.)

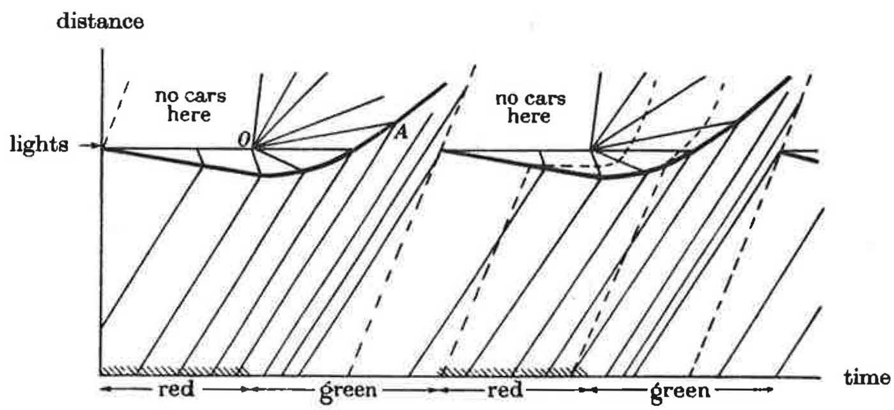



Figure 16. Wave pattern for traffic lights of capacity sufficient to admit the incoming flow.

If the incoming flow begins to exceed the capacity of the controlled junction, the shock waves of figure 16 do not get clear of it in time; each then collides with the shock wave sent out at the beginning of the next stopped period. They unite, and in turn collide with the next shock wave, and so on. If the excess incoming flow is maintained, these collisions (of shock waves, not cars!) must occur farther and farther back, and thus become a less and less significant feature of the situation. When this has happened the residual behaviour indicated by the theory is quite simple. Each shock wave (figure 17), on being formed at the stopping point, reduces the full flow $q_m$ to rest. As it moves backwards (e.g. at $G$) the traffic it stops is travelling more slowly. From $D$ onwards (figure 17) however, it does not reduce the flow completely to rest. Finally, at a very large distance behind the stopping point (e.g. at $A$) little reduction in vehicle speed occurs at shock waves and their effect has almost been ironed out into a typical bottleneck crawl.

The quantitative details of this familiar oscillating-speed crawl behind a choked controlled junction can be obtained by a device similar to that used above in the unchoked case. At a point in figure 17 such as $A$, a distance $x$ behind the stopping point, let the values of $k$, $q$ and $c$ carried by the waves which run into the shock wave at $A$ be distinguished by suffixes 1 (for the first) and 2 (for the second). Then the difference in the number of vehicles crossing the whole of each of these waves is equal to the flow across the stopping point during the time $T_g$; in symbols,

$$(q_1 - k_1 c_1)\frac{x}{(-c_1)} - (q_2 - k_2 c_2)\frac{x}{(-c_2)} = q_m T_g. \tag{18}$$

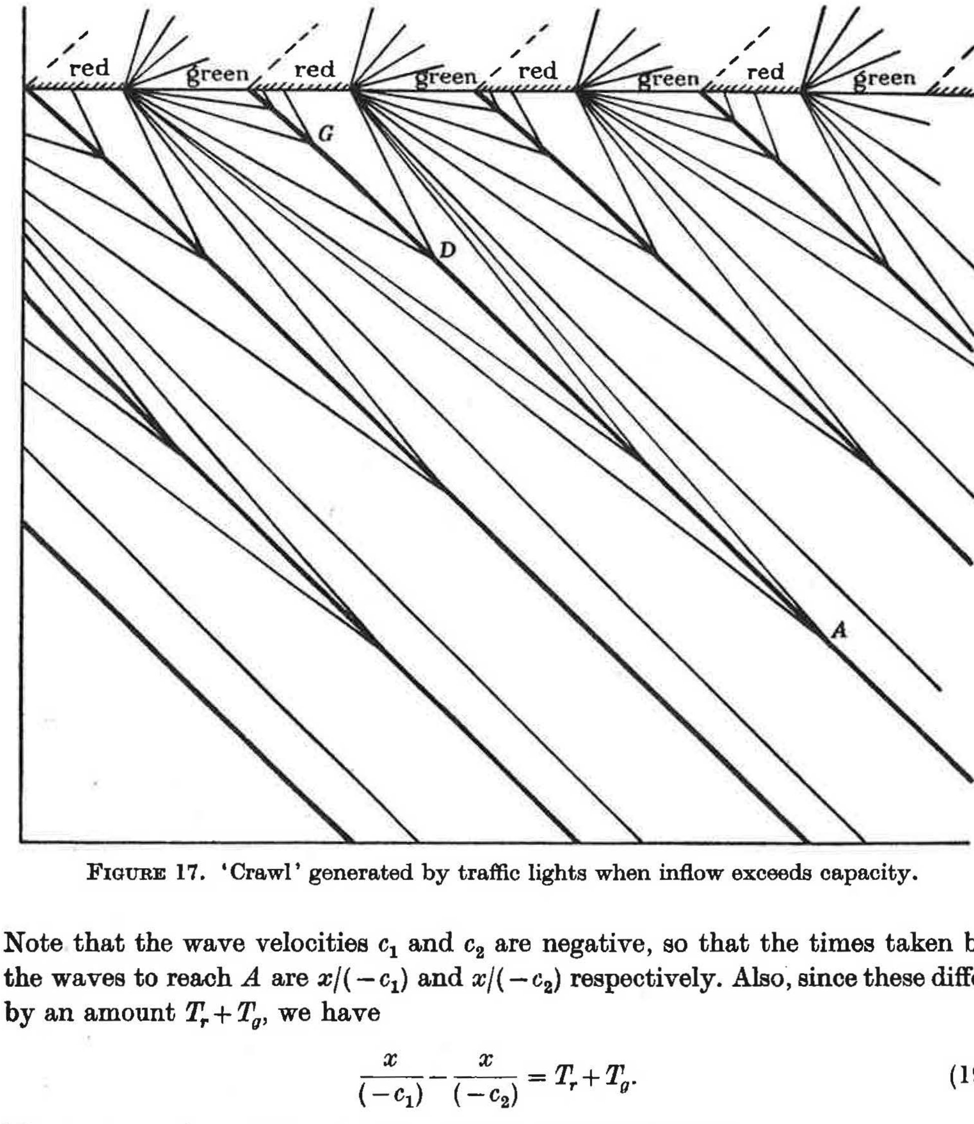



FIGURE 17. 'Crawl' generated by traffic lights when inflow exceeds capacity.

Note that the wave velocities $c_1$ and $c_2$ are negative, so that the times taken by the waves to reach $A$ are $x/(-c_1)$ and $x/(-c_2)$ respectively. Also, since these differ by an amount $T_r + T_g$, we have

$$\frac{x}{(-c_1)} - \frac{x}{(-c_2)} = T_r + T_g. \tag{19}$$

Eliminating $x$ from (18) and (19), we can write the result as

$$k_1 - \frac{1}{c_1}\left(q_1 - \frac{q_m T_g}{T_r + T_g}\right) = k_2 - \frac{1}{c_2}\left(q_2 - \frac{q_m T_g}{T_r + T_g}\right). \tag{20}$$

Geometrically, this means that a line drawn across the flow-concentration curve at a level corresponding to the capacity of the controlled junction (figure 18) will have the property that tangents drawn to the curve (e.g. $AB$ and $AC$, or $DE$ and $DF$) from points on the line have slopes equal to the slopes of waves meeting at points on the shock wave (e.g. $A$ or $D$ in figure 17) where there is a transition between the states represented by the two points of contact of the tangents (e.g. $B$ and $C$, or $E$ and $F$). When one of the tangents cannot be drawn, the point $F$ (where $q = 0$ and $k = k_j$) must be used as an end-point instead (see e.g. $GH$ and $GF$ corresponding to the point $G$ in figure 17).

The use of the theory on a small scale, which has been illustrated in the above discussion of the flow behind a controlled junction, is open to many objections, which are discussed below, one by one.

First, the time taken for each vehicle to accelerate to its desired speed is ignored, whereas it may not be negligibly small compared with the time scale of the process as a whole (say, with the period $T_g$ of permitted flow). This is especially true of the front vehicle, which is supposed to pursue a path at constant speed $v_F$, but

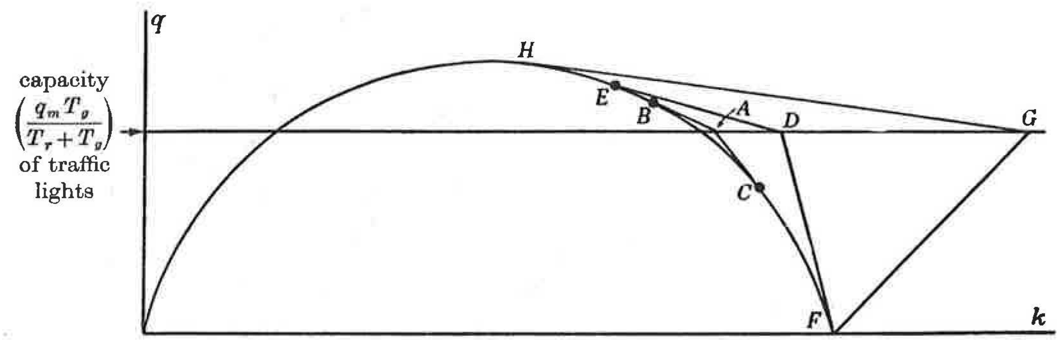



FIGURE 18. Geometrical construction for the flow of figure 13.

actually has to accelerate from rest up to this speed, by which time it is a certain distance, say $v_F T_l$, behind the path in question, and subsequently remains at such a distance behind it. This difficulty has been met in earlier, queue-theoretic, discussions of traffic light behaviour (Clayton 1941; Wardrop 1952) by regarding the vehicle as 'losing' a time $T_l$ after the lights have gone green. Its final path is that which it would have if it accelerated to speed $v_F$ *instantaneously* after an initial delay $T_l$. It is possible, therefore, that the present theory may be reasonably correct provided that the period of stoppage $T_r$ is taken to include this 'lost time' $T_l$, which must in turn be deducted from the period of permitted flow $T_g$. The 'lost time' is of the order of 5 to 10 s. The existence of this lost time is an important argument for keeping the periods of stoppage and permitted flow fairly long, so as to achieve a total flow at the junction as near to $q_m$ as possible. Conversely too great a period increases average vehicle delay, and Wardrop (1952) has shown that there is in any given case a cycle length which renders this average delay a minimum.

A second objection is that the theory ignores the fluctuations in inflow over times comparable with $T_g$ or $T_r$. It is just these fluctuations which lead to the phenomena (alternating quiet and 'busy' periods) studied in the theory of queues. Another way of phrasing the objection is to say that the times $T_g$ and $T_r$ are not large compared with the time $\tau$ needed (§2) to obtain smooth mean values of flow and concentration. This objection is certainly valid under relatively easy traffic conditions. It seems likely, however, that when the road is 'crowded' in the sense used in this paper the general picture of the starting flow given above may be relevant, the variations of inflow serving only to alter the positions of the shock waves at any instant.

A third objection is that certain measurements of traffic stopped and started indicate that under these conditions the mean concentration may be far less, and the mean speed far greater, for a given mean flow, than the values taken from the flow-concentration curve under more nearly steady conditions. Measurements on these lines known to the author include an unpublished set made at the Road Research Laboratory in 1954 and a study of flow in the road tunnel under the Meuse at Rotterdam (Aangenendt, Van Gils & Boost 1951). Both sets of results are summarized in figure 19, and a smooth curve drawn through all the points. Dr Smeed has suggested to the authors that the cause of the discrepancy might be variation in the acceleration of vehicles: if many vehicles in the queue cannot match the acceleration of those in front, the mean headway will exceed the minimum value tolerable under steady conditions. The present authors regard this and other causes mentioned above as serious limitations on the quantitative accuracy of their theory, but find the magnitude of the observed departures rather greater than they would expect from such a cause. The front vehicles certainly have an opportunity to accelerate very fast, which may not be allowed for adequately by the theory of 'lost time'; but in a long queue the acceleration required of the vehicles towards the rear is very moderate, and few of them can be incapable of it.

It must be remembered, on the other hand, that the mean flows and concentrations recorded in figure 19 were each measured *at a fixed point*, and according to the theory (see, for example, figure 16) the flow and concentration are changing so rapidly at a point that such a method can at most obtain an average of a large range of values. The method of measurement (§2) shows that in fact a time-average of each would be taken. It is easy to see that such an averaging process would in practice lead to a 'flow-concentration curve' somewhat like that of figure 19. Far behind the stopping point the mean would be taken over a period of very high values of concentration (and low values of flow) and (after the shock wave has passed forward) a longer period of very low values of concentration (and only moderate values of flow). The means would then be well in the left-hand half of the area under the true flow-concentration curve, and near the bottom. But, near the stopping point, there would be a longer period before the shock wave moves forward, and for much of this period the flow would be near its maximum. When averaged with a shorter period of low concentration this would give points in the middle of the area under the curve, somewhat below the top.

Future experiments will perhaps show whether or not this is the major cause of the discrepancy revealed in figure 19.* In the meantime, it is perhaps worth

* Very recent work by Wardrop has already gone some way towards confirming this. By taking means over very short distances (only twice the headway) and replotting the flow-concentration curve, he obtains points lying on a curve which is at least parallel (instead of perpendicular!) to the 'headway' curve of figure 19, although somewhat below it.

noting one or two directions in which the present theory could be improved in its application to small-scale flows. First, there is the 'blend with statistical ideas' suggested in § 1, but this is too difficult to be treated briefly, and the compounding of this blend is postponed to a later paper.

A second extension is to exclude a 'diffusion' effect due to the fact that each driver's gaze is concentrated on the road in front of him, so that he adjusts his speed to the concentration slightly ahead. This gives a dependence of flow on concentration gradient, which leads to an effective diffusion exactly as noted in part I, § 6. Such diffusion 'spreads out' the shock waves; in fact, drivers do not reduce speed instantaneously at shock waves, because they see them coming.

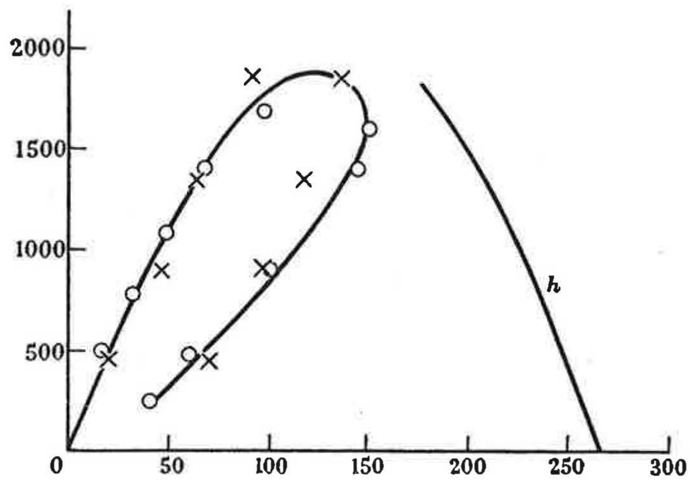



FIGURE 19. Observed mean flows and mean concentrations in traffic stopped and started. ○, Rotterdam tunnel; ×, Road Research Laboratory; *h*, 'headway' curve.

A third extension is to include an 'inertia' effect due to the fact that a driver must apply accelerator or brake to reach his desired speed and neither is instantaneously effective.

When both the last-named extensions have been applied, one reaches an equation of motion of a general form

$$\frac{\partial q}{\partial t} + c\frac{\partial q}{\partial x} + T\frac{\partial^2 q}{\partial t^2} - D\frac{\partial^2 q}{\partial x^2} = 0, \tag{21}$$

where $T$ is the inertial time constant for adjustments of speed, and $D$ is the diffusion coefficient, or decrement of flow for unit concentration-gradient. This is very similar to the equation governing waves in rivers (part I) when higher-order effects are taken into account. The new terms may be expected to introduce similar additional effects in traffic flow on a small scale to those found in certain river flows. In particular, something analogous to 'roll waves' might sometimes arise, in which a uniform flow is unstable and tends to degenerate into a succession of rapid accelerations and even more rapid retardations. This sometimes happens to a long convoy of vehicles which are expected to keep in line, even when the front vehicle maintains a uniform speed.

The behaviour of the flow behind a controlled junction could in principle be evaluated on the basis of an equation such as (21), but until the experimental information is clearer such extensions of the theory would seem to be premature.

## REFERENCES

Part I

1. Greenshields, B. D., "A Study of Traffic Capacity." *HRB Proc.*, 14: Pt. I, 448-474 (1934).
2. Richards, P. I., "Shock Waves on the Highways." *Oper. Res.*, 4: No. 1, 42-51 (1956).

Part II

1. Aangenendt, J. J. M., Gils, J. F. L., and Boost, A. G. M., "Studies of Traffic and of the Conditions Which Influence It." 9th Cong. of the Permanent Internat. Assoc. of Road Congresses, Sec. 2, Qun. III, Rep. No. 42, Paris (1951).
2. Charlesworth, G., "Methods of Making Traffic Surveys Especially 'Before and After' Studies." *Jour.* Inst. Hwy. Eng., 1: 42 (1950).
3. Clayton, A. J. H., "Road Traffic Calculations." *Jour.* Inst. of Civil Eng., 16: 247-284 (1941).
4. Glanville, W. H., "The Study of Road Traffic." Inst. Civil Eng., joint summary meeting with Inst. Civil Engineers (Ireland), pp. 41-68 (1949).
5. Glanville, W. H., "Road Safety and Road Research." *Jour.* Royal Soc. of Arts, 99: 114-192 (1951).
6. Glanville, W. H., "Road Research and Its Bearing on Road Transport." *Jour.* Inst. of Transp., 25: p. 186 (1953).
7. Greenshields, B. D., "A Study of Traffic Capacity." *HRB Proc.*, 14: 448-474 (1934).
8. Lighthill, M. J., and Whitham, G. B., "On Kinematic Waves. I. Flood Movement in Long Rivers." *Roy. Soc. Proc.* Ser. A, 229: 281-316 (1955).
9. Normann, O. K., "Results of Highway Capacity Studies." *Pub. Roads,* 23: 57-81 (June 1942).
10. Normann, O. K., and Taragin, A., "Resume of Previously Published Material on Highway Capacity." Bur. of Pub. Roads, Washington, D. C. (1942).
11. Normann, O. K., and Walker, W. P., "Highway Capacity, Practical Applications of Research." *Pub. Roads,* 25: 201-235 (Oct. 1949).
12. Smeed, R. J., "Traffic and Traffic Studies in the U. S. A." *Roads and Road Constr.,* 30: 138-142, 166-169 (1952).
13. Tanner, J. C., "A Problem of Interference Between Two Queues." *Biometrika,* 40: 58 (1953).
14. Wardrop, J. G., "Some Theoretical Aspects of Road Traffic Research." *Inst. of Civil Eng. Proc.,* Pt. II, 1: 325 (1952).
15. Wardrop, J. G., and Charlesworth, G., "A Method for Estimating Speed and Flow of Traffic From a Moving Vehicle." *Inst. of Civil Eng. Proc.,* Pt. II, 3: 158 (1954).
16. Whitham, G. B., "The Flow Pattern of a Supersonic Projectile." *Commun. on Pure and Appl. Math.,* 5: 301 (1952).

# Chapter 2

# CAR FOLLOWING AND ACCELERATION NOISE

E. W. Montroll
International Business Machines Corporation

and

R. B. Potts
University of Adelaide, Australia

## Contents

*Section*

# Chapter 2
# CAR FOLLOWING AND ACCELERATION NOISE

## 2.1 INTRODUCTION

Traffic phenomena are more a part of the behavioral than the physical sciences, for they result from the response of humans to various stimuli. Certain stimulus response equations can be analyzed, however, in the same manner that physicists analyze dynamic equations of motion.

The average speed or travel time for smooth safe driving on a given road depends on many phenomena (weather, mechanical condition of vehicles, driver behavior patterns, curves, hills, pedestrians, etc.). Two factors determine the maintenance of a smooth safe trip—the motion of an isolated vehicle and the interference of vehicles with each other.

Theoretically, traffic can be considered as the behavior of an assembly of vehicles which are influenced by their environment and by each other. Each vehicle is capable of either acceleration or deceleration. The "traffic problem" concerns the large-scale motions of these vehicles at high density. In this state they are forced to follow each other in lanes and they have only occasional opportunities to pass. Traffic theory in this regard then is the study of the acceleration and decleration patterns of these vehicles and the flows resulting when they are regulated in various ways.

## 2.2 THE ISOLATED VEHICLE

When a car is driven on an open road in the absence of traffic, the driver generally attempts, consciously or unconsciously, to maintain a rather uniform velocity, but he never quite succeeds. His acceleration pattern, as a function of time, has a random appearance. An acceleration distribution function can be easily obtained from such a pattern. This distribution is essentially normal. The random component of the acceleration pattern is called "acceleration noise" (*4, 5, 7*).

A measure of the smoothness or jerkiness of the driving is then given by the dispersion $\sigma$ of the acceleration noise. The mathematical definition of this quantity is

$$\sigma^2 = \frac{1}{T}\int_0^T [a\,(t)]^2\,dt \tag{2.1}$$

in which $a(t)$ is the acceleration (positive or negative) at time $t$, and $T$ is the total running time. Alternatively, if one considers that the acceleration is sampled at successive time intervals, $\Delta t$, then

$$\sigma^2 = \frac{1}{T}\sum [a\,(t)]^2\,\Delta t \tag{2.2}$$

The dispersion, or standard deviation, $\sigma$, is simply the root-mean-square of the acceleration, and it has the dimensions of acceleration. Its values are usually quoted in ft/sec$^2$ or as a fraction or multiple of $g = 32$ ft/sec$^2$.

Runs made on a section of the General Motors test track (an almost perfect roadbed) by four operators while driving in the range of 20 to 60 mph yielded normal acceleration noise distributions with standard deviates of $0.01g \pm 0.002g$, which are about 0.32 ft/sec$^2$. This dispersion increases at extreme speeds greater than 50 mph or less than 20 mph.

The acceleration noise of a driver will vary considerably as he drives on different roads or under different physiological or psychological conditions. The acceleration noise observed in a run in the Holland Tunnel of the New York Port Authority (with no traffic interference in the lane in which the run was made) was 0.73 ft/sec$^2$. Although the roadbed of the Holland Tunnel is quite good, the narrow lanes, artificial lighting and confined conditions induce a tension in a driver which is reflected in the doubling of his acceleration noise dispersion from its perfect road value. Preliminary studies of the acceleration noise associated with runs on poorly surfaced, winding country roads indicate that dispersions of 1.5 to 2 ft/sec$^2$ are not unusual.

Both transverse and longitudinal acceleration noises exist, but no measurement of the transverse (left-right) noise has been made. The latter would be large on winding roads and in the pattern of drivers who change lanes frequently while driving in heavy traffic. Both components of the noise would be large in the case of an intoxicated or fatigued driver or in situations in which the attention of the driver is shared between the road and his traveling companions. Noise measurements have not yet been made in these situations.

The dispersion of the acceleration noise of a vehicle was first measured by Herman *et al.* (*4*) by using an accelerometer to record on photographic film the car's acceleration as a function of time. From an analysis of the curve, the value of the dispersion $\sigma$ was determined. Although preliminary results were obtained by this method, the reduction of the data was rather tedious. Apparatus for automatically recording the acceleration in a form which can be converted to digital data suitable for computer input has been developed by Herman and his group. This apparatus enables accurate

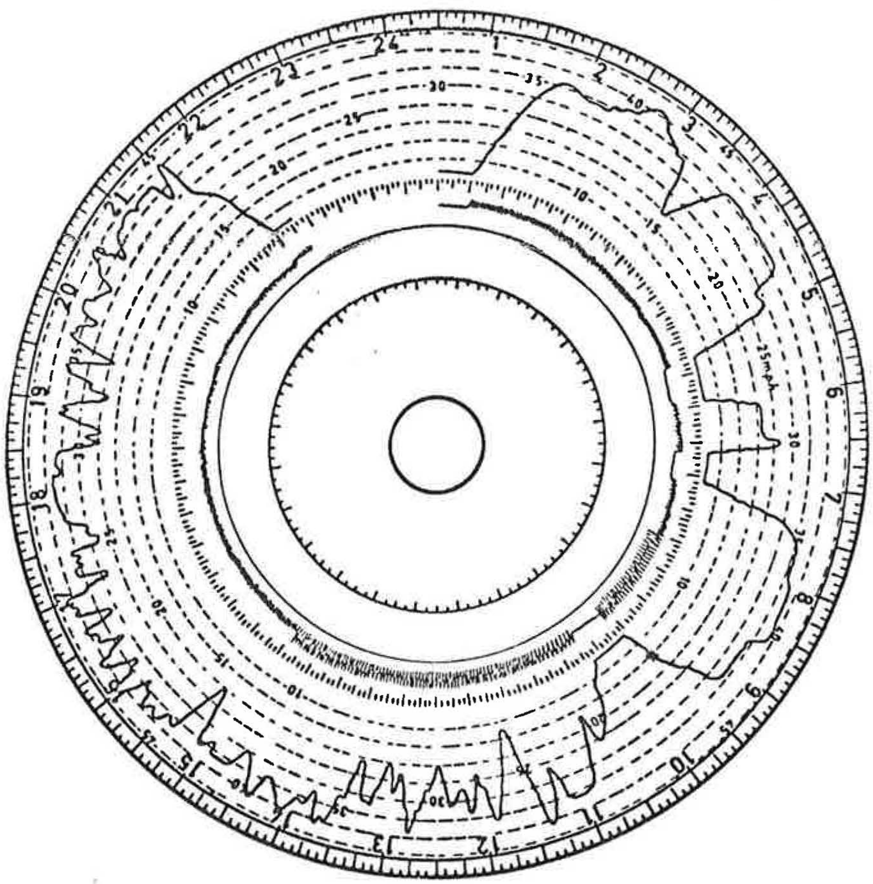


Figure 2.1. Sketch of a recording obtained on the circular chart of Kienzle TCO 8F tachograph. The concentric circles give the speed in mph; the scale on the outer circumference is in minutes. The inner trace is formed by an additional stylus whose mode of vibration is chosen by the driver by operating a key on the tachograph. The record illustrates a period of comparatively smooth driving with some stops (medium acceleration dispersion) followed by frequent accelerations and brakings (large acceleration dispersion).

Table 2.1 Value of $n^2/\Delta t$

| $\Delta t$ | $n = 1$ | $n = 2$ | $n = 3$ | $n = 4$ | $n = 5$ | $n = 6$ | $n = 7$ | $n = 8$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 4.00 | 9.00 | 16.00 | 25.00 | 36.00 | 49.00 | 64.00 |
| 2 | 0.50 | 2.00 | 4.50 | 8.00 | 12.50 | 18.00 | 24.50 | 32.00 |
| 3 | 0.33 | 1.33 | 3.00 | 5.33 | 8.33 | 12.00 | 16.33 | 21.33 |
| 4 | 0.25 | 1.00 | 2.25 | 4.00 | 6.25 | 9.00 | 12.25 | 16.00 |
| 5 | 0.20 | 0.80 | 1.80 | 3.20 | 5.00 | 7.20 | 9.80 | 12.80 |
| 6 | 0.17 | 0.67 | 1.50 | 2.67 | 4.17 | 6.00 | 8.17 | 10.67 |
| 7 | 0.14 | 0.57 | 1.29 | 2.29 | 3.57 | 5.14 | 7.00 | 9.15 |
| 8 | 0.13 | 0.50 | 1.13 | 2.00 | 3.13 | 4.50 | 6.13 | 8.00 |
| 9 | 0.11 | 0.44 | 1.00 | 1.78 | 2.78 | 4.00 | 5.44 | 7.11 |
| 10 | 0.10 | 0.40 | 0.90 | 1.50 | 2.50 | 3.60 | 4.90 | 6.40 |
| 11 | 0.09 | 0.36 | 0.82 | 1.45 | 2.27 | 3.27 | 4.45 | 5.82 |
| 12 | 0.08 | 0.33 | 0.75 | 1.33 | 2.08 | 3.00 | 4.08 | 5.33 |
| 13 | 0.08 | 0.31 | 0.69 | 1.23 | 1.92 | 2.77 | 3.77 | 4.92 |
| 14 | 0.07 | 0.29 | 0.64 | 1.14 | 1.79 | 2.57 | 3.50 | 4.57 |
| 15 | 0.07 | 0.27 | 0.60 | 1.07 | 1.67 | 2.40 | 3.27 | 4.27 |
| 16 | 0.06 | 0.25 | 0.56 | 1.00 | 1.56 | 2.25 | 3.06 | 4.00 |
| 17 | 0.06 | 0.24 | 0.53 | 0.94 | 1.47 | 2.12 | 2.88 | 3.76 |
| 18 | 0.06 | 0.22 | 0.50 | 0.89 | 1.39 | 2.00 | 2.72 | 3.56 |
| 19 | 0.05 | 0.21 | 0.47 | 0.84 | 1.32 | 1.89 | 2.58 | 3.37 |
| 20 | 0.05 | 0.20 | 0.45 | 0.80 | 1.25 | 1.80 | 2.45 | 3.20 |

estimations of the acceleration dispersion.

An inexpensive and simple method of estimating the dispersion employs the Kienzle TCO8F model tachograph* with a speed recording range of 0 to 45 mph. The speed is recorded by a stylus on a circular chart which revolves once in 24 min. A typical record is shown in Figure 2.1. The inner trace is formed by an additional stylus that can vibrate in any of three modes of vibration. The choice of the mode is decided by the position of a tachograph key which can be operated by the driver. It enables him to indicate when he passes selected points on the highway. A stylus for recording distance traveled was not used, as the mileometer on the tachograph was more suitable and accurate.

Inasmuch as times are proportional to angles on the circular chart, it is a simple matter to use a protractor to measure the travel time TT, the stopped time ST and the running time RT (= TT − ST). A special analyzer is available from the tachograph manufacturers which allows the record to be mounted on a protractor and viewed through a magnifying glass. The acceleration dispersion was estimated by approximating Eq. 2.1 by

$$\sigma^2 \approx \frac{1}{T} \sum \left( \frac{\Delta u}{\Delta t} \right)^2 \Delta t \qquad (2.3)$$

or

$$\sigma^2 = \frac{(\Delta u)^2}{T} \sum \frac{n^2}{\Delta t} \qquad (2.4)$$

in which $\Delta t$ is the time taken for a change $n\,\Delta u$ in speed, $n$ being an integer and $\Delta u$ a small speed interval taken constant throughout the measurement of the record. The time $T$ is taken as the running time RT and not the travel time. For a chart recording speeds in the 0- to 45-mph range, a value $\Delta u = 2.5$ mph proved to be convenient. The record is first marked at speed intervals $n\,\Delta u$, as indicated at the beginning of the record in Figure 2.1. The chart is then placed in an analyzer and successive values of $\Delta t$ are measured. It is convenient to use a table of values of $n^2/\Delta t$ (such as Table 2.1) to enable the value of $n^2/\Delta t$ to be cal-

* The Kienzle tachograph is distributed under the name ARGO in the United States. Other tachographs are manufactured by VDO and Wagner. Various models are available. Some have circular charts; others use paper wound on spools. Models with slow-moving charts are used by trucking and bus companies. Those with fast-moving charts are ideal for many traffic engineering purposes.

culated progressively on a desk calculator. To illustrate the method, the values of $n$, $\Delta t$ and $n^2/\Delta t$ for the beginning of the marked record in Figure 2.1 are:

| $n$ | $\Delta t$ | $n^2/\Delta t$ |
|---|---|---|
| 8 | 20 | 3.20 |
| 4 | 15 | 1.07 |
| 2 | 10 | 0.40 |
| 0 | — | 0 |
| 1 | 16 | 0.06 |
| 1 | 6 | 0.17 |
| 0 | — | 0 |
| 2 | 14 | 0.29 |

If $\Delta t$ is in seconds, the running time $T$ in seconds, and $\Delta u = 2.5$ mph, then $(\Delta u)^2 = \left(2.5 \times \frac{22}{15}\right)^2 \approx 13.44$ ft$^2$/sec$^2$, which when inserted in Eq. 2.4 gives $\sigma$ in ft/sec$^2$.

Some of the advantages of this method of measuring the acceleration dispersion are:

1. The equipment is inexpensive.
2. The chart requires no processing.
3. The chart forms a convenient permanent record of the test run.
4. Travel times, stopped times and running times are easily measured from the chart.
5. Small speed fluctuations are ignored.

The main disadvantages are:

1. Each record takes up to 30 min to analyze.
2. The accuracy of the determination of $\sigma$ is only about 10 percent.

It must be emphasized that the acceleration dispersion $\sigma$ is suggested as a useful traffic parameter, enabling the comparison of different traffic situations. Although the error is about 10 percent, the estimated value is consistently smaller than the exact value because the subdivision of the record into speed intervals, which are multiples of $\Delta u$, essentially replaces the speed-time curve by a set of linear segments. In any case the many factors contributing to the acceleration noise denote that its dispersion varies from run to run, and the usual care must be taken to design a set of experiments with a sufficient number of runs so that significant statistical tests can be made on the results.

## 2.3 LAW OF CAR FOLLOWING AND VARIATION OF FLOW WITH DENSITY

In this section the effect of the road is neglected, and consideration is given only to the interaction between cars. Consider a line of traffic so dense that passing is impossible and the driver of each vehicle is forced to drive slower than he would on his own volition. Also suppose that the road is excellent, so that the acceleration pattern of each vehicle depends more on the behavior of its predecessor than on its own natural acceleration noise.

The acceleration of the $n$th vehicle at time $t$ can be expected to depend on various relative characteristics of the $(n-1)$st and the $n$th vehicles. Some of these characteristics are relative velocity and separation distance. The manner in which one vehicle follows another is referred to as the *law of following* (*1*, *7*).

Several qualitative features of the law are self-evident. First, *a moving line of traffic must not amplify small disturbances.* That is, if the first vehicle in the line slows down slightly and then speeds up to his old rate, this slight perturbation must not be amplified as it is transmitted down the line to the extent that a collision occurs far behind the point of perturbation or that the cars sufficiently far back must stop to avoid collisions. Secondly, the law of following must not be such that a strong perturbation such as a sudden stop cannot sometime cause a rear-end collision, for such collisions occur rather frequently. *Responses are never instantaneous.* A certain time $t_1$ is required for a driver to notice that his relative speed and separation distance with his predecessor have changed. A time $t_2$ is required to decide on the proper response to a variation. A time $t_3$ is required for the vehicle to act on the response. In practice, $t_1 + t_2 + t_3$ is about 1.5 sec.

As a standard from which perturbations are to be measured, consider a hypothetical line of traffic moving at constant velocity $u$ with all cars separated by a distance $s$ (= distance from the front bumper of one car to the front bumper of the car behind it). The traffic in Figure 2.2 is postulated to be moving to the right and $X_n(t)$ is the position of the $n$th car at the time $t$. Then, if the origin is chosen as the location of the front bumper of the first car at time $t = 0$,

$$X_n(t) = u\,t - (n-1)s \qquad (2.5)$$

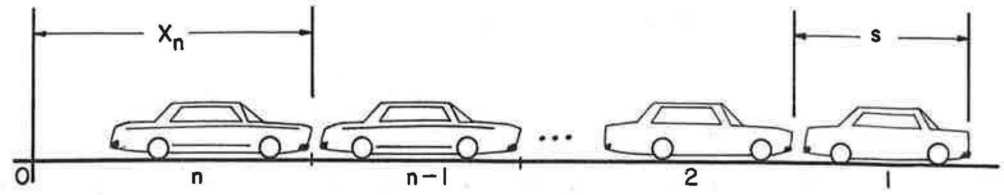



Figure 2.2. Postulation of moving vehicles.

Of course, cars in a real stream of traffic do not move with constant velocity, nor is the separation distance fixed. Let $x_n(t)$ be the deviation from $X_n(t)$ of the location of the $n$th car at time $t$ and let $y_n(t)$ be the actual location. Then

$$y_n(t) = x_n(t) + X_n(t) \tag{2.6}$$

and the velocity of the $n$th car is

$$u_n(t) = \dot{x}_n(t) + u \tag{2.7}$$

while the acceleration is

$$a_n(t) = \ddot{y}_n = \ddot{x}_n \tag{2.8}$$

Suppose that the line of traffic flows almost in the described manner so that $x_n(t)$ and $\dot{x}_n(t)$ are very small for all $n$. By accelerating and decelerating, each driver makes small compensations to arrive at the steady-stream velocity and spacing. Now examine several possible laws of following to see which might be realistic and then try to compare with experimental results. In the limit of very small $x_j(t)$ and $\dot{x}_j(t)$, three possibilities might exist:

(a) The $n$th driver accelerates or decelerates by an amount proportional to the deviation in relative separation from the desired amount $s$. That is,

$$a_n(t) = \ddot{x}_n(t) = \mu\,[x_{n-1}(t) - x_n(t)] \tag{2.9}$$

in which the parameter $\mu$ would be determined from observations on relative motions of cars in the traffic stream.

(b) The $n$th driver accelerates by an amount proportional to the difference in relative velocity of $n$th and $(n-1)$st cars, giving

$$a_n(t) = \dot{u}_n(t) = \alpha\,[u_{n-1}(t) - u_n(t)] \tag{2.10}$$

If the $(n-1)$st car is moving faster than the $n$th, the $n$th driver accelerates to compensate and reduce velocity differences and vice versa when $u_{n-1}(t) < u_n(t)$, the parameter $\alpha$ being chosen to be positive.

(c) A linear combination of the previous two laws:

$$a_n(t) = \mu\,[x_{n-1}(t) - x_n(t)] + \alpha\,[u_{n-1}(t) - u_n(t)] \tag{2.11}$$

All these laws are *linear* laws, which might be appropriate only for *small* deviations from the desired state of traffic. The response of the $n$th driver is proportional to a deviation for which he wishes to compensate. The parameters $\alpha$ and $\mu$ are called sensitivities of the response to the deviations. Large values of $\alpha$ and $\mu$ correspond to strong compensation, and small values correspond to weak compensations. Experiments have been performed to determine whether these possibilities are sensible. Before resorting to experimental evidence, however, determine if any of these laws can be ruled out on the basis that they violate the requirement that a line of traffic must not be an amplifier of small disturbances.

A standard way to investigate the effect of disturbances and of stability of linear systems is to make a harmonic (frequency) analysis of the disturbance to see how individual frequency components are propagated through the system. Assuming that the deviation of the motion of the lead car in a platoon is the source of the disturbance, its motion can then be harmonically analyzed. When this is done in law (a), it turns out that a resonance exists at frequency $\omega = \mu^{1/2}$. That is, any frequency components at frequencies near $\mu^{1/2}$ are amplified strongly by the traffic, the law of amplification of the amplitude of the $\omega$ component being $[1 - \omega^2/\mu]^{-n}$. On the other hand, law (b) damps out a disturbance as

$$[1 + \omega^2/\alpha^2]^{-n} \qquad (2.12)$$

for the $n$th car behind the source of the disturbance. Hence, law (b) is a reasonable one to investigate further while law (a) is not. If one investigates mixed laws such as (c) or any other law in which the acceleration is proportional to the difference in $i$th derivatives of the separation distance between two successive vehicles, he finds resonances (instabilities) in those laws which contain terms with even values of $i$. Inasmuch as it is doubtful that a driver could be sensitive to third derivatives, one is left with only law (b) as a possible one for investigation.

Before law (b) is compared with experimental data, additional features of the law must be examined. It will be recalled from the discussion at the beginning of this section that responses are never instantaneous. Even though law (b) may be suggestive for further consideration, it should be amended to take into account the time lag between the time of the actual development of a disturbance and the moment of effective response; therefore, law (b) should now read

$$a_n\,(t + \Delta) = \dot{u}_n\,(t + \Delta) = \alpha\,[u_{n-1}\,(t) - u_n\,(t)] \qquad (2.13)$$

in which the velocities on the right side are to be taken at time $t$ to influence the acceleration of the left at time $t + \Delta$. When time lags are incorporated into linear systems, instabilities may result. If one reacts too strongly (large $\alpha$) to an event which occurred too far in the distant past (large response lag $\Delta$), the situation at the moment of response may have changed to the point where the response is actually in the wrong direction. Hence, when lags are long there should be weak responses to insure stability. In fact, a line of traffic following Eq. 2.13 is stable, not amplifying small disturbances, only when

$$2\,\alpha\,\Delta < 1 \qquad (2.14)$$

A disturbance of unit amplitude is propagated back to the $n$th car so that its amplitude at arrival is equal to or less than

$$[1 + (\omega^2/\alpha^2)\,(1 - 2\,\alpha\,\Delta)]^{-n} \qquad (2.15)$$

It should be noted that a resonance appears when Eq. 2.14 is violated.

Before comparing Eq. 2.13 with experimental data, it is worth trying to extend the formula slightly so that it is applicable to cases in which, for some reason, rather large gaps have formed between cars. Clearly, when the separation distance is large one will not drive as sensitively as he would in a bumper-to-bumper situation. Hence $\alpha$ should depend on the separation distance in such a way that when two successive vehicles are separated by an enormous distance no interaction exists between them at all. One possible law is that the sensitivity $\alpha_0$ should be inversely proportional to the car spacing (distance between cars plus car length) so that

$$a_n\,(t + \Delta) = \ddot{y}_n\,(t + \Delta) = \alpha_0 \left\{ \frac{\dot{y}_{n-1}\,(t) - \dot{y}_n\,(t)}{y_{n-1}\,(t) - y_n\,(t)} \right\} \qquad (2.16)$$

in which $\alpha_0$ is a measure of sensitivity.

A number of car-following experiments were performed on the General Motors test track, as well as in the Holland and Lincoln Tunnels in New York. Each of a number of drivers using an instrumented car was told to follow a lead car as he would in normal city driving. In each case a continuous record was taken of the acceleration of the second car $a(t)$, as well as the relative velocities $u_r\,(t)$ and spacing $s(t)$ of the two cars. For each driver a best value of $\alpha$ and $\Delta$ was obtained in

$$a(t + \Delta) = \alpha_0\,[u_r\,(t)/s(t)] \qquad (2.17)$$

which is equivalent to Eq. 2.16 so that

$$\sum_t\,[a(t + \Delta) - \alpha_0 u_r(t)/s(t)]^2 = \min \qquad (2.18)$$

The results (5) of the car-following experiments are summarized in Table 2.2. The correlation coefficients for the best values of $\alpha_0$ and $\Delta$ were usually greater than 0.9, and for some drivers as high as 0.97. If Eq. 2.16 were exact and no experimental error existed in the data, the correlation coefficients would be 1. Some deviation from 1 must be expected because the acceleration noise contribution to $a(t)$ has been omitted. There is some variation in the values of $\alpha_0$

Table 2.2 Summary of Car-Following Experiments

| Locality | Number of Drivers | $\alpha_0$ (mph) | $\Delta$ (sec) |
|---|---|---|---|
| General Motors test track | 8 | 27.4 | 1.5 |
| Holland Tunnel | 10 | 18.2 | 1.4 |
| Lincoln Tunnel | 16 | 20.3 | 1.2 |

and $\Delta$ for different drivers. For example, in the General Motors test track experiments, $\Delta$ varied from 1.0 to 2.2 sec, with one-half the drivers having $\Delta$ values between 1.4 and 1.7. It would be interesting to find these constants on a given road for a large number of drivers. This would enable one to obtain reliable statistics on personal variations between drivers. In applying Eq. 2.16 to a line of traffic, it is assumed that all drivers have the same characteristics; namely, the average ones.

An interesting consequence of the law of following (Eq. 2.13) is that the formula for the rate of propagation of a disturbance down a line of traffic (in cars per second) is $n/t = \alpha$.

Although a line of traffic is stable to small perturbations, it is well known that most rear-end collisions are due to local instabilities in which one or more cars are unable to compensate for large disturbances ahead of them. It can be shown that no such local instabilities would occur in the law of following if the inequality $\alpha e \Delta < 1$ were satisfied, a condition rarely exhibited in follow-the-leader experiments.

Although Eq. 2.17 was derived to form a basis for the law of following of one vehicle by another, it can also be employed to relate the flow rate of single-lane traffic to the traffic density (*3*). The flow rate $q$ (say in vehicles per hour) is the product of the density $k$ (cars per mile) and the velocity $u$ (miles per hour). Thus, $q = u\,k$. Qualitatively the equation of state of the traffic, the name given to the flow-versus-density relation, can be expected to have the form given in Figure 2.3. When there are no cars on the road ($k = 0$) the flow rate is zero. At close packing (bumper to bumper) where $k = k_j$, the density is greatest, but no cars

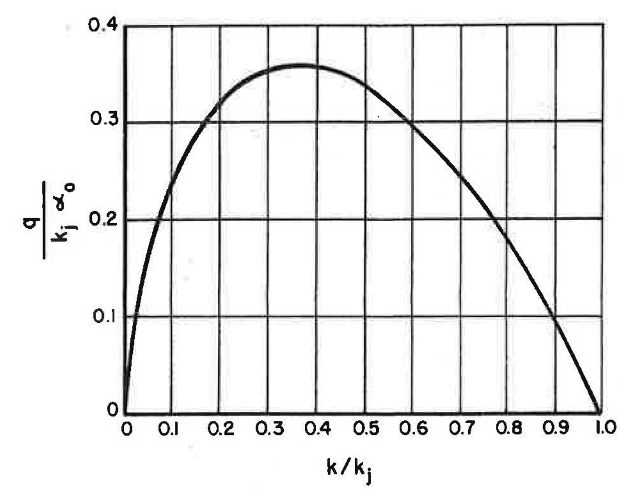



Figure 2.3. Normalized traffic flow versus density as obtained from Eq. 2.21. Curve compares with data obtained by Greenberg from experiments in the Lincoln Tunnel.

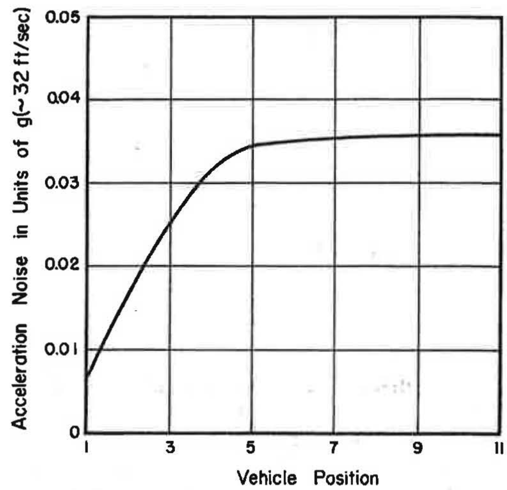



Figure 2.4. Acceleration noise of vehicles at different locations in a platoon.

can move ($u = 0$). At some intermediate density, a maximum flow rate exists.

Eq. 2.16 can be integrated to yield

$$u_n(t+\Delta) - u'_n = \alpha_0 \log [y_{n-1}(t) - y_n(t)]/s'_n \qquad (2.19)$$

in which $s'_n = y_{n-1} - y_n$ at time when velocities are $u'_n$. Now choose $s'_n$ to be the close packing bumper-to-bumper distance. Because there is no motion under this condition, $u'_n = 0$ and

$$u_n(t+\Delta) = \alpha_0 \log [y_{n-1}(t) - y_n(t)]/s'_n \qquad (2.20)$$

Now suppose that the traffic flow has become steady. Then its average velocity at time $t$ is about the same as that at $t + \Delta$ ($\Delta$ being about 1.5 sec). Therefore, $u_n(t+\Delta)$ can be replaced by the average velocity $u$, and $[y_{n-1}(t) - y_n(t)]$ can be replaced by the average separation distance, which is the reciprocal of the average density, $k^{-1}$. Actually $u$ is the arithmetic mean velocity and $k$ the geometric mean density. Hence,

$$u = \alpha_0 \log_e k_j/k \qquad (2.21)$$

in which $k_j$ is the density at close packing ($k_j = 1/s'_n$).

The flow rate $q$ is then given by

$$q = u\,k = k\,\alpha_0 \log_e k_j/k \qquad (2.22)$$

This function, plotted in Figure 2.3 compares with experimental data taken in the Lincoln Tunnel in New York. From a large sample of more than 24,000 vehicles (*7*) in the Holland Tunnel, the best fit value of $\alpha_0$ was found to be 18.95 mph, which is to be compared with $\alpha_0 = 18.2$ mph obtained in car-following experiments in the same tunnel (Table 2.2). This provides a good check for the theory.

Notice that $\alpha_0$ is the velocity which gives a maximum flow rate. It has been observed that $\alpha_0$ is small under hazardous driving conditions, such as poor lighting or narrow roadway with two lanes in tunnels, whereas it is large on good roads such as freeways with no turns. Because the expensive parts of a highway system, such as bridges and tunnels, are frequently its bottlenecks, the traffic engineer should make $\alpha_0$ as large as possible to increase the maximum possible flow rate and to regulate traffic so that for a given $\alpha_0$ this maximum is achieved.

## 2.4 ACCELERATION NOISE OF A VEHICLE IN TRAFFIC

In Section 2.2, the acceleration noise of an isolated vehicle was discussed. In Section 2.3, several simple car-following laws for traffic in the absence of acceleration noise were exhibited. Clearly, the total acceleration noise of a vehicle in traffic is a superposition of its natural noise and its response to that of its predecessors through the law of following. In stable, smooth-flowing traffic the effect of the natural noise of a given vehicle dies out as it is propagated down the line. The total acceleration noise of vehicles at different locations in a platoon has been measured by Herman and Rothery (*6*) (see Fig. 2.4). It is noted that traffic has broadened the acceleration distribution function so that the dispersion far down the platoon is about three times that of the lead car, which is effectively moving freely on the road. Figure 2.4 also shows that in the absence of any violent disturbances the influence of the noise of a single vehicle is dampened out by the time the signal of its motion has propagated down to the fifth or sixth car behind it. Traffic broadens the acceleration distribution, the broadening being smaller for the conservative driver who is satisfied to follow the stream than for the "cowboy" who by weaving attempts to drive 5 to 10 mph faster than the stream. This is shown in Figure 2.5 for traffic on Woodward Avenue in Detroit (*4*).

The traffic broadening is not large for smoothly flowing traffic, but the dispersion increases rapidly at the onset of congestion. For stop-and-go traffic the dispersion is small because cars are unable to accelerate to appreciable speeds.

The broadening of the acceleration distribution by traffic depends on the parameters of the law of following. The acceleration of the $n$th car at time $t$ is a superposition of its natural acceleration noise and its response to the motion of its predecessor. In smoothly moving traffic the separation distance varies only slightly from the equilibrium distance $s$. Hence, Eq. 2.16 can be linearized so that addition of the natural acceleration $\beta\ (t)$ gives

$$\dot{u}_n\ (t + \Delta) = \alpha\ [u_{n-1}\ (t) - u_n\ (t)] + \beta\ (t) \tag{2.23}$$

in which

$$\alpha = \alpha_0 / s \tag{2.24}$$

The $\beta\ (t)$ is a random function whose value at time $t$ is not specified. It is determined by its distribution function $f(a)$ so that $f(a)\ da$ is the probability that $\beta\ (t)$ has a value between $a$ and $a + da$ at time $t$. For simplicity, assume that $\beta\ (t)$ has the same distribution for all drivers on the road of interest. One can use the standard methods of the theory of Brownian motion to determine the statistical differences of properties of $a_n\ (t) = \dot{u}_n\ (t)$ from those of $\beta\ (t)$ in terms of $\alpha$ and $\Delta$. If the acceleration noise is peaked in the low frequency range, one finds that the dispersion $\sigma$ of the distribution function of $a_n\ (t)$ (as $n \rightarrow \infty$; *i.e.*, for cars far from the beginning of a platoon) is related to the dispersion $\sigma_0$ of $\beta\ (t)$ by

$$\sigma = \sigma_0 / (1 - 2\,\alpha\,\Delta)^{1/2} \quad \text{if } 2\,\alpha\,\Delta < 1 \tag{2.25}$$

The stability condition (Eq. 2.14) again makes its appearance. The closer the traffic reaches the limit of stability ($2\,\alpha\,\Delta \rightarrow 1$) the larger the traffic broadening of the acceleration noise.

If Eq. 2.24 is substituted in Eq. 2.25, the average spacing may be expressed as

$$s = 2\,\alpha_0\,\Delta / [1 - (\sigma_0/\sigma)^2] \tag{2.26}$$

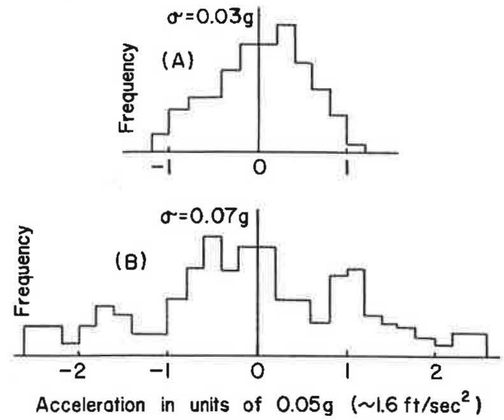



**Figure 2.5.** Acceleration distribution functions for a driver (A) moving with a traffic stream at approximately 35 mph and (B) attempting to drive 5 to 10 mph faster than the stream average.

This equation was checked with the Holland Tunnel observations of Herman, Potts and Rothery. The traffic broadening of the acceleration noise dispersions $\sigma/\sigma_0$ in the tunnel varied from about 1.50 to 1.75, depending on the density during the experiment. The value of $\alpha_0$ was determined by fitting Eq. 2.15 to the observed flow-versus-density curve for the tunnel. The average time lag of 1.5 sec, which was observed in car-following experiments, was substituted in Eq. 2.26, as was the observed ratio $\sigma/\sigma_0$. The computed values of $s$ were then converted into appropriate densities ($s = 1/k$), which were compared with the observed densities made at the same time as $\sigma/\sigma_0$ was determined. These calculated values generally did not deviate from the measured ones by more than 10 or 15 percent.

## REFERENCES

1. Chandler, R. E., Herman, H., and Montroll, E. W., "Traffic Dynamics: Studies in Car Following." *Oper. Res.*, 6: 2, 165-184 (1959).
2. Edie, L. C., Foote, R. S., Herman, R., and Rothery, R. W., "Analysis of Single Lane Traffic Flow." *Traffic Eng.*, pp. 21-27 (Jan. 1963).
3. Gazis, D. C., Herman, R., and Potts, R. B., "Car-Following Theory of Steady-State Traffic Flow." *Oper. Res.*, 7: 4, 499-505 (1959).
4. Herman, R., Montroll, E. W., Potts,

R. B., and Rothery, R. W., "Traffic Dynamics: Analysis of Stability in Car Following." *Oper. Res.*, 7: 6, 86-106 (1959).
5. Herman, R., and Potts, R. B., "Single-Lane Traffic Theory and Experiment." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 120-146 (1961).
6. Herman, R., and Rothery, R. W., "Microscopic and Macroscopic Aspects of Single Lane Traffic Flow." *Oper. Res.*, Japan (in press).
7. Jones, T. R., and Potts, R. B., "The Measurement of Acceleration Noise—A Traffic Parameter." *Oper. Res.*, 10: 6, 745-763 (1962).
8. Montroll, E. W., "Acceleration Noise and Clustering Tendency of Vehicular Traffic." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 147-157 (1961).

# Chapter 3

# QUEUEING THEORY APPROACHES

D. E. Cleveland
Texas A&M University

and

D. G. Capelle
Automotive Safety Foundation

## Contents

# 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

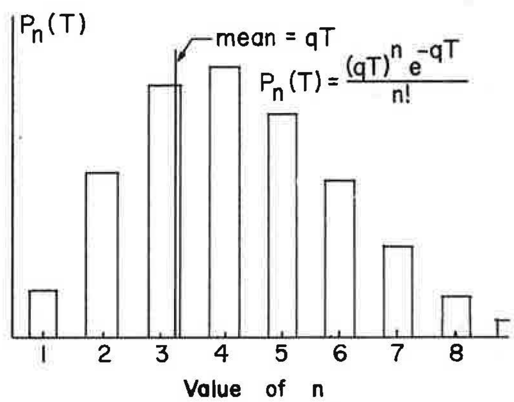



Figure 3.1. Poisson distribution.

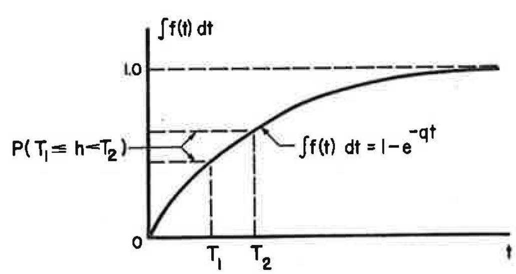



Figure 3.2. Cumulative exponential distribution.

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),

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

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

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

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

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

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

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

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):

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

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

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

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

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

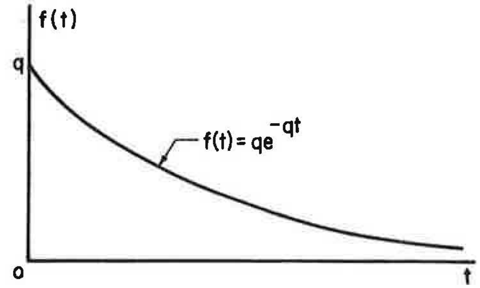



Figure 3.3. Exponential distribution.

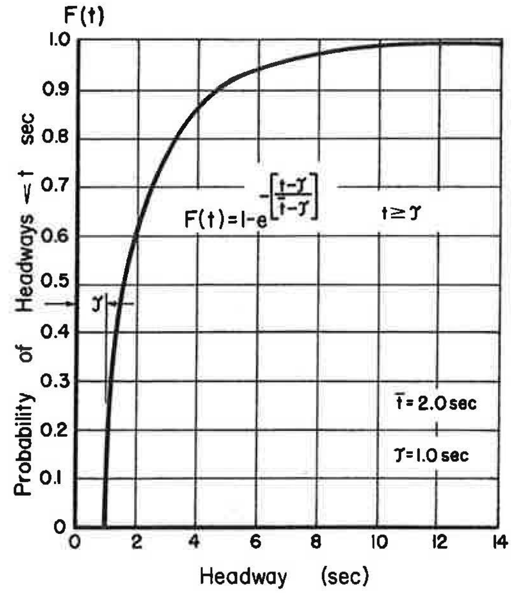



Figure 3.4. Shifted exponential distribution.

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

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 $\tau$. This, in effect, states that the probability of a gap between successive vehicles of less than $\tau$ 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, $\tau_2$ is the minimum headway of constrained vehicles, and $\alpha$ 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 $\tau_1$ 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, $\tau$, should be considered improbable, but not impossible. The exponential and hyper-exponential distributions, on the other hand, represent curves which find their maximum probabil-

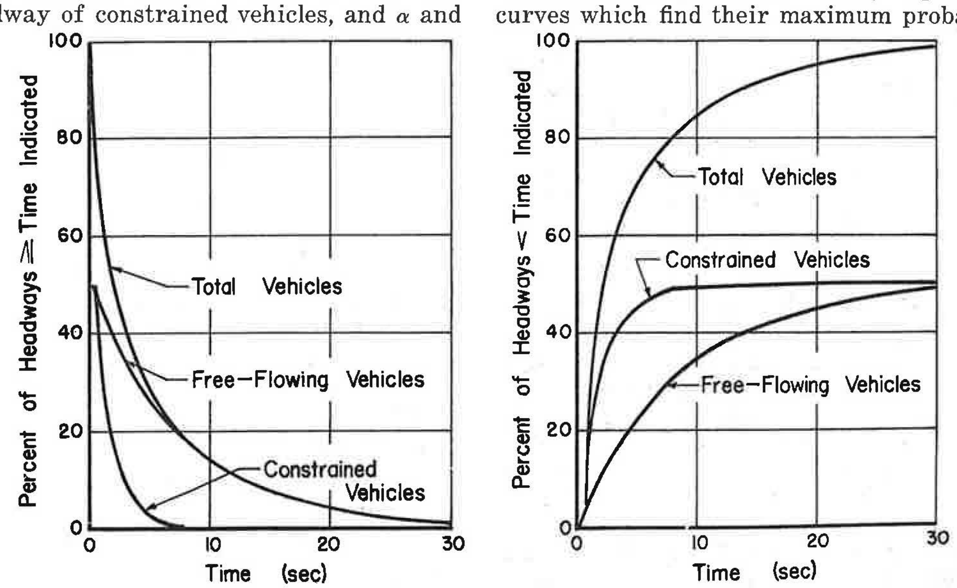



Figure 3.5. Schuhl's composite exponential distribution.

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

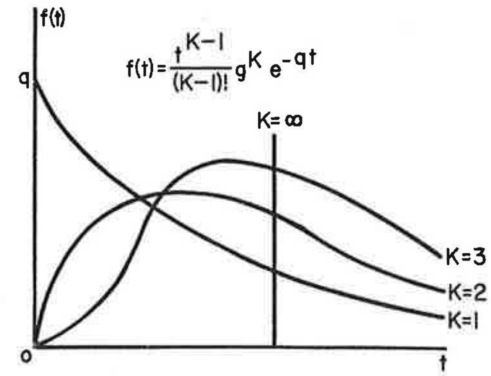



Figure 3.6. The Erlang gap distribution.

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 $\bar{t}$,

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

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

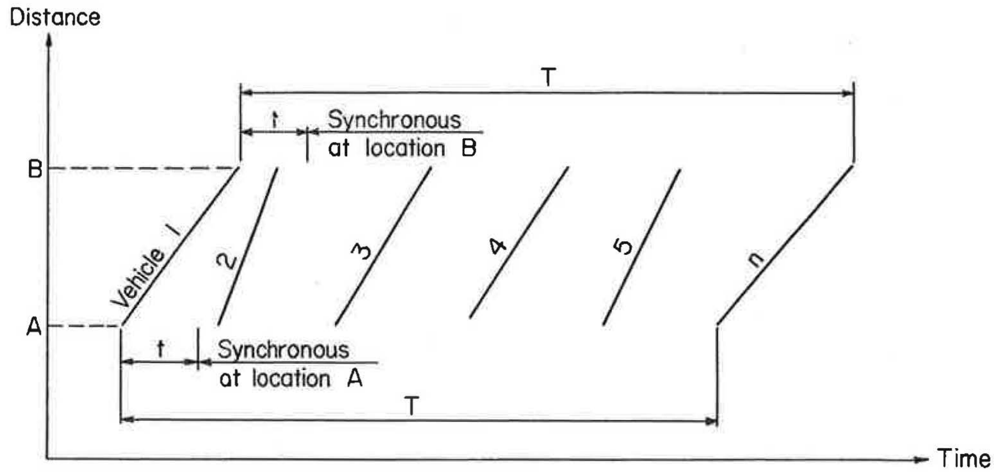



**Figure 3.7.** Time-space diagram illustrating synchronous counting procedures at two locations: A, regular arrivals, and B, Poisson arrivals.

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

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

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

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*):

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

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

$$P(n \mid r) = \frac{e^{-nBq}(nBq)^{n-r}}{(n-r)!}\left(\frac{r}{n}\right) \quad n = r, r+1, \ldots \qquad (3.29)$$

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

$$P_n = \frac{n^{n-1}}{n!} r^{n-1} e^{-rn} \tag{3.30}$$

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:

$$P_n = (m+1)\,(m+1)!\,\frac{(n-1)!}{(m+n+1)!} \tag{3.31}$$

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 $\tau$ 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, $\tau$, was defined by Raff as that value of lag which has the property that the number of accepted lags shorter than $\tau$ is the same as the number of rejected lags longer than $\tau$. This is shown in Figure 3.8.

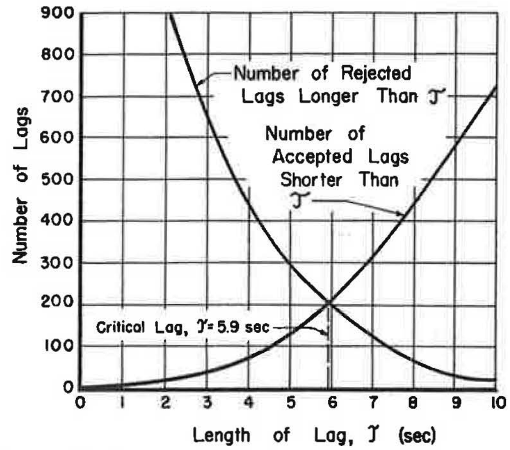
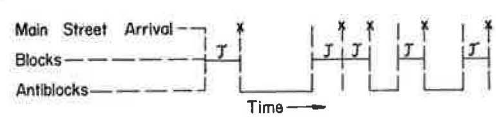



Figure 3.8. Typical distribution of accepted and rejected lags.

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 $\tau$. The time which is more than $\tau$ before the passage of the next car is considered to be in antiblocks as shown in the following:



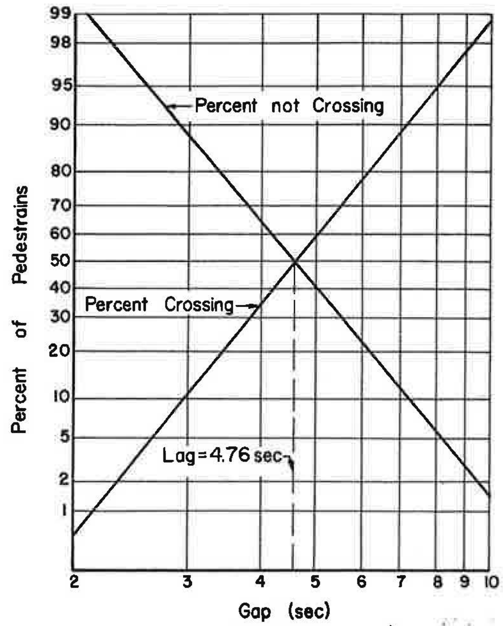



Figure 3.9. Pedestrian gap acceptance.

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

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

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

in which $\tau$ and $\lambda$ are the parameters of the translated exponential distribution. $\tau$ is the minimum acceptable gap, and $\lambda$ 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.

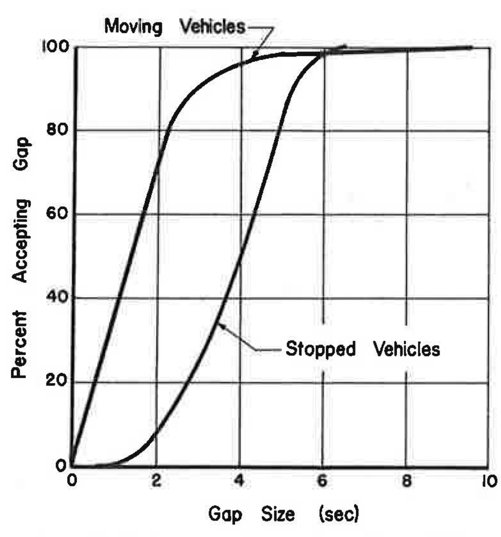



Figure 3.10. Gap acceptance of merging vehicles.

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, $\tau$, they state that the probability of a driver accepting a gap of size $H$ after the $i$th 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.

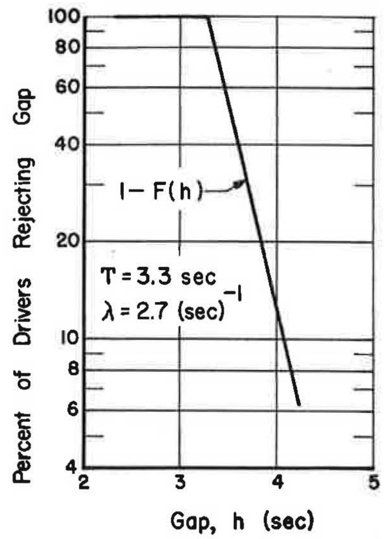



Figure 3.11. Crossing gap rejection.

### 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 $\lambda$, the average interval between arrivals is $1/\lambda$. If the service rate of the system is called $\mu$, 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 $\lambda$ customers per unit of time and where service times are independent and distributed exponentially with a mean rate $\mu$. 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 $\Delta t$ 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 $\Delta t$.
(b) The system was in state $n-1$ at $t$ and one customer arrived in $\Delta t$.
(c) The system was in state $n+1$ at $t$ and one customer departed in $\Delta t$.

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:

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

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

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

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

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

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

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

Because the sum of all probabilities is 1,

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

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

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

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

and

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

The traffic intensity, $\rho$, 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

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

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

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

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

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

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

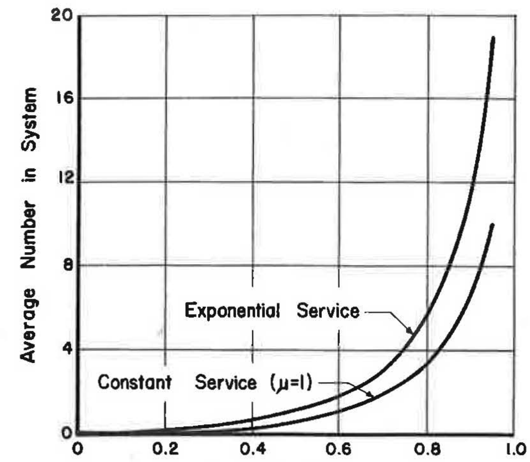



Figure 3.12. Average number in system as a function of traffic intensity.

a $\rho$ of approximately 0.8 or more is reached and then increases rapidly.

The variance of the number in the system is

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

This relationship, plotted in Figure 3.13,

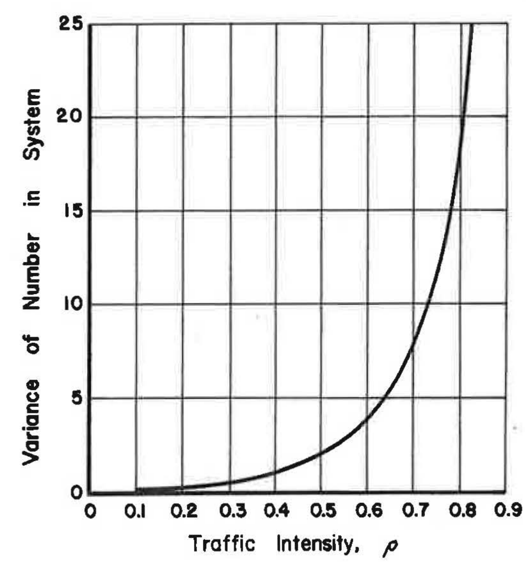



Figure 3.13. Variance of number in system as a function of traffic intensity.

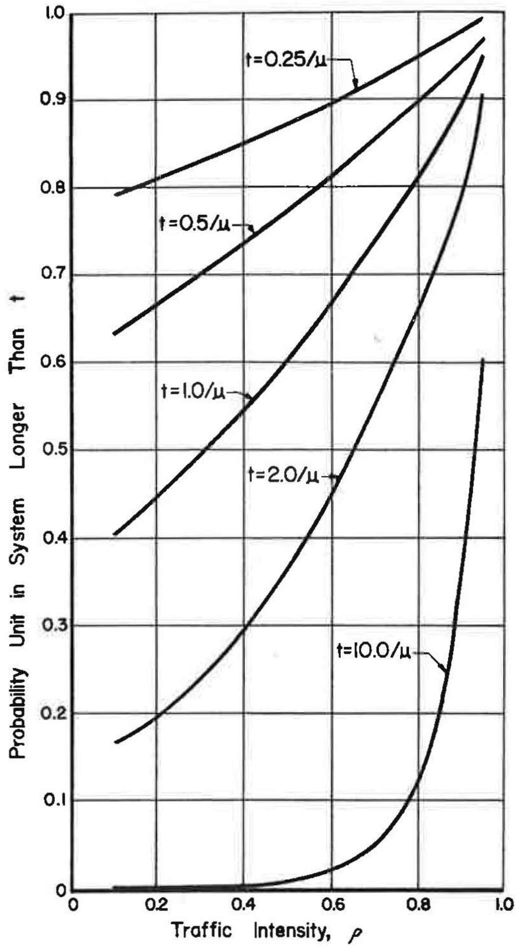



Figure 3.14. Time spent in system.

shows a wide variation in the number in the system at greater values of $\rho$.

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 $\tau$ (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

$$P_d = 1 - e^{-q\tau} \qquad (3.48)$$

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

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

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

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

Adams observed pedestrian delay at five London locations and computed $\tau$ 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 $\tau$ 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, $\tau$, 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:

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

in which $q$ is the main-street flow and $\tau$ 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 $\tau$ is shown in Figure 3.17, in

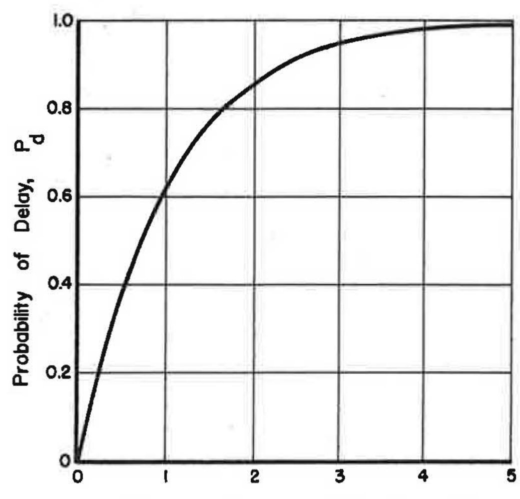



Figure 3.15. Probability of pedestrian delay.

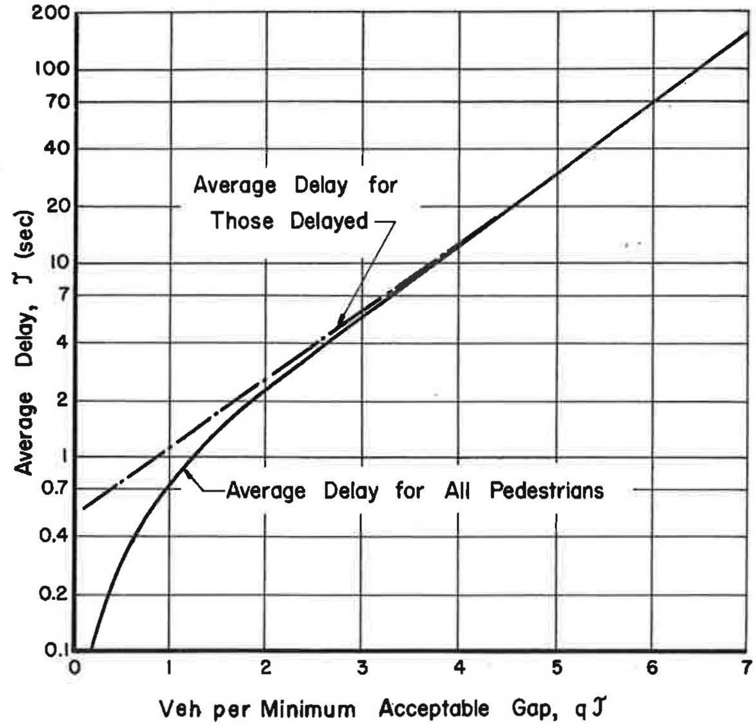



Figure 3.16. Average delay to pedestrians.

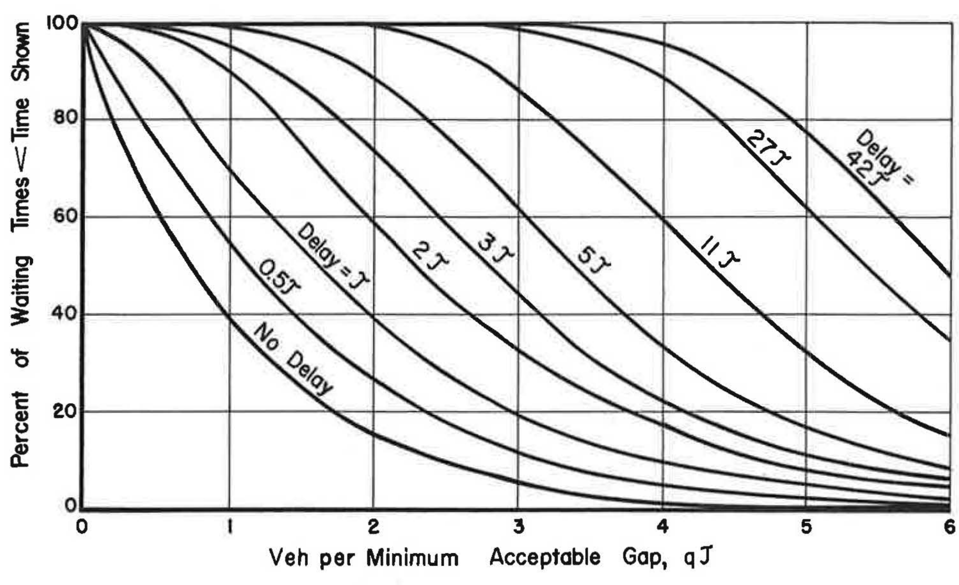



Figure 3.17. Cumulative pedestrian delay.

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

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

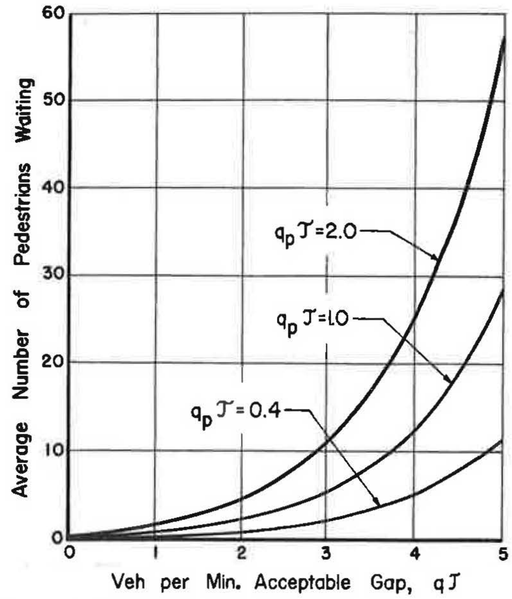



Figure 3.19. Average number of pedestrians waiting to cross street.

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

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

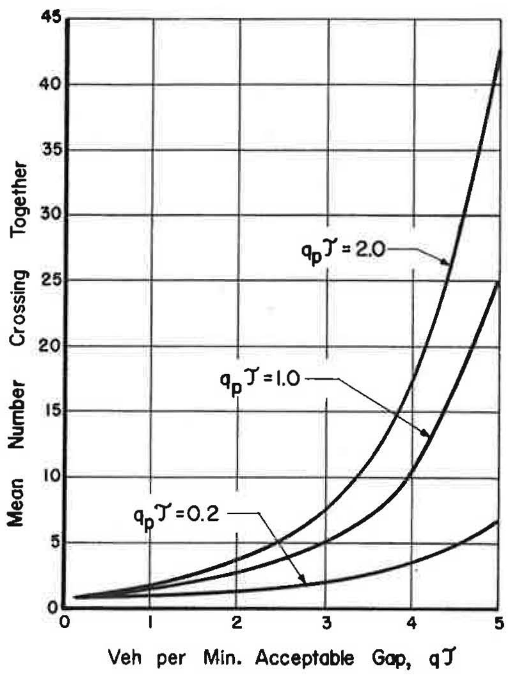



Figure 3.18. Average number of pedestrians crossing together.

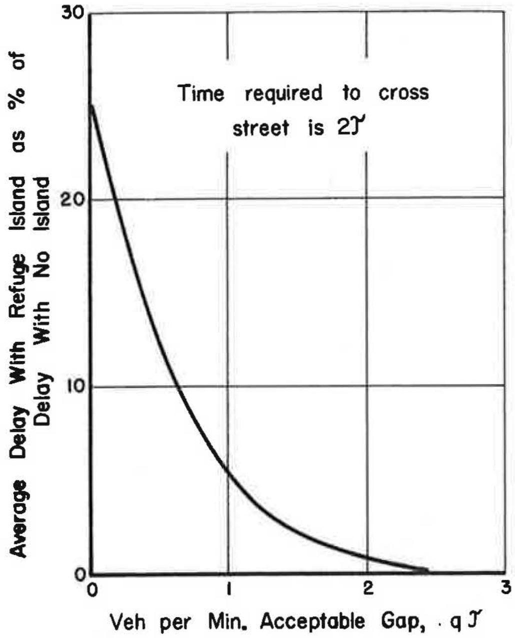



Figure 3.20. Pedestrian delays with and without refuge island.

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

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

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 $\tau$ 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

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

in which $\tau$ 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

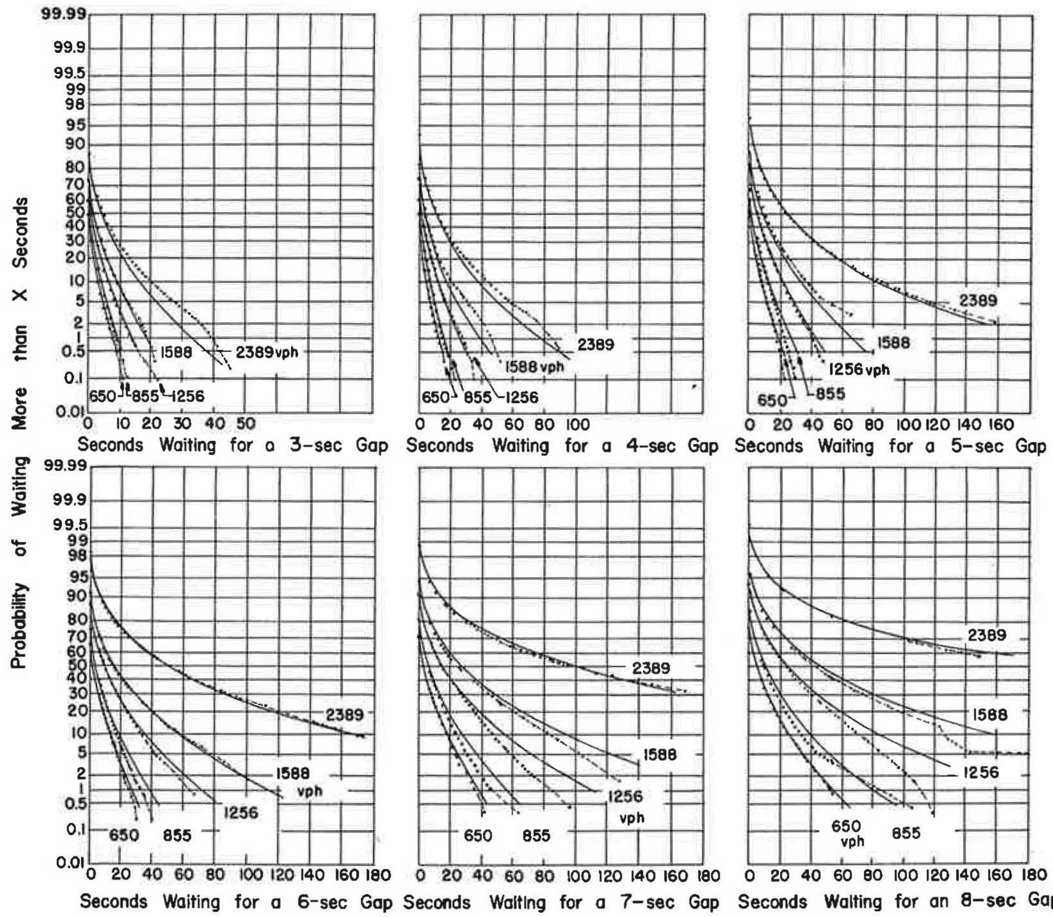



**Figure 3.21. Probability of waiting various times for specified gaps at several traffic volume rates.**

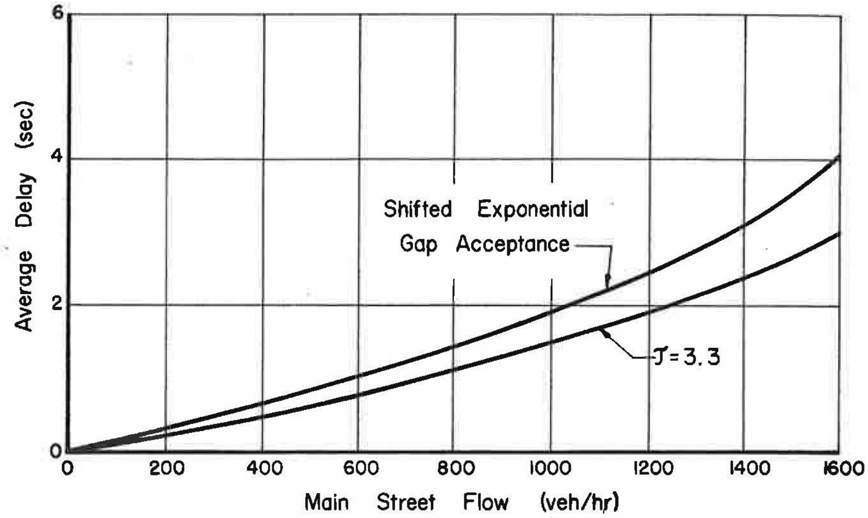


Figure 3.22. Average delay crossing a street.

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

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 $\tau_1$, and a stopped vehicle requires a gap of $\tau_2$ ($\tau_1 \leq \tau_2$), the mean delay is

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

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-

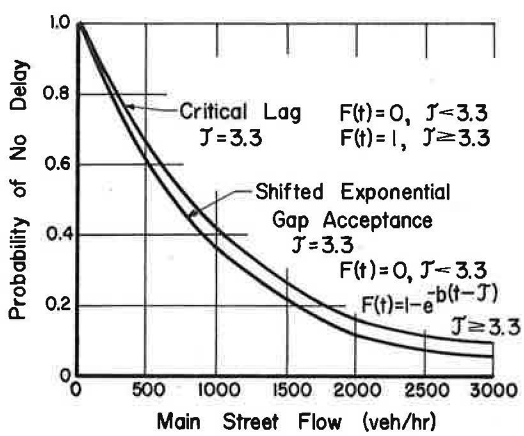


Figure 3.23. Probability of no delay at a signalized intersection.

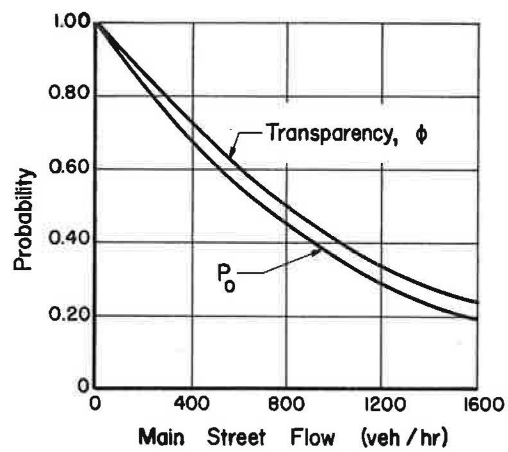


Figure 3.24. Transparency and probability of no delay at a signalized intersection.

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 $\lambda$, Miller derived an expression for the mean waiting time

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

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, $\mathrm{E}(t) = 1/2 \times 1/40\ (10 + 10)^2 = 5$ sec.

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

$$\mathrm{P}_0 = (1 - q\,\bar{t})\, e^{-\lambda\tau} \qquad (3.62)$$

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

$$\mathrm{P}_0 = (1 - 10/50)\, e^{-0.25} = 0.622 \qquad (3.63)$$

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

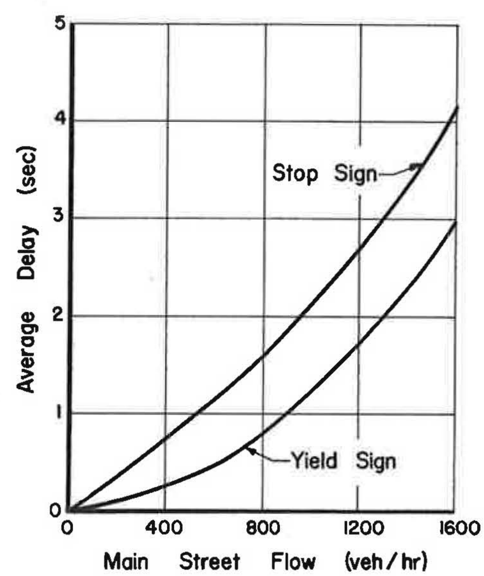



Figure 3.25. Average delay for side-street vehicles.

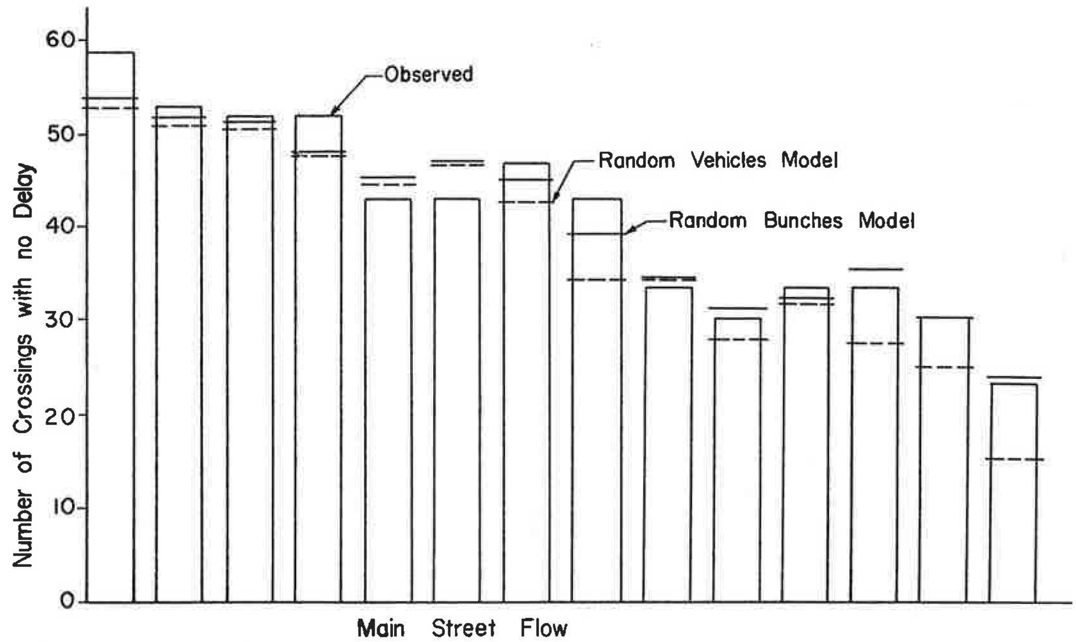



Figure 3.26. Comparison of undelayed crossing opportunities.

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 $\tau$. The distribution of lengths of antiblocks is assumed to be geometric, or

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

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 $\mathrm{f}(b)$ and the average block length is $\mathrm{E}(b)$. They found expected waiting time by developing expected queue lengths at blocked and antiblocked points. The expected queue length is

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

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

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

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.

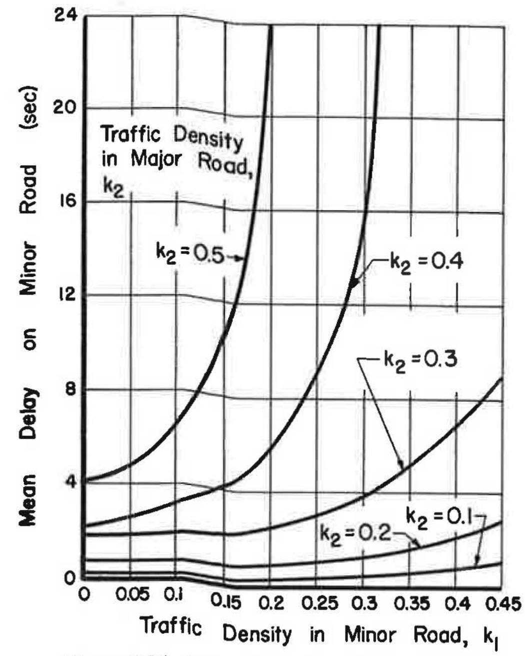



Figure 3.27. Intersection delay for $\tau = 2$ sec.

### 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:

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

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

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

The mean delay per cycle is

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

The mean delay per vehicle is expressed as

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

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

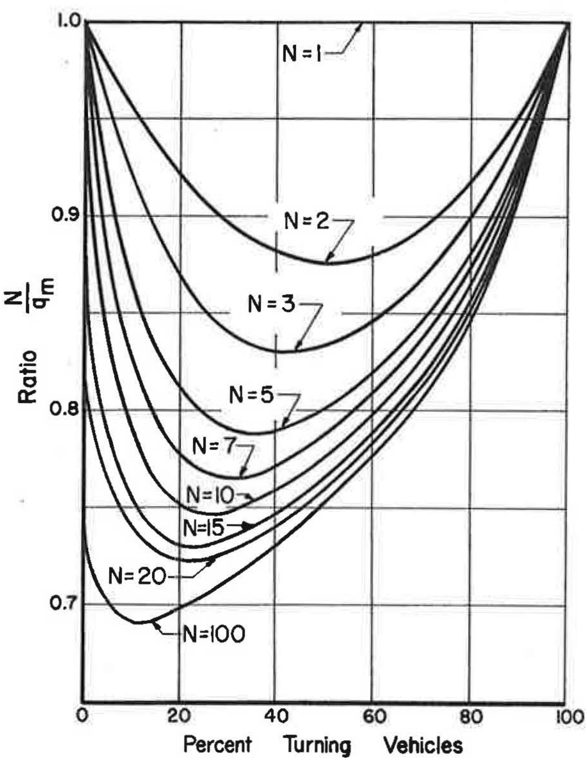



Figure 3.28. The capacity of left-turn movements.

$N$ 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

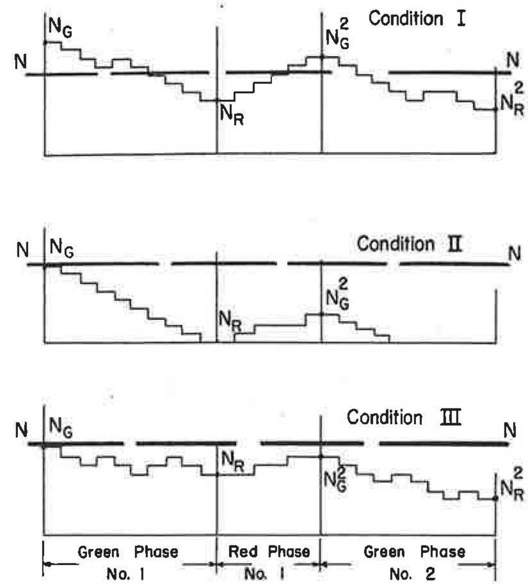



Figure 3.29. Overflow conditions at a signalized intersection.

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.

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 $\rho$ 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

$$E(N) = \frac{qR}{1 - qh} \tag{3.75}$$

The mean square of the queue length is

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

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

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

and the mean square time for the queue to pass is

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

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.

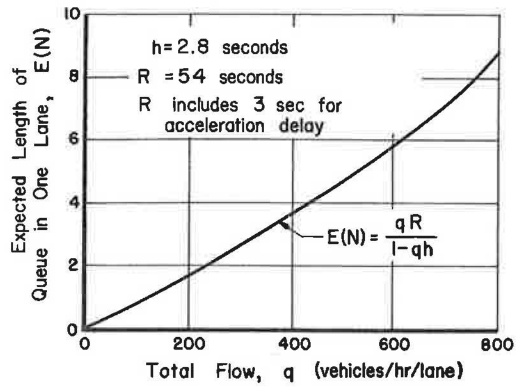



Figure 3.30. Average queue length at a signalized intersection.

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 $q = \frac{500}{2(3{,}600)} = 0.07$ veh/sec and a mean queue length $E(N) = \frac{0.07\,(54)}{1 - (0.07)(2.8)} = \frac{3.75}{0.806} = 4.65$ 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:

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

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

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

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.

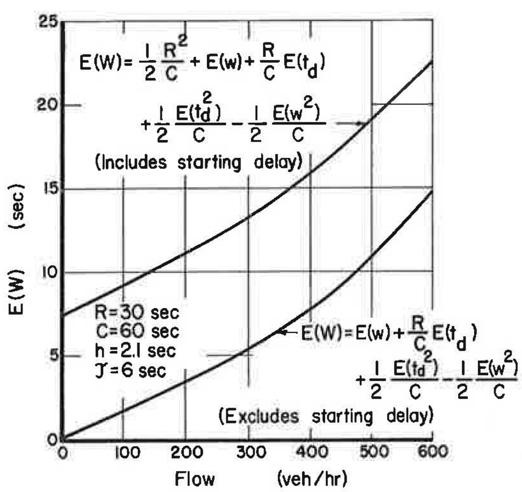



Figure 3.31. Relationship between expected left-turn delay and opposing flow.

(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:

$\tau$ = 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

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

the mean square wait is

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

the average time required for the opposing queue to pass is

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

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

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

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 $i$th approach per unit red time assigned to the $i$ traffic stream, ft/sec;
$K_i = x_i\left(\frac{1}{\bar{u}_i} + \frac{1}{\bar{U}_i}\right)$, 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 |
| $\beta_1 = 80/5 = 16\ \text{sec}^2$ | $\beta_2 = 60/5 = 12\ \text{sec}^2$ |

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 $n$th cycle, $C_n$. The fundamental equations are:

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

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

and

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

in which

$C_n$ = $n$th 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

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

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
$\mathrm{P}(I/T)$ = 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, $\mathrm{P}(I/T)$. 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

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

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

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

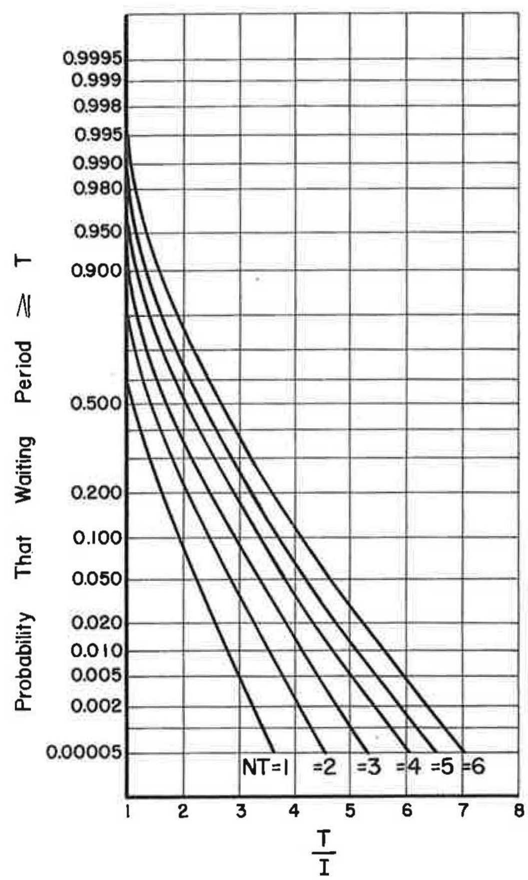



Figure 3.32. Probability of waiting period lasting longer than time $T$ for various vehicle intervals and numbers of vehicles per unit time in opposing stream.

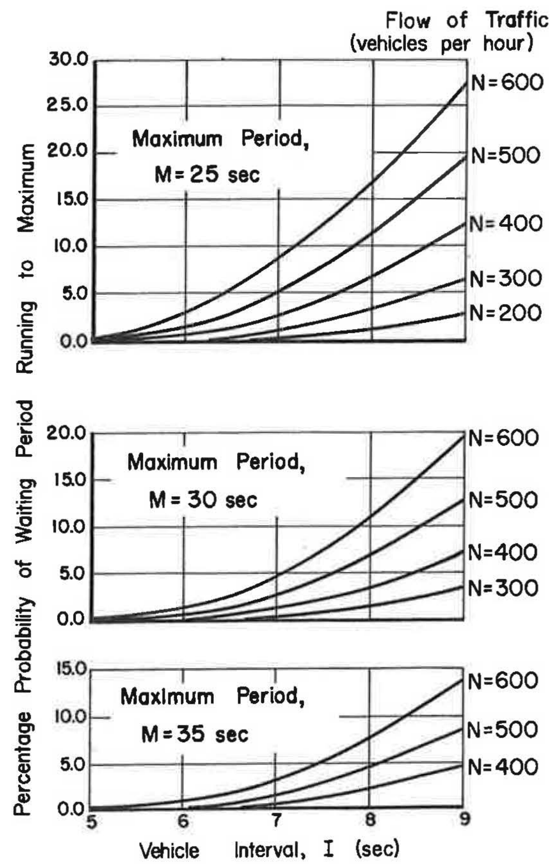



Figure 3.33. Probability of waiting period running to maximum for various values of the maximum period, the vehicle interval, and the main-stream traffic intensity.

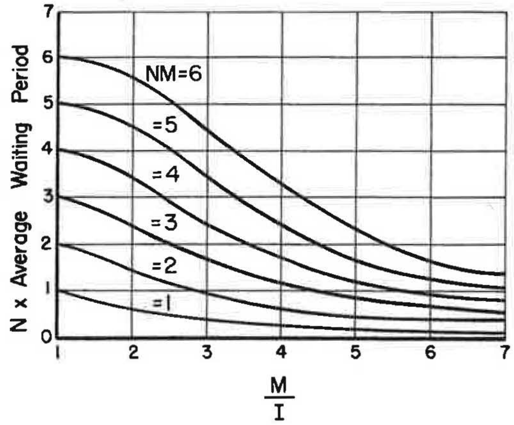



Figure 3.34. Average waiting period for maximum period, $M$, vehicle interval, $I$, and number of vehicles per unit time in opposing stream, $N$.

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

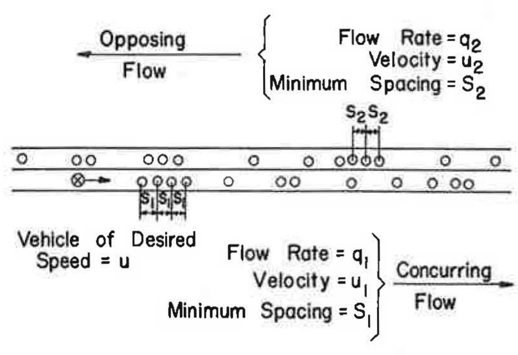



Figure 3.35. Two-lane roadway, showing the corresponding assumed terms in the Tanner model.

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 $\mathrm{E}(u)$ 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 $\mathrm{E}(u)$ in terms of the average waiting time behind all vehicles, $\mathrm{E}(t_w)$, which included zero waits. The expression

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

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

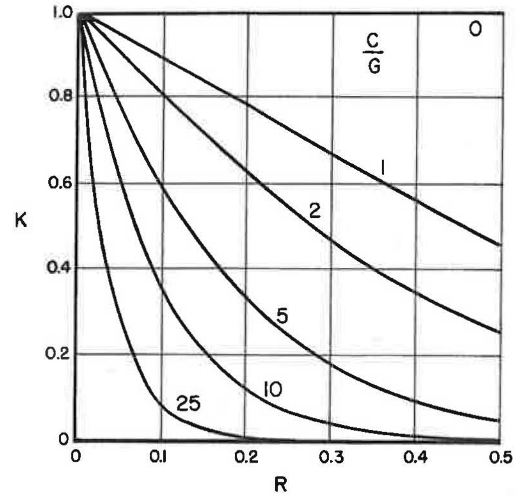



Figure 3.36. Relationship between $K$ and the parameters $R$ and $C/G$.

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

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 $\mathrm{E}(t_w)$ 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 $\mathrm{E}(t_w)$ in Eq. 3.98 gives an expression for the average speed $\mathrm{E}(u)$ 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.

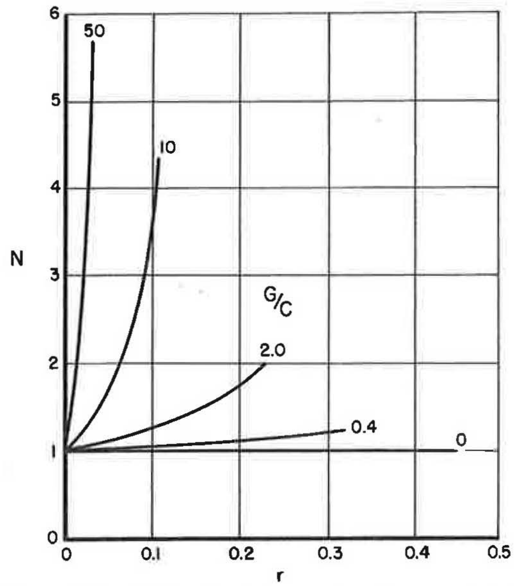



Figure 3.37. Relationship between N and the parameters r and G/C.

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 $\varphi$, 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 $\varphi$ and log $q_2$ for the field data he used from a study in Sweden.

This model assumes that there is an intensity of $\rho$ 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, $\rho_1/\rho$, 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 $\sigma$.

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, $\rho_1/\rho$ may be determined by

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

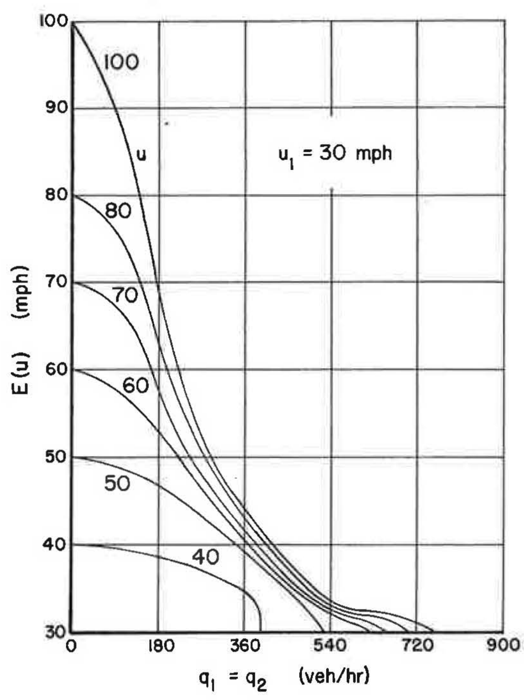



Figure 3.38. Effect of traffic when equally divided between directions for various values of u (two-way traffic).

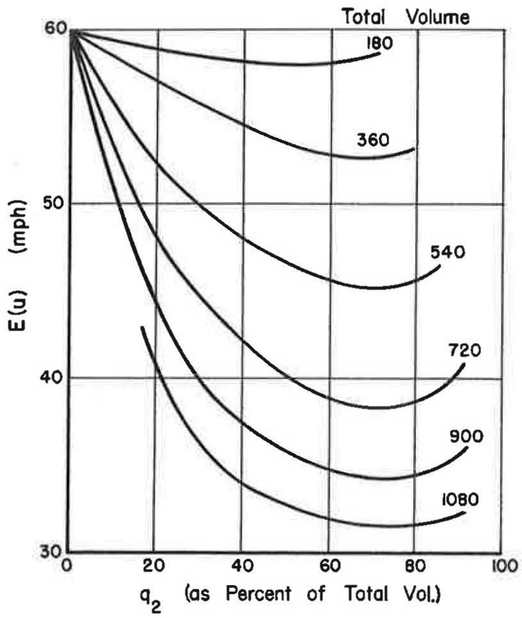



Figure 3.39. Effect of varying proportion of opposing traffic for various levels of total traffic.

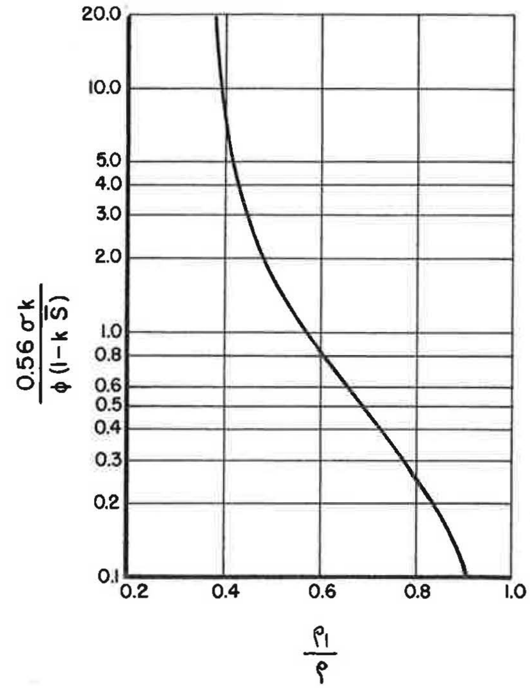



Figure 3.40. Solutions of Eq. 3.101.

Figure 3.40 shows the solutions of Eq. 3.101 for various values of $\dfrac{0.56\,\sigma k}{\varphi\,(1 - k\bar{S})}$. For values of $\rho_1/\rho$ 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

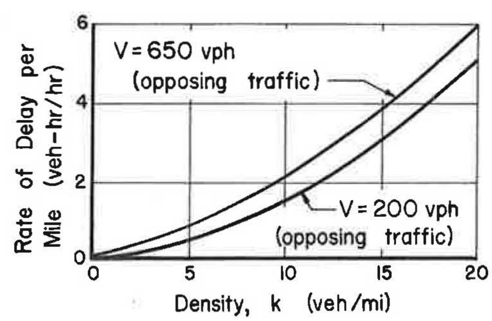



Figure 3.41. Rate of delay versus density.

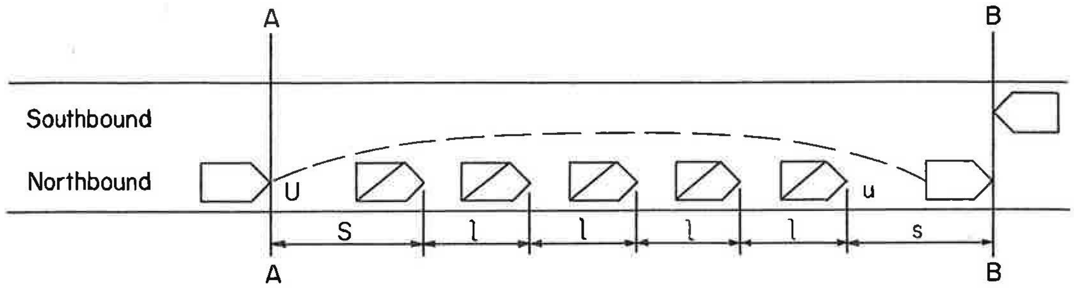



Figure 3.42. Passing $n$ continuous slower-moving vehicles on a two-lane highway.

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

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

in which

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

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

$$t = \frac{S + s}{u} \tag{3.106}$$

$V'_1$ = 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 $\tau_n$ (second term of Eq. 3.103). The value $\tau'_n$ 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 $\tau_n$ is the sum of the time required to pass, $\tau'_n$, 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

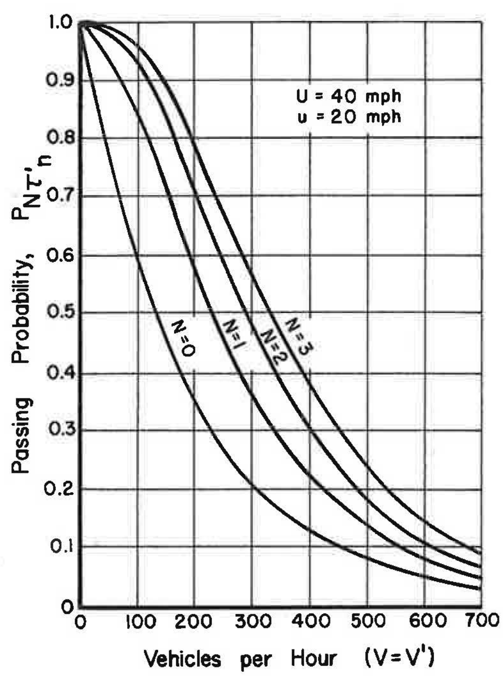



Figure 3.43. Passing probability on a two-lane highway.

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 $\tau$ 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 $\tau$ 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-

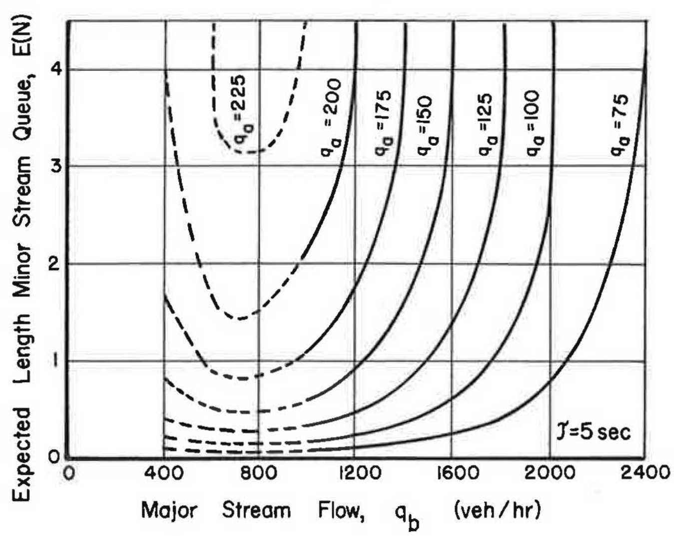



Figure 3.44. Relationship of the average minor stream queue length and the major stream flow rate. Dashed curves represent range in which this model cannot be expected to yield a reasonable approximation of the average queue length.

come first-served basis with the appearance of a minimum acceptable gap length of $\tau$.

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).

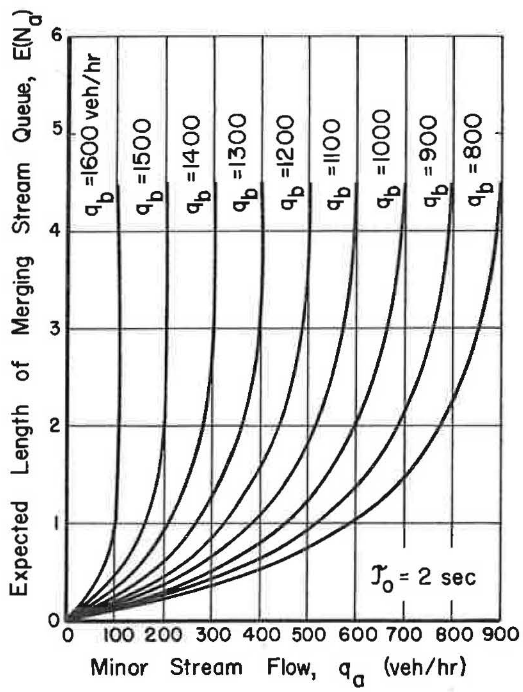



**Figure 3.45.** Expected length of queue in the merging stream, in terms of the average flow of the major and minor streams.

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

$h_i$ = $i$th 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);

$\alpha$ = number of vehicles of the merging stream which merge into the $i$th 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 $\tau_0$, 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

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

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.

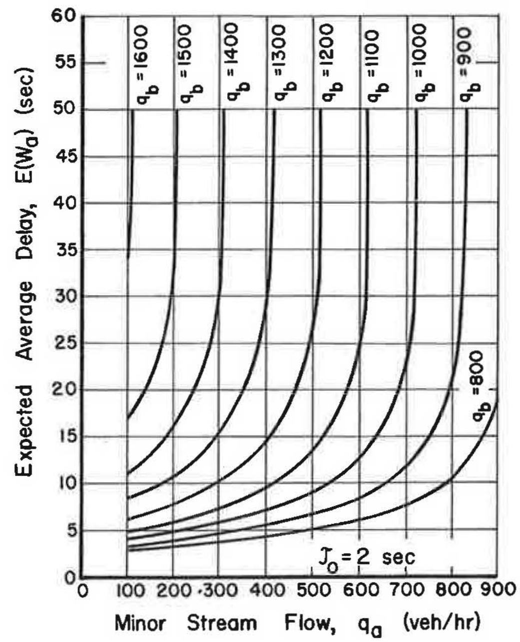



**Figure 3.46. Expected average delay to a vehicle in the merging stream, in terms of the average flow of the major and minor streams.**

### 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

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

in which

$\lambda_i$ = the flow in direction $i$;
$\mu_i$ = 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 $i$;
$M_i = \mathrm{E}(e^{-\lambda_j t_i})$;
$m_i = \mathrm{E}(t_i^2)$;
$\alpha = M_i + M_j - M_i\,M_j$;
$d_i = \lambda_i + \lambda_j - \lambda_i\,M_i$;
$T_i$ = the time required for a vehicle moving in direction $i$ to pass through the one-lane section;
$B_i = \dfrac{1}{\lambda_i}\left[\exp\,[\lambda_i\,(T_i - 1/\mu_i)] - 1\right]$

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 $\mu_i$ and $\mu_j$ 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, $\rho$, is less than unity; that is, the mean service

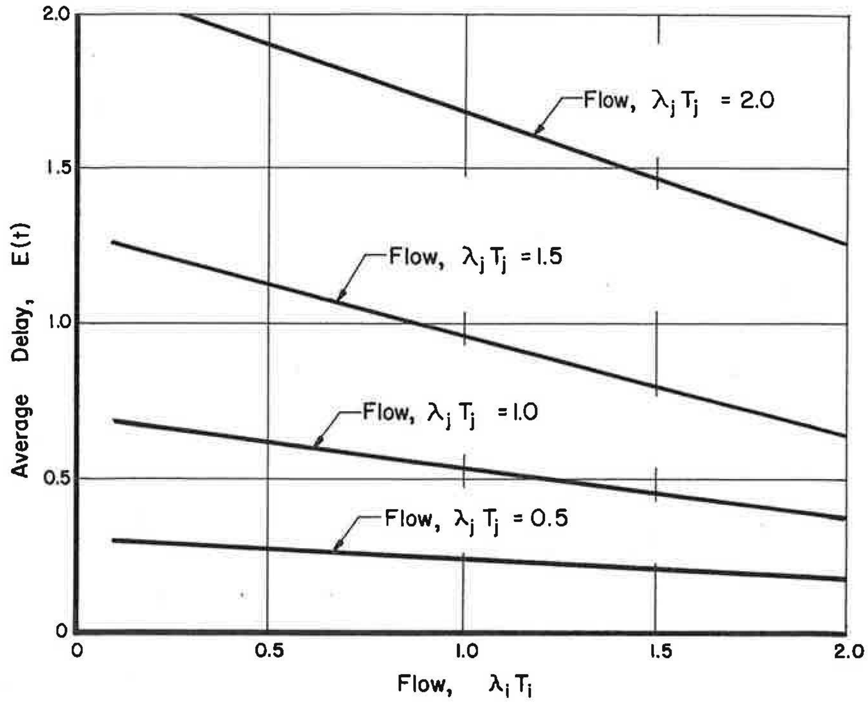



Figure 3.47. Average delays at a one-lane bottleneck for varying traffic flows and passage time through the bottleneck (see note in text).

rate $\mu$ exceeds the mean rate of arrivals $\lambda$. 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, $\lambda$, exceeds the mean service rate, $\mu$, 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 $\rho_0$ and an expected number of vehicles in the system of $E(n)$. Now suppose that immediately the traffic intensity increases to $\rho_1$, 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:

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

Similar treatment of variances yields

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

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 $\lambda t$ and $\mu t$ for Poisson arrivals and negative exponential service times, the number waiting in the system at time $t$ may be expressed by

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

and their variance by

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

in which $\rho_0$ is the initial traffic intensity ($\rho_0 = \lambda_0/\mu$) and $\lambda_0$ 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

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

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 $E\,[n(60)] = \frac{3/4}{1-3/4} + \frac{4}{3}\left(\frac{3}{2} - 1\right) 60 = 43$; $\mathrm{Var}\,[n(60)] = \frac{3/4}{(1-3/4)^2} + \frac{4}{3}\left(\frac{3}{2}+1\right)60 = 212.$

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

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 $\rho_1$ 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

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

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

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 $E(T) = \left[\frac{43}{4/3} - \frac{3/4}{2(1-3/4)}\right] \Big/ (1-3/4) = 123$ 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

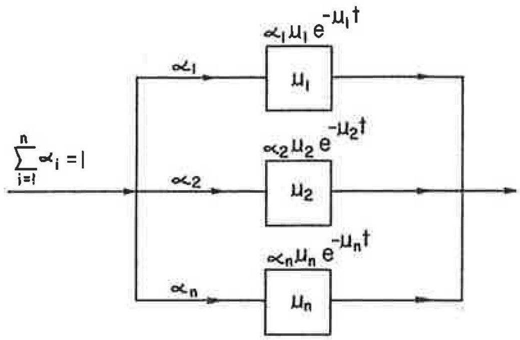



Figure 3.48. Exponential service facilities in parallel (hyper-exponential).

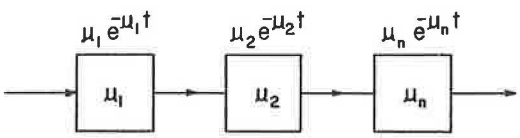



Figure 3.49. Exponential service facilities in series (Erlang, if $\mu_1 = \mu_2 = \ldots = \mu_n$).

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 $\mu_i$ acting on some fraction of the total number of arrivals, $\alpha_i$ (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

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

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 $K = 1$,

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

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 $\mu$ 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:

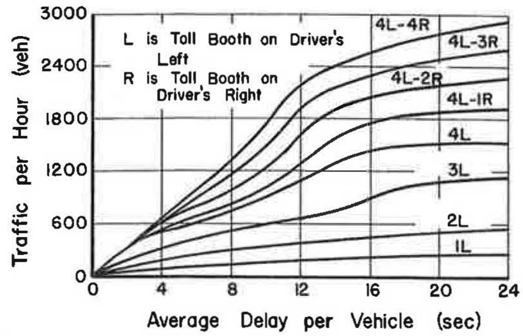



Figure 3.50. Average delay for various volumes and toll booth combinations.

"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

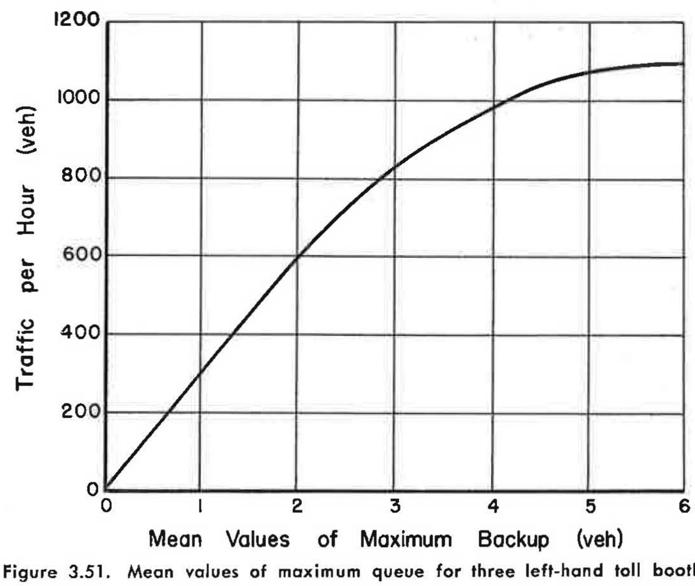



Figure 3.51. Mean values of maximum queue for three left-hand toll booths.

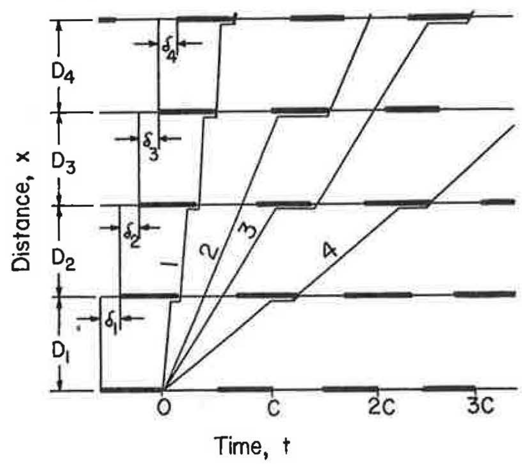



Figure 3.52. Trajectories of cars passing through a series of traffic lights (heavy bars represent red phases).

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 $\delta$, and is

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

in which

$Z$ = the fraction of the cycle change;
$C$ = the cycle length;
$D$ = the distance between signals; and
$\delta$ = 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 $D/\delta$, 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:

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

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 $\rho$ 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, $\rho$, is also equal to $\frac{\lambda E(t)}{N}$, $\lambda$ 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, $\rho$.

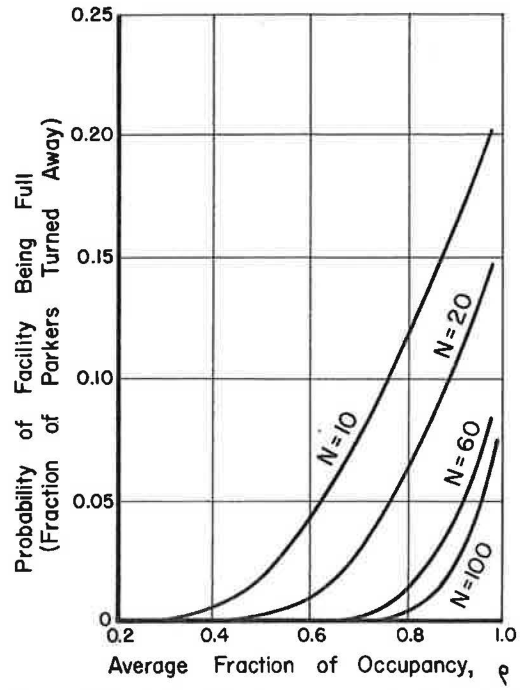



Figure 3.53. Probability of a parking facility with N spaces being full.

## 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).

Chapter 4

# SIMULATION OF TRAFFIC FLOW

D. L. Gerlough
Planning Research Corporation

## Contents

*Section*

# Chapter 4

# SIMULATION OF TRAFFIC FLOW

## 4.1 INTRODUCTION

Simulation of vehicular traffic on high-speed computers has attracted considerable interest within recent years.

Simulation is a technique which enables the study of a complex traffic system in the laboratory rather than in the field. It is usually faster and less expensive than the testing of a real system and, in many cases, enables study of system characteristics prior to construction of the facility.

The verb, *to simulate*, is defined in the dictionary as "to assume the appearance of, without the reality; to feign."

Various types of simulation have been used in engineering for many years. The scale model has been of considerable value in the study of structures, hydraulic systems, aerodynamic systems, etc. Even some controlled experiments may be considered simulation in that they are controlled in such a manner as to remove one or more important variables of system operation. One important technique of simulation is the study of analogous systems; that is, systems having the same mathematical relationship as the system in question. An example is the study of the oscillations in an electrical system in order to learn the behavior of vibrations in a mechanical system.

In the study of traffic simulation it is possible to represent traffic of the particular characteristic desired and in the quantities desired, whereas to obtain the characteristic in the field may be very difficult. In traffic studies, furthermore, there may be a substantial hazard in conducting field trials, whereas the same situation can be studied by simulation without risk.

## 4.2 USE OF COMPUTERS IN SIMULATION

Since World War II, there has been an increasing use of computers to speed and simplify the mathematical processes involved in simulation. These computers are of two general types: analog (continuous-variable) and digital (discrete-variable). Both types have been used for simulation, and it has become practice to characterize the simulation technique as analog or digital, depending on the way the problem is formulated and the type of computer needed for the formulation.

### 4.2.1 Analog Simulation

An analog computer is one in which computation is performed by varying the state of some physical element in which the variables are continuous. One class of such computers has the integrator as its principal component. Nearly all analog computers are electronic, and integrators are of the so-called "operational amplifier" type.

When analog simulation is used, all parts of the system must be simulated simultaneously. Each component or function of the system must be simulated by one or more components in the computer. This requires the addition of more computer equipment as the system simulated becomes more complex. For small systems this is not serious, but for the study of large systems the addition of more simulator elements can become expensive. In many cases, further additions may not be feasible because there is a practical limit to the number of integrators that will work together satisfactorily. The accuracy of the analog computer is limited to the accuracy of the physical components involved, and the error can seldom be made smaller than 1 part in 10,000.

### 4.2.2 Digital Simulation

Digital simulation is characterized by the use of a digital computer. Whereas the analog computer must handle all elements of the simulation simultaneously (in parallel), the digital computer handles elements of the simulation one after another (in series). In this case, an increase in system complexity results in an increase in the time required for computation. Although accuracy in simulation may not necessarily be important, it is possible to reach any degree of accuracy desired by doing more computation. In analog simulation, the mathematical models used must be those which involve differential equations or which can be made to look as though they involve differential equations. With digital simulation, however, it is possible to use models described in words rather than mathematical terms.

## 4.3 SIMULATION TECHNIQUES

The following steps are normally required in the simulation of any system:

(a) Definition of problem.
(b) Formulation of a model, including the selection of a figure of merit (measure of effectiveness).
(c) Preparation of the computer "program" which will implement the model.
(d) Conducting experimental runs of the simulated system (including experimental design to determine the number of runs and the parameter values to be used).
(e) Interpretation of results.

Items (a), (d) and (e) are known to any engineer familiar with the techniques of experimental investigation. Item (c) (programming) is a function of the computer used and is frequently routine if the model contains sufficient detail. However, item (b), formulation of a model, is a critical factor in the development of simulation techniques and is emphasized in the following sections.

### 4.3.1 Characteristics of the Model

The model is a statement of the problem with only important features of the system to be studied included. A model has been defined (*16*) as "a logical description of a physical system originating in the mind of the investigator and adequately accounting for what he considers to be significant behavior." Characteristics of a system should be stated by mathematical equations when possible. If data are not known or a suitable mathematical statement is not possible, the behavior of the system is described in words. There may be parts of the system which involve random or stochastic variables. These are treated by what are known as Monte Carlo techniques.

Applying these procedures to a traffic problem, the flow of vehicles on a given network of streets is simulated, subject to established rules of conduct and controls. Then, if a random sample of traffic flow is introduced into the network, the effect of control devices and other variables may be observed. These random samples may be deduced and prepared from a combination of empirical data and theoretical considerations, or solely from empirical data.

Important elements of any model are:

(a) Statement of the behavior of each of the components and inputs of the system. This will include probability distributions of any random phenomena.
(b) Selection of one or more measures of effectiveness (criteria) by which the performance of the system is to be judged.
(c) Statement of any particular assumptions, simplifications or dissections of the model which may be necessary to permit adaptation of the model to a particular computer.

### 4.3.2 Model Formulation

In the design or improvement of any highway facility, the engineer is required to establish basic rules by which the effectiveness of the design or improvement can be measured. This is accomplished by formulating the simulation model in such a manner that the measure of effectiveness is expressed as a function of the variables of the system.

**4.3.2.1 Block Diagrams.** The formulation of a model is greatly aided by the construction of a functional diagram incorporating certain postulates. In the formulation of a model of both vehicular and pedestrian traffic, one takes the viewpoint of an observer standing beside a road. To such an observer, the vehicle arrival rate, the traffic interferences encountered, and the direction taken by cars on leaving the area appear to be random processes. Once a functional diagram has been drawn, it serves as a

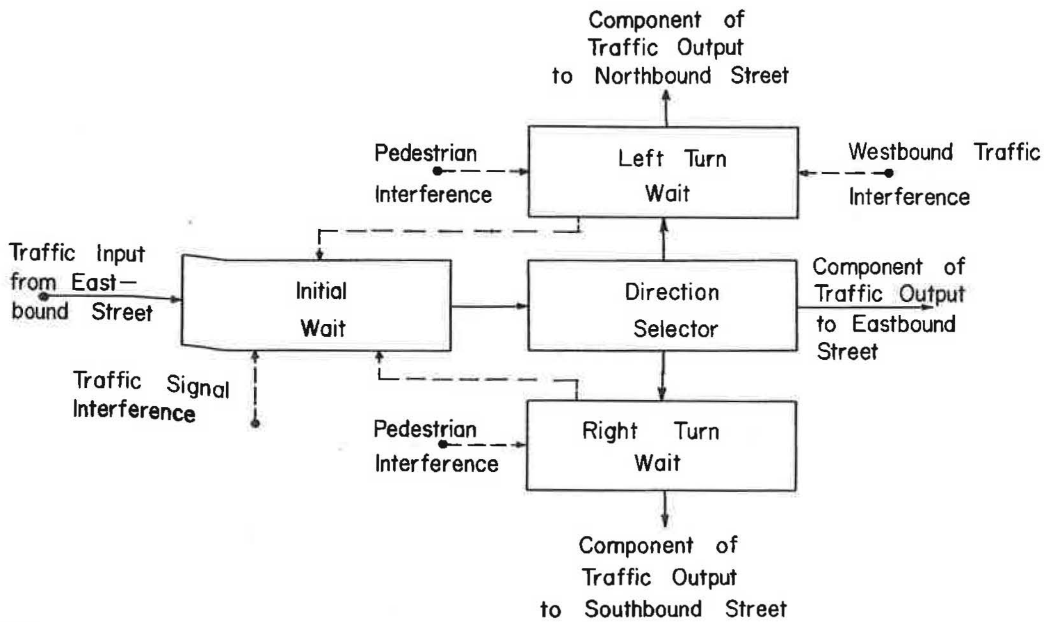



Figure 4.1. Block diagram of one approach of a generalized intersection having one lane in each direction.

guide in the writing of mathematical expressions, in the statement of a set of "rules of the road," and in the design of suitable units for a simulator.

For example, consider Figure 4.1, which is a functional or block diagram for a typical intersection. This diagram is based on the following postulates:

(a) The intersection is signalized.
(b) There is one lane in each direction.
(c) Pedestrians are present.
(d) Right turns are permitted only on green.
(e) Left turns are permitted.

The diagram is drawn for only one-quarter of the intersection—that representing traffic entering the intersection from the eastbound approach. Traffic entering from other approaches would be represented by similar diagrams.

In Figure 4.1, solid lines represent the flow of traffic; broken lines represent the flow of interference information. It may be seen that arriving vehicles pass into an initial waiting zone. Vehicles are delayed here if the signal is red or, since one-lane operation is considered, if the car ahead encounters delay in performing a turn. Having passed the initial delay, the car enters a direction selector, which determines whether the car is turning right, turning left, or going straight ahead. Cars going straight ahead proceed without further delay. Cars turning right may be delayed by pedestrians. Cars turning left may be delayed by oncoming traffic or pedestrians. The broken lines from the right- and left-turn wait blocks to the initial wait block represent the feedback of information to prevent a car from entering the intersection when a vehicle is delayed within the intersection.

4.3.2.2 Measure of Effectiveness. In most studies it is desirable to have a criterion or figure of merit by which to judge the behavior of the system as parameters are varied. If optimum performance is sought, there must be some quantity which is to be optimized. Several such figures are worthy of consideration:

(a) Average time for a vehicle to traverse the system.
(b) Percentage of vehicles required to travel at speeds less than their desired speeds.
(c) Average number of seconds delay per car-mile.

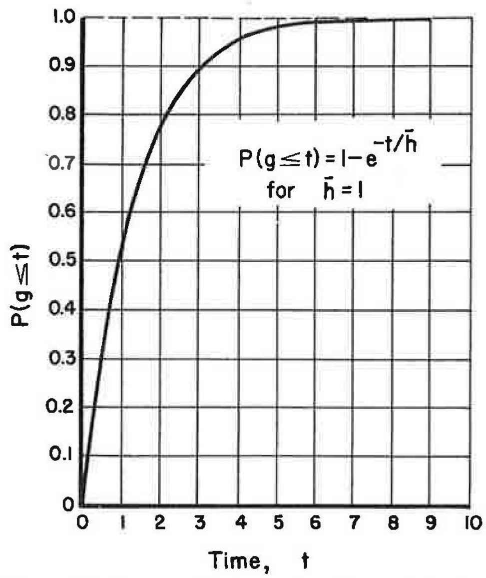



Figure 4.2. Exponential distribution of gaps (inter-arrival times). Ordinate is the probability of a gap $t$ or less; $\bar{h}$ is the average inter-arrival time.

(d) Number of lane changes per car-second.

(e) A figure of merit, $B$, to be defined as follows:

Let

$u_d$ = desired speed of any vehicle, in ft/sec;

$u_a$ = actual speed of the same vehicle, in ft/sec (it is assumed that $u_a \leq u_d$);

$\Delta$ = the distance lost in 1 sec by a vehicle traveling at $u_a$ instead of $u_d$; and

$t$ = the time, in sec, lost by a vehicle during each second that it is required to travel at a reduced speed.

Then

$$\Delta = 1u_d - 1u_a = u_d - u_a \qquad (4.1)$$

and

$$t = \frac{\Delta}{u_d} = \frac{u_d - u_a}{u_d} = 1 - \frac{u_a}{u_d} \qquad (4.2)$$

If during the $i$th second $n_i$ vehicles are traveling at reduced speeds, the total time lost, $\tau$, in car-seconds, will be

$$\tau = \sum_i \sum_j n_{ij}\, t_{ij} \qquad (4.3)$$

in which $j$ denotes the various reduced-speed conditions.

Now let

$T$ = the duration of a particular run, in sec; and

$q$ = the flow of traffic, in cars/sec.

Then define $B$ as follows:

$$B = \frac{\tau}{Tq} \qquad (4.4)$$

(Note: The units are sec/car.)

**4.3.2.3 Probability Distributions of Traffic Input Characteristics.** Simulation is valuable for the study of traffic flow because it enables the engineer to incorporate the random nature of traffic into the simulation models. A simulation model, therefore, must include a description of the variables to be treated as random.

There are many variables associated with a traffic system. Some are associated with characteristics of vehicles, some with characteristics of the roadway, and others with characteristics of drivers. Nearly all are of a statistical nature. Due to a lack of knowledge of the distributions and laws of interaction of traffic system components, the

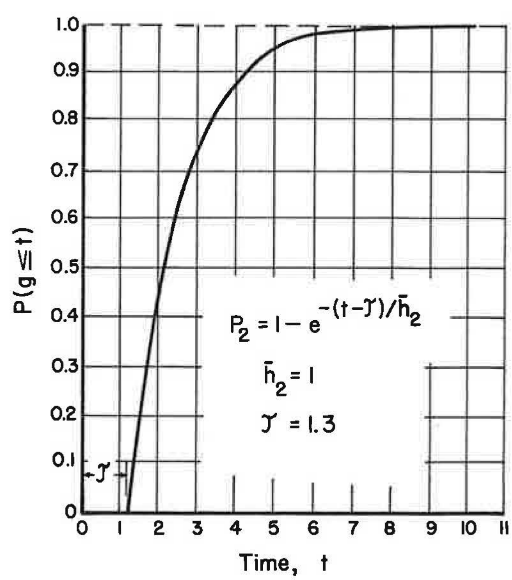



Figure 4.3. Exponential distribution of gaps shifted from origin. Ordinate is the probability of a gap $(t - \tau)$ or less. Average inter-arrival time is $(\bar{h}_2 + \tau)$.

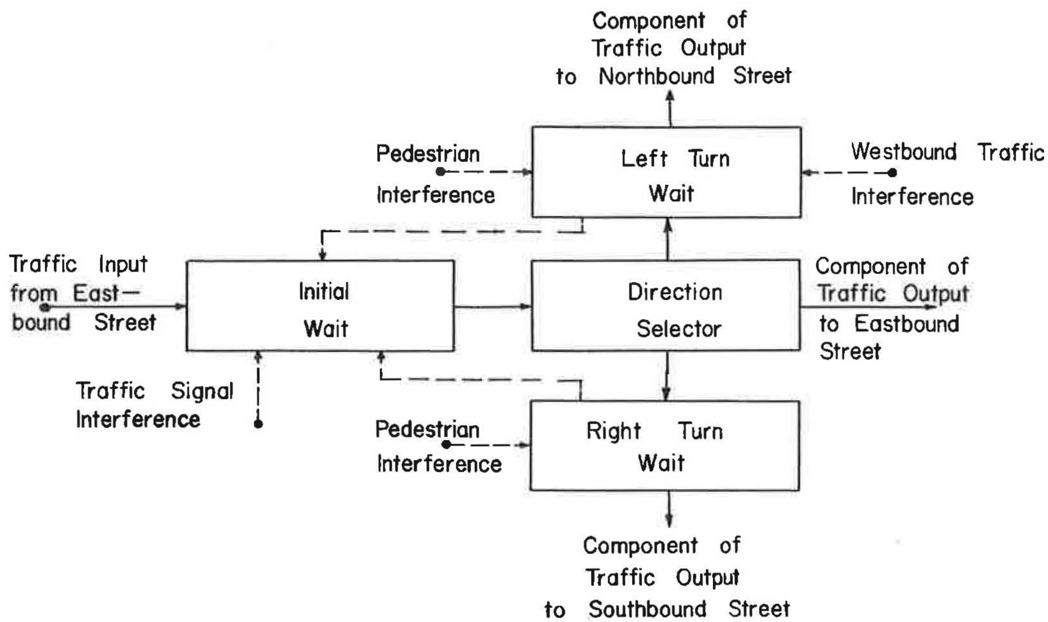



Figure 4.1. Block diagram of one approach of a generalized intersection having one lane in each direction.

guide in the writing of mathematical expressions, in the statement of a set of "rules of the road," and in the design of suitable units for a simulator.

For example, consider Figure 4.1, which is a functional or block diagram for a typical intersection. This diagram is based on the following postulates:

(a) The intersection is signalized.
(b) There is one lane in each direction.
(c) Pedestrians are present.
(d) Right turns are permitted only on green.
(e) Left turns are permitted.

The diagram is drawn for only one-quarter of the intersection—that representing traffic entering the intersection from the eastbound approach. Traffic entering from other approaches would be represented by similar diagrams.

In Figure 4.1, solid lines represent the flow of traffic; broken lines represent the flow of interference information. It may be seen that arriving vehicles pass into an initial waiting zone. Vehicles are delayed here if the signal is red or, since one-lane operation is considered, if the car ahead encounters delay in performing a turn. Having passed the initial delay, the car enters a direction selector, which determines whether the car is turning right, turning left, or going straight ahead. Cars going straight ahead proceed without further delay. Cars turning right may be delayed by pedestrians. Cars turning left may be delayed by oncoming traffic or pedestrians. The broken lines from the right- and left-turn wait blocks to the initial wait block represent the feedback of information to prevent a car from entering the intersection when a vehicle is delayed within the intersection.

**4.3.2.2 Measure of Effectiveness.** In most studies it is desirable to have a criterion or figure of merit by which to judge the behavior of the system as parameters are varied. If optimum performance is sought, there must be some quantity which is to be optimized. Several such figures are worthy of consideration:

(a) Average time for a vehicle to traverse the system.
(b) Percentage of vehicles required to travel at speeds less than their desired speeds.
(c) Average number of seconds delay per car-mile.

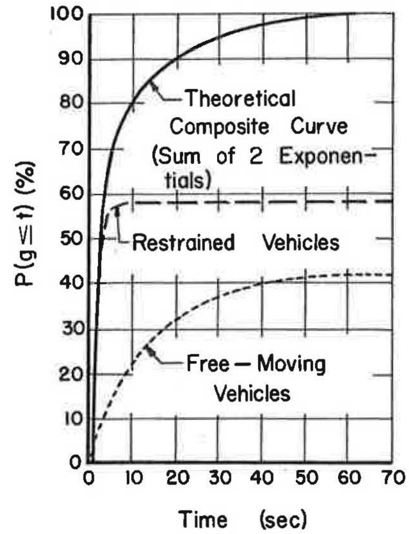



Figure 4.4. Composite exponential curve for inter-arrival times.

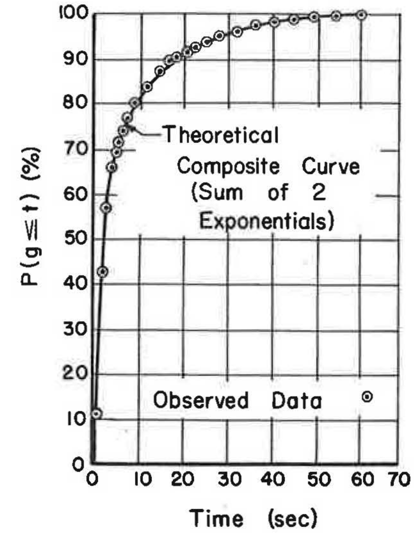



Figure 4.5. Composite exponential curve fitted to observed data.

traffic engineer must usually fit the distributions to observations of the over-all system. The variables usually observed are (in order of ease of measurement): flow (rate), inter-arrival times, and speeds. If given the distribution of one variable, it is sometimes possible through simulation to determine the distribution of another.

Greenshields and others (*26, 24*) have shown that with low to moderate flow and with a sufficient number of lanes so that vehicles can pass at will, vehicle arrivals follow the Poisson distribution. Thus inter-arrival times follow the exponential distribution of Figure 4.2. If cars are constrained so that they cannot pass, short inter-arrival times cannot occur. It is possible to use a shifted exponential distribution, as shown in Figure 4.3, to describe this situation.

Schuhl (*30*), however, has suggested that a traffic stream may be regarded as a mixture of constrained and unconstrained cars. Unconstrained cars can be represented by an exponential through the origin; constrained cars, by a shifted exponential (see Figs. 4.4 and 4.5). The composite is the sum of these two exponentials. Kell (*29*) has given considerable attention to techniques for calculating parameters of composite exponential curves for a given flow.

Haight, Whisler, and Mosher (*28*) have shown that under heavy traffic flows arrivals follow the "generalized Poisson" distribution. This suggests that inter-arrival times would follow the Erlang distribution. This distribution has not, however, been previously used for simulation.

The inter-arrival times of cars entering a signalized intersection after the green "go" has been displayed have been studied by Greenshields (*26*) and others. Although

Table 4.1 Time Intervals Between Cars Entering a Signalized Intersection[1]

| Vehicle Number | Interval Between Vehicles (sec.) | Entrance Time[2] (sec.) |
|---|---|---|
| 1 | — | 3.8 |
| 2 | 3.1 | 6.9 |
| 3 | 2.7 | 9.6 |
| 4 | 2.4 | 12.0 |
| 5 | 2.2 | 14.2 |
| 6 | 2.1 | 16.3 |
| 7 | 2.1 | 18.4 |
| 8 | 2.1 | 20.5 |
| 9 | 2.1 | 22.6 |
| 10 | 2.1 | 24.7 |

[1] After Greenshields (*26*).
[2] From change of light to green.

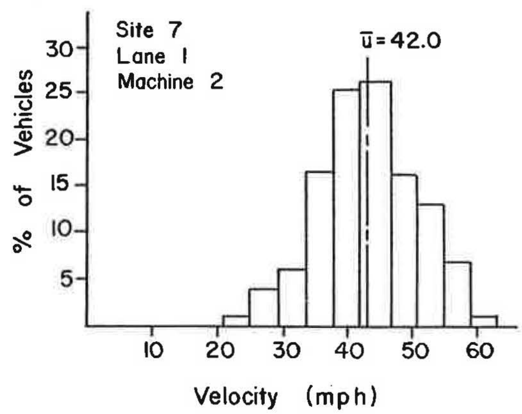



Figure 4.6. Typical speed distribution curve.

these times are properly represented by distributions, some investigators have used the mean values given by Greenshields (Table 4.1).

Speed distributions have been found to be substantially normal. The mean and the standard deviation of speeds decrease as the flow increases (*18*). Figure 4.6 is a typical distribution of observed speeds.

Many traffic variables do not lend themselves to representation by theoretical distributions. These variables must be treated by empirical curves fitted to observed data. The manner in which traffic distributes itself among the lanes of a multilane freeway is an example of a variable which requires an empirical curve. Figure 4.7 shows the use of third-order polynomials fitted to em-

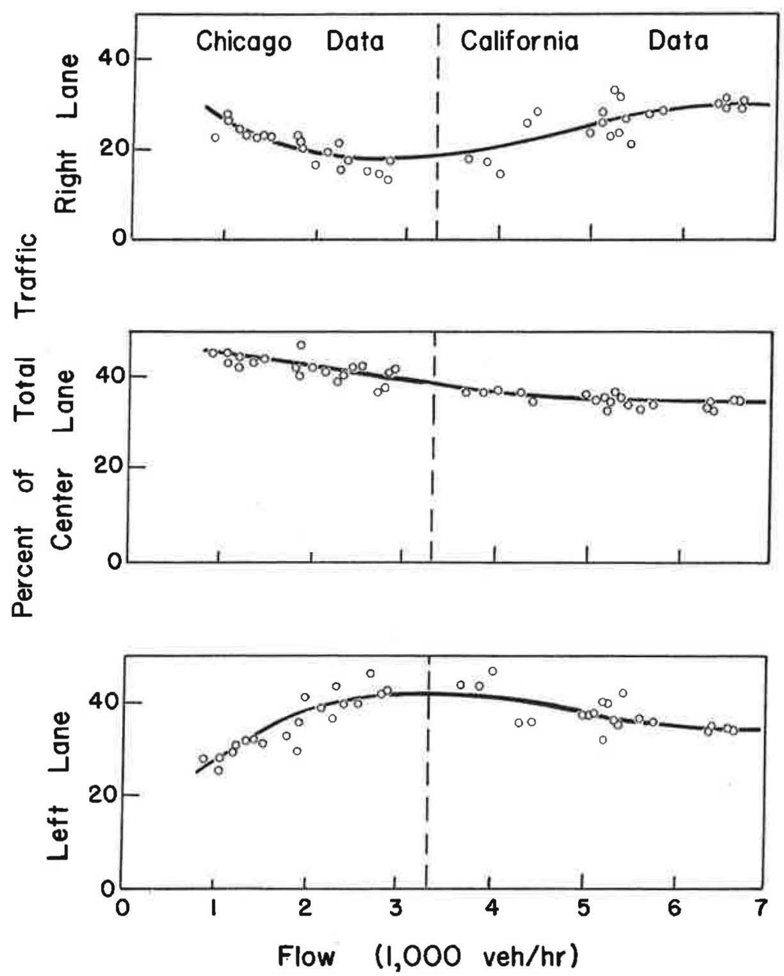



Figure 4.7 Distribution of traffic among lanes of a three-lane freeway.

pirical data from several three-lane freeways (*18*).

**4.3.2.4 Monte Carlo Methods.** The generation of any random phenomenon by a digital computer requires the cumulative probability distribution of the phenomenon. Figure 4.8 is an illustration of the generalized random phenomenon. Using a random fraction which has been previously generated, the distribution is entered along the probability axis and the corresponding value is obtained from the variable axis. Thus, the generation of random phenomena may be divided into two parts: generation of random fractions (having a uniform distribution), and conversion to random "deviates" (having the desired distribution). "Deviate" is the term used by the statistician to denote the departure from the mean. The following are Monte Carlo methods which are used for generating random phenomena.

*Generation of Random Numbers.*—Several investigators have found that the simplest and most reliable method for the generation of pseudo-random numbers by computers is as follows (*32*):

An assumed starting number is multiplied by an appropriate multiplier. The low order, or less significant, half of the product is taken as the random number. The second random number is formed by using the first as a starting number and the same multiplier as before, etc. This technique is usually stated as

$$R_m = \rho \, R_{m-1} \text{ Mod } b^n \qquad (4.5)$$

in which

$R_m$ = the $m$th random number;
$\rho$ = the multiplier;
$n$ = number of digits in a normal word on the particular computer used;
$b$ = number base of computer;
Mod $b^n$ = instruction to use only the low order or less significant half of the full ($2n$-digit) product (the remainder after dividing the product by $b^n$ the maximum integral number of times); and
$R_o$ = any odd number selected as a starting number.

The multiplier $\rho$ may be selected by taking a base which is prime relative to the number base of the computer and raising it to the highest power which can be held by one word of the computer. Table 4.2 gives a few appropriate multipliers.

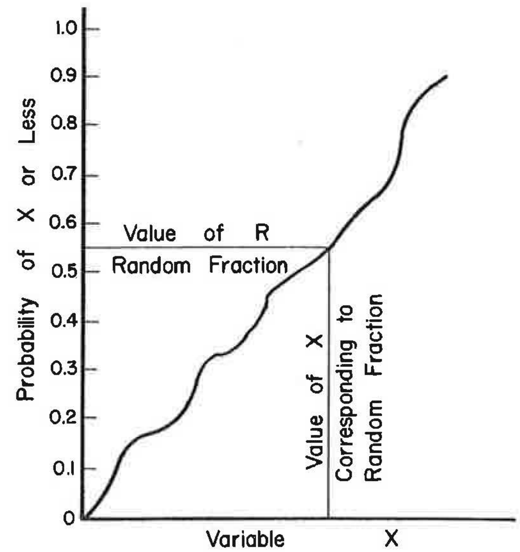



Figure 4.8. Generalized cumulative probability distribution.

The numbers resulting from this generation technique form a series of pseudo-random numbers; that is, the numbers are generated in a non-random manner, but behave as though they were random. Tests by several investigators indicate no evidence that the numbers are non-random. Random numbers so generated may be interpreted either as random integers with the point at the extreme right, or as random fractions with the point at the extreme left. If the random numbers are interpreted as fractions, the result is a rectangular or uniform probability distribution.

*Conversion to Desired Distribution.*—Conversion of random fractions to random

Table 4.2 Multipliers for Random Number Generation

| Number Base and Capacity of Computer, $b^n$ | Random Number Multiplier, $\rho$ | Typical Computer with This Base |
|---|---|---|
| $10^{10}$ | $7^{11}$ | IBM 650 |
| $2^{35}$ | $5^{13}$ | IBM 709 |
| $2^{36}$ | $5^{13}$ | |

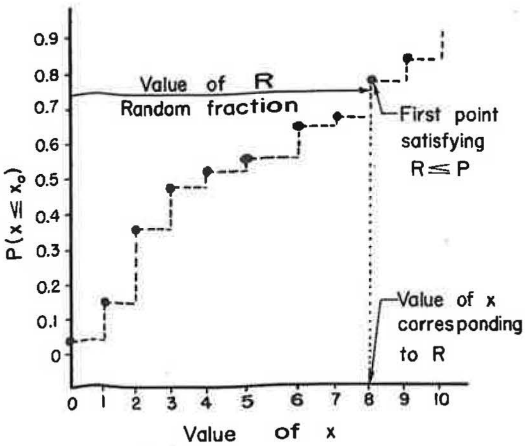



Figure 4.9. Cumulative probability distribution. Points represent tabular values; broken line, distribution for discrete values of $x$.

deviates of the desired form through the use of the cumulative probability distribution can be accomplished in a variety of ways. If the cumulative probability distribution is a continuous function, which can be represented by an equation, the operation of inserting the random fraction and obtaining the random deviate is purely a matter of calculation. In certain specialized cases it is faster, and may even be more accurate, to obtain the random deviates in an indirect trial-and-error, Monte Carlo method. Such a technique has been used by Gerlough (*25*, Appendix C) for generation of exponential arrivals.

If the cumulative probability distribution is represented not by an equation but by tabular data, the generation of the random deviates may be performed by a form of table-lookup operation. In Figure 4.9 tabular values of a cumulative distribution are indicated by points. The distribution for discrete values of the random deviate, $x$, is shown as a broken line. The random fraction is compared with the ordinates of the various points until the first point is found which satisfies the condition $R \leq P_i$, in which $P_i$ is the probability (ordinate) value of the $i$th point. The value of $x$ for the $i$th point is taken as the desired deviate.

If the random deviate $x$ is a continuous distribution, the tabular values may be interpolated along straightline segments between the points, obtaining a unique value of $x$ for each value of random fraction.

*Poisson Distribution.*—The usefulness of the Poisson distribution for describing various phenomena, including traffic, is discussed in Chapter 3. The cumulative Poisson distribution is expressed by

$$P(n) = \sum_{i=0}^{n} \frac{m^i e^{-m}}{i!} \qquad (4.6)$$

in which $P(x)$ is the probability of $n$ or fewer events during interval $t$, and $m$ is the average number of events during intervals of length $t$.

The generation of random deviates (arrivals, for instance) which follow the Poisson distribution must be carried out by a trial-and-error process. First a random fraction, $R$, is generated. The cumulative Poisson distribution is then formed, term by term, using Eq. 4.6. At each step the cumulation is compared to $R$. When the first value of $P(n)$ satisfying the relationship $P(x) \geq R$ is found, the corresponding value of $n$ is taken as the random variate (number of arrivals). If many Poisson events are to be generated, it may be more time saving to compute a table which lists the probability for each value of arrival. The number of arrivals is then obtained by a table look-up process. The table is entered with the random number (probability) as an argument, and the number of arrivals is read.

The flow diagram of a computer program for accomplishing this generation is given by Gerlough (*25*).

*Exponential Distribution.*—Many phenomena characterized by sequences of arrivals, as in traffic situations, may be treated by means of the exponential distribution

$$P(g \geq t) = e^{-t/\bar{h}} \qquad (4.7)$$

in which

$g$ = gap between successive arrivals, in time units;
$t$ = time, usually in seconds;
$\bar{h}$ = average time spacing between arrivals (may be thought of as the abscissa of the center of gravity of the area under the exponential curve);
$1/\bar{h}$ = flow (number of arrivals per unit time, the unit time being the same as that used for $t$); and
$P(g \geq t)$ = probability that $g \geq t$.

Eq. 4.7 expresses the probability that the spacing between arrivals is equal to or greater than the specified time. The cumulative exponential probability distribution complement of this relationship is

$$P(g < t) = 1 - e^{-t/\bar{h}} \qquad (4.8)$$

which is shown graphically in Figure 4.2. By considerations related to the Monte Carlo method, it is possible to simplify this expression to

$$P(t) = 1 - e^{-t/\bar{h}} \qquad (4.9)$$

in which $t$ is taken as the time spacing between arrivals.

Solving Eq. 4.9 for $t$ gives

$$t = -\bar{h} \log (1 - P) \qquad (4.10)$$

By substituting the random fraction, $R$, for $(1 - P)$, it is possible to solve for $t$. Flow diagrams for computer programs employing two different methods of solution are given by Gerlough (*25*).

*Shifted Exponential Distribution.*—When all vehicles are free-flowing—that is, can pass at will—the exponential distribution appears to describe time spacings adequately. However, when vehicles are flowing in platoons or are constrained so that they cannot pass at will, a modified solution must be used.

From observations it is known that there is a certain minimum headway, $\tau$, which can be maintained by vehicles. This may be stated: "The probability of a gap between successive vehicles of less than $\tau$ is zero." This phenomenon may be represented by an exponential curve shifted to the right by an amount $\tau$, or

$$P_2 = 1 - \exp\left[-(t - \tau)/\bar{h}_2\right] \qquad (4.11)$$

which is plotted in Figure 4.3 with $\bar{h}_2 = 1$. In Eq. 4.11, $\bar{h}_2$ is the average time spacing measured from the point where the curve intersects the $t$ axis. The average spacing between vehicles, $t_2$, is the average time spacing measured from origin.

Thus,

$$t_2 = \bar{h}_2 + \tau$$

or

$$\bar{h}_2 = t_2 - \tau$$

so that

$$P_2 = 1 - \exp\left[-(t - \tau)/(t_2 - \tau)\right]$$

Solving Eq. 4.11 for $t$ gives

$$t = \bar{h}_2\left[-\log(1 - P)\right] + \tau \qquad (4.12)$$

Again, substituting the random fraction, $R$, for $(1-P)$ permits solution for $t$, once values are available (or assumed) for $\bar{h}_2$ and $\tau$.

### 4.3.3 Implementation of Intersection Model

Formulation of the model beyond the basic block diagram state is determined by the type of computer to be used for implementation of the model. The next two sections explain how the block diagram of Figure 4.1 is implemented for use in analog and digital computers.

**4.3.3.1 Analog Computer Implementation of Intersection Model.** The following description of analog simulation of the intersection is taken from Mathewson, Trautman and Gerlough (*16*).

"For approximate studies to give general effects in a large network, it may be feasible to use a continuous-variable computer. A suitable computer for this purpose may be built around the operational-amplifier type of integrator, which may be used as the storage element of a wait circuit as shown in Figure 4.1. Figure 4.10 shows the wait and the direction selector sections. In this model individual vehicles are not recognized. Instead, more gross effects are considered. It is not necessary, therefore, to provide separate right-turn and left-turn waits. These effects may be lumped into the initial wait, and provided for by the response of the Output Rate Generator to the various interferences.

"In Figure 4.10, relay 1 operates during red. Relay 2 operates when there are cars waiting at the intersection. Relay 3 operates when there is interference. Thus, if the signal is green, if no vehicles are waiting and if there is no congestion, there is a direct connection between the input and output, and there is no voltage on either input to the integrator. If there is interference from opposing traffic or from pedestrian (relay 3 energized), or if there is delay caused by cars waiting to enter the intersection (relay 2 energized), the input is fed to 'storage' in the integrator. When

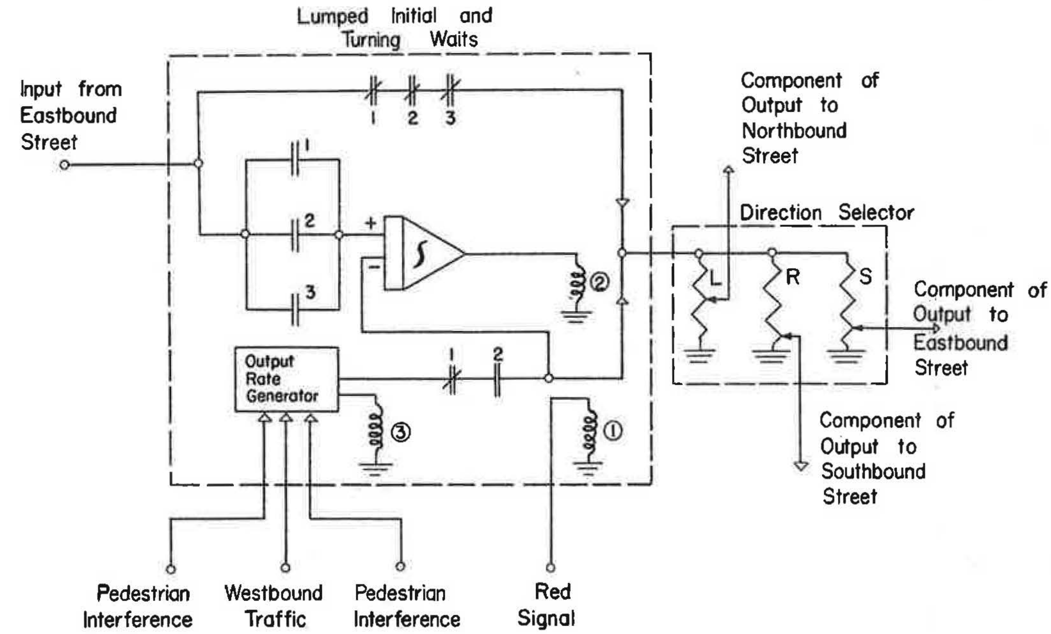



Figure 4.10. Portion of continuous-variable simulator, showing elements for representation of traffic entering from the west (see Fig. 4.11 for notation).

the green appears (relay 1 de-energized), traffic is released (subtracted from 'storage') at a rate in approximate inverse proportion to the interference, as established by the Output Rate Generator.

"The streets between intersections may be simulated by means of magnetic tape delays. These consist of a magnetic tape having recording and playback heads operating continuously. The time for the tape to travel from the recording head to the playback head constitutes the travel time of the vehicles."

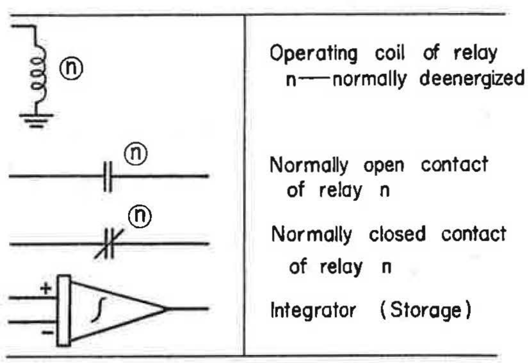



Figure 4.11. Symbols for continuous-variable simulator of Figure 4.10.

4.3.3.2 Digital Computer Implementation of Intersection Model. The following description is from Goode, *et al.* (*11, 13*).

"For digital simulation, the intersection is assumed to consist of a crossblock (Figure 4.12) having streets 22 ft wide and lanes approximately 400 ft long. The vehicles traveling through it are assumed to average 18 ft in length, being in any case more than 11 ft and less than 22 ft long. They travel in the lanes at 30 mph when not obstructed, pass through the intersection under the control of a traffic signal, and turn right, left, or go straight ahead according to the 'desires of the driver.' These desires are determined by a random selection procedure. The position of a car in the block is given by the position of the midpoint of the front bumper. To simplify the model, cars are not allowed to pass one another, and interferences from parked cars and pedestrians are assumed to be negligible.

"Each intersection is to consist of a crossblock, two of which are shown in Figure 4.12. Each approach of this crossblock includes, in addition to the intersection area, an approach path from the previous intersection. The two-dimensional strip is idealized to a line, and the position of the car in the lane is idealized to a corresponding point on the line. A lane is then a sequence of 35 points in which a car jumps from point to point. Along the approaches, the points are approximately 11 ft apart. Within the intersection the points are closer together to permit slower movement of the cars.

"In this model, each lane entering the intersection is broken into four segments or paths lying within the intersection: one traversed by cars turning right, one by cars going straight, and two traversed in succession by cars turning left. These paths are called $\rho$, $\alpha$, $\lambda$ and $\bar{\lambda}$, respectively. The paths, also, are considered to be sets of points and are shown for a single lane in Figure 4.13. The end point of $\lambda$ (point 9) is of special importance and is called the left-turn zone.

"Cars move down the idealized lanes and paths by jumping from one point to the next. When a car is moved, it jumps and, thereby, covers the distance between two adjacent points every quarter second.

"In the computer, each lane is represented by a binary register. The points of the lane, or intersection path, are associated one-to-one with digit position in the corresponding registers. To represent the distribution of cars in the model at a particular instant of time, it is only necessary to specify the presence or absence of a car for each point of the model. This is done by having ones in the digit position corresponding to points at which there is a car, and zeros otherwise. To move a car from one point to the next, the digit concerned is extracted to determine whether a car is present. If one is present, then the logical rules govern its position in the scheme of things.

"Cars enter a given approach at a point 40 (Figure 4.13). They are generated by a process making use of psuedo-random numbers. At the end of each quarter-second interval, a random number subroutine generates a number between 0 and 1 for lane S, say. This number is then compared to the number $m$*. If it is less than $m$, a car is generated for lane S. If it is greater than $m$, no car is generated. Thus, by changing the value of $m$ the average number of cars per hour entering the lane can be controlled. The resulting distribution approximates the Poisson.

* $m$ is the average number of cars arriving per quarter second. (An alternate statement is that $m$ is the fraction of all quarter seconds which contain cars.)

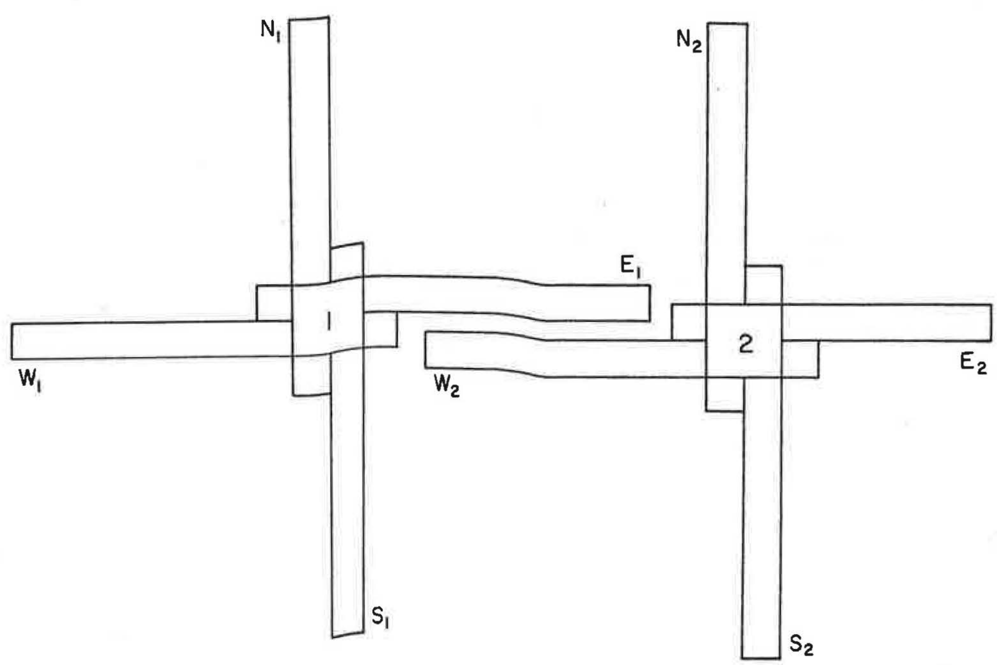



Figure 4.12. Crossblock representation of two intersections.

"In order to avoid piling cars on top of one another, which would occur whenever cars are generated at successive quarter-second intervals, cars are first put in a register known as the backlog. In this register, cars are merely counted and not put in relative positions. The contents of the register indicate the number of cars waiting to enter the lane at a point remote from the lane. As space becomes available, cars are moved from the back into the lane S. Similar procedures are carried out for the other three lanes.

"When cars leave the end-point of a path in the intersection, they pass through the exit block ($E_1, E_2$) and are dropped from consideration in the single-intersection model.

"At an actual intersection, an observer of traffic cannot tell which way a particular car will turn. However, he may know the probabilities for a right turn, left turn, or for going straight ahead. This characteristic of traffic is simulated by associating a turn register with each lane. It can be thought of as representing the turn indicator of the car nearest the intersection in that lane. If the turn register contains a 1, the car nearest the intersection will turn right. A 0 indicates straight ahead; a 2, a turn left. After the car has made its turn, the turn register must be set to indicate the turn condition of the next car. A mode similar to that of generating cars is used to generate the random numbers which decide whether turns take place.

"The traffic signal is simulated in the computer by a light register. The register contains a 0 if the light is red, a 1 if the light is green, and a 2 if the light is amber. The duration of each portion of the light is controlled by counting the number of quarter seconds during which it has been continuously in that phase, and changing to the next indication when the counter has reached the specified value. The duration of red and green for the north-south and east-west lanes are parameters of the program and can be set to any value desired. The duration of amber, however, is fixed.

"Behind the specific rules the computer obeys are the following general principles:

"1. Cars approaching the intersection give the right-of-way to cars which are in the intersection, but not in the left-turn zone.

"2. Cars in the left-turn zone give right-of-way to cars which will cross their paths.

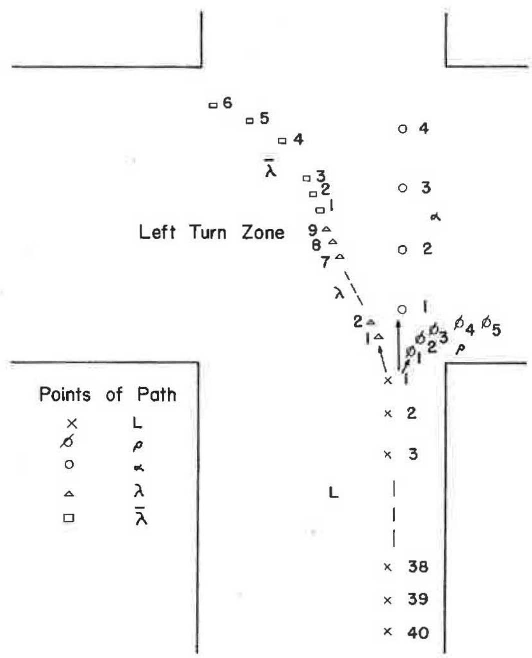



Figure 4.13. Computer "crossblock."

"The cars on $\alpha$, $\rho$, and $\bar{\lambda}$ (Figure 4.13) are first considered and are moved up one point. (All cars travel at 30 mph.) This can be done without any consideration of the light or traffic because cars are not allowed to enter these paths until the way is clear for their complete traverse.

"A car approaching the left-turn zone is likewise automatically moved up. If there is a car, A, in the left-turn zone, the light is checked. If it is red, A completes the turn. If the light is green or amber, the computer examines the right- and straight-ahead-paths of the opposing lane (*e.g.*, lane $N_1$ if A is turning left from lane $S_1$, Fig. 4.12). If any of these paths contains a car, A remains where it is. If they are empty, the traffic in the opposing lane is examined. If there is no car within 55 ft of the intersection (that is, if there are zeros corresponding to the first five points), the car continues its left turn. If there is a car within 55 ft, the turn register for the car's lane is examined. If it indicates that the nearest car is to turn left, A completes the turn; otherwise, it remains in the left-turn zone.

"If the car is at point 1, and ready to enter the intersection, the light is checked. If the light is red, the car remains at point 1; if it is green, the computer examines the turn register. If a right turn is indicated, the right-turn path and the paths intersecting it are examined, and if clear, the right-turn path is entered. The digit corresponding to point 1 is made zero, and the digit corresponding to the first point of the right-turn path is made 1; that is, the car turns right. If the left turn is specified, the left-turn path and the paths intersecting it are examined. If empty, the car proceeds; otherwise, it remains at point 1. Similar treatment is given to the 'straight ahead' direction.

"A car at point 2 of a lane always moves to point 1, and a car at point 3 advances unless there is a car at point 1 or a car entering the intersection from this lane. These facts are determined by checking to see if 0's or 1's are associated with the points of these intersection paths.

"Cars farther back in the lanes follow rules designed to maintain a distance of at least 55 ft between the front bumpers of moving cars.* This is deemed a reasonable minimum distance for cars traveling at 30 mph. A moving car will approach a stopped car until there is a distance of 22 ft between their front bumpers.

"A measure of effectiveness used in the one-intersection model was the average delay experienced by cars at the intersection. Inasmuch as the minimum time for negotiating the course was known, the average delay for cars in a given lane was obtained by finding the average actual time needed to go from the far end of the lane (point 40) through the intersection, and subtracting the minimum time for negotiating the course. The average time needed to pass through the lane and intersection is approximated by counting the 1's (that is, the cars) in the lane and its associated intersection path every quarter second. These counts are totaled and the sum divided by the number of cars leaving the lane's intersection paths. Parameters are varied to study the average delay as it is affected by increase in right or left turns, changes in the light cycle, and changes in the rate of cars being generated."

It is thus evident that digital simulation of traffic flow can include more details of the behavior of individual cars than is possible with analog simulation.

### 4.3.4 Simulation of Freeway Traffic Flow

The formulation of a model for freeway traffic flow must include a description of system behavior in terms of rules of the road and provide methods for the implementation of these rules within a computer.

**4.3.4.1 Rules of the Road.** One possible set of rules for a four-lane divided freeway in a section without interchanges is as follows:

(a) Each vehicle proceeds in either the right or left lane at its desired speed or the maximum allowable speed until it encounters another vehicle in the same lane. Encountering consists of coming so close that during the next increment of time the spacing between the encountered and encountering vehicles would become less than safe spacing.

* It is possible to select other rules for spacing of the cars. In this early work of Goode's, uniform distance headway was selected because it could be easily implemented.

(b) On encountering another vehicle in the same lane, the encountering vehicle, if it is in the right lane, examines the lane to its left. If the encountering vehicle is in the left lane, it examines the lane to its right. A lane change is made if it is safe to do so.* If possible, the encountering vehicle maintains its same speed after the lane change. If a change of lane is not safe, the encountering vehicle decreases its speed to that of the encountered vehicle. The deceleration may be at a specified rate or selected from a probability distribution.

(c) During each time increment, all vehicles in the left-hand lane look for opportunities to move to the right.

(d) During each time increment, all vehicles traveling at speeds less than their desired speeds look for opportunities to increase their speeds.

**4.3.4.2 Representation.** When the flow of any type of discrete objects is considered, there are two general ways in which the objects may be represented for purposes of simulation. With the first method, physical representation, one or more binary digits are assigned to represent the presence, position and, perhaps, size of the item or vehicle to be simulated. Certain areas of the computer memory are assigned and organized in such a way as to represent the flow network, and the groups of binary digits representing the items are caused to flow in the network by suitable manipulations. This technique is primarily useful with binary computers. The use of physical representation within the computer is complex and slow. Techniques for this process are discussed in the literature (*5*).

The second technique, memorandum representation, consists of recording all conditions pertaining to a given vehicle. Usually this can be done by using a single coded word, the parts of which can be extracted and interpreted by suitable computer routines. This technique is applicable to computers of any number base.

The memorandum method is easier to understand and to program for the computer than is the physical method. It also requires less computer time. The status of each vehicle is kept in the memory circuit of the computer. The data for each vehicle must include position (distance from starting or reference point), lane, desired speed, actual speed, and turn requirements (points at which turns are to be made) if there are ramps within the simulated test section. Other desirable information would include length or class of vehicle, normal acceleration and deceleration rates, passing characteristics of drivers, and time at which each vehicle entered the system.

Distance along the roadway is quantified, using a unit known as a "unit block." This is usually a fractional part of the length of a vehicle. Time is also quantified.

One computer word is frequently assigned to each vehicle, as shown in Figure 4.14, to conserve computer storage space.

A register, or memory cell, of the computer is used as a clock counter. The clock counter is advanced at each simulation cycle, and the memoranda of vehicles and other data are scanned to determine whether:

(a) It is time for a new vehicle to enter the system.

(b) Encountering will take place during the next unit of time.

(c) It is safe to pass.

(d) It is time to pull to the right.

The time for vehicle entry is determined by generating an inter-arrival time, a process discussed in section 4.3.2.4. This time interval is added to the time of arrival of the previous vehicle to determine when the next vehicle should enter. Comparison with the clock counter determines whether it is time for a new vehicle to enter the system. When a new vehicle does enter the system,

* In simplified simulations, a lane change would always be made if it is safe to do so. In more sophisticated simulations, a random selection procedure can be used to determine if a permissible lane change is actually made.

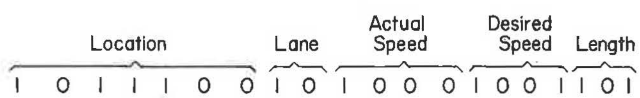
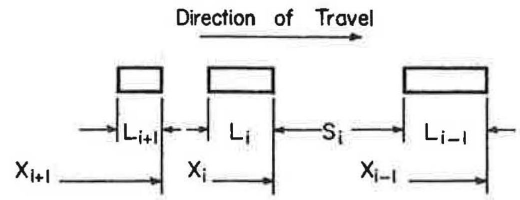



Figure 4.14. Typical arrangement of information in code word.




$X_i$ = Distance of ith car from reference (at time t);

$L_i$ = Length of ith car;

$u_i$ = Speed of ith car (at time t);

$S_i\big|_t$ = Space ahead of ith car (at time t) $= X_{i-1} - L_{i-1} - X_i$; and

$S_i\big|_{t+\Delta t}$ = Estimated space ahead of ith car at time $(t + \Delta t)$

$$= \left[ X_{i-1} + u_{i-1} \Delta t \right] - L_{i-1} - \left[ X_i + u_i \Delta t \right]$$

Figure 4.15. Method of testing for safe distance between vehicles.

it is assigned a desired speed, vehicle type, turn requirements, etc., by reference to the appropriate distributions.

Techniques such as those shown in Figure 4.15 can be used to determine whether encountering will take place during the next unit of time. Similar manipulations can be developed for answering other questions.

Once the questions have been answered, each vehicle can be advanced by changing the record to show its position one time unit later. This is done by multiplying the vehicle's speed by the length of the unit of time and adding the product to present position. The recording of data which serve as a measure of system performance is an important part of each cycle.

The memorandum method may be varied by allocating one memory cell to each unit block (*3*). If a vehicle is in a given unit block, the data concerning the vehicle may be stored in the corresponding memory cell. Thus it is not necessary to otherwise record the position and lane of the vehicle.

4.3.4.3 Scanning. It is not possible to examine, within the digital computer, all parts of the system simultaneously. Furthermore, time is divided into discrete units. For these reasons, some method of scanning must be used.

There are two general scanning methods. "Periodic scanning" consists of scanning and up-dating the entire system once during

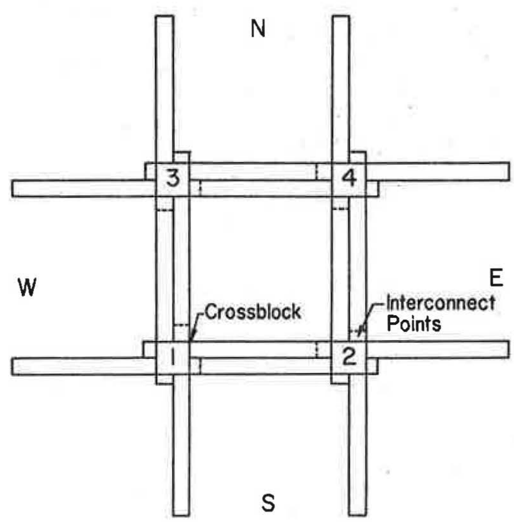



Figure 4.16. Intersection network formed from four crossblocks.

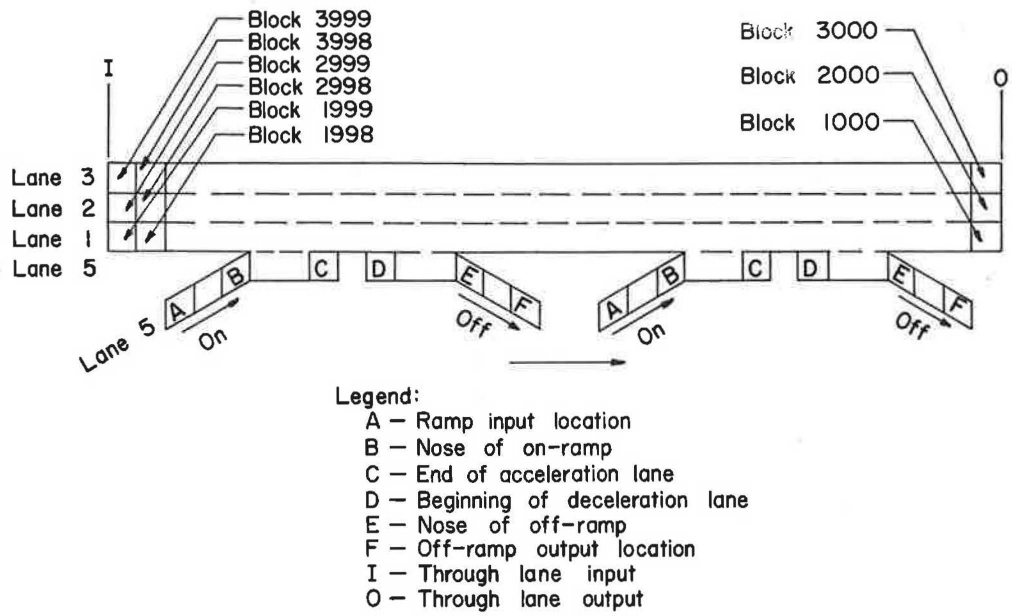



**Figure 4.17.** Organization of roadway layout for programming of computer simulation.

each unit of time. This technique is straightforward and is usually easy to program. "Event scanning" consists of determining the next event of significance by extrapolation and moving the clock to this next event without any intervening scans. This method increases computing speed by a factor of about 10, but it usually requires greater program complexity (*22*).

There are many variations or combinations between these two extremes of scanning. It is possible, for instance, to incorporate event scanning into certain portions of a program if it is basically written in the periodic scan mode.

## 4.4 EXAMPLES OF TRAFFIC STUDY BY SIMULATION

Published simulation work to date consists of studies of traffic flow on the following types of facilities:

*Signalized Streets*
- Individual intersection with single-lane approaches.
- Small network of intersections.

*Freeways*
- Section isolated from entrances and exits.
- Section involving two entrances and two exits.

*Tunnels*
- Lane changing not permitted.

Four of these studies are summarized in the following sections.

### 4.4.1 Signalized Intersection Delays

Webster (*34*) studied delays at isolated signalized intersections. He developed methods for predicting the average delay of vehicles on each approach and set forth manual procedures for determining the condition of minimum delay.

### 4.4.2 Network of Intersections

Goode and True (*13*) simulated a network involving four intersections. The model consisted of four crossblocks connected together as shown in Figure 4.16. Simulation runs were made on an IBM 704 equipped with a cathode ray tube display unit, providing a visual output. It was possible to take a motion picture, which showed traffic moving forward, stopping for a red light, moving through an intersection on the green, and continuing to another intersection. This motion picture did much to promote the acceptance of simulation among traffic engineers.

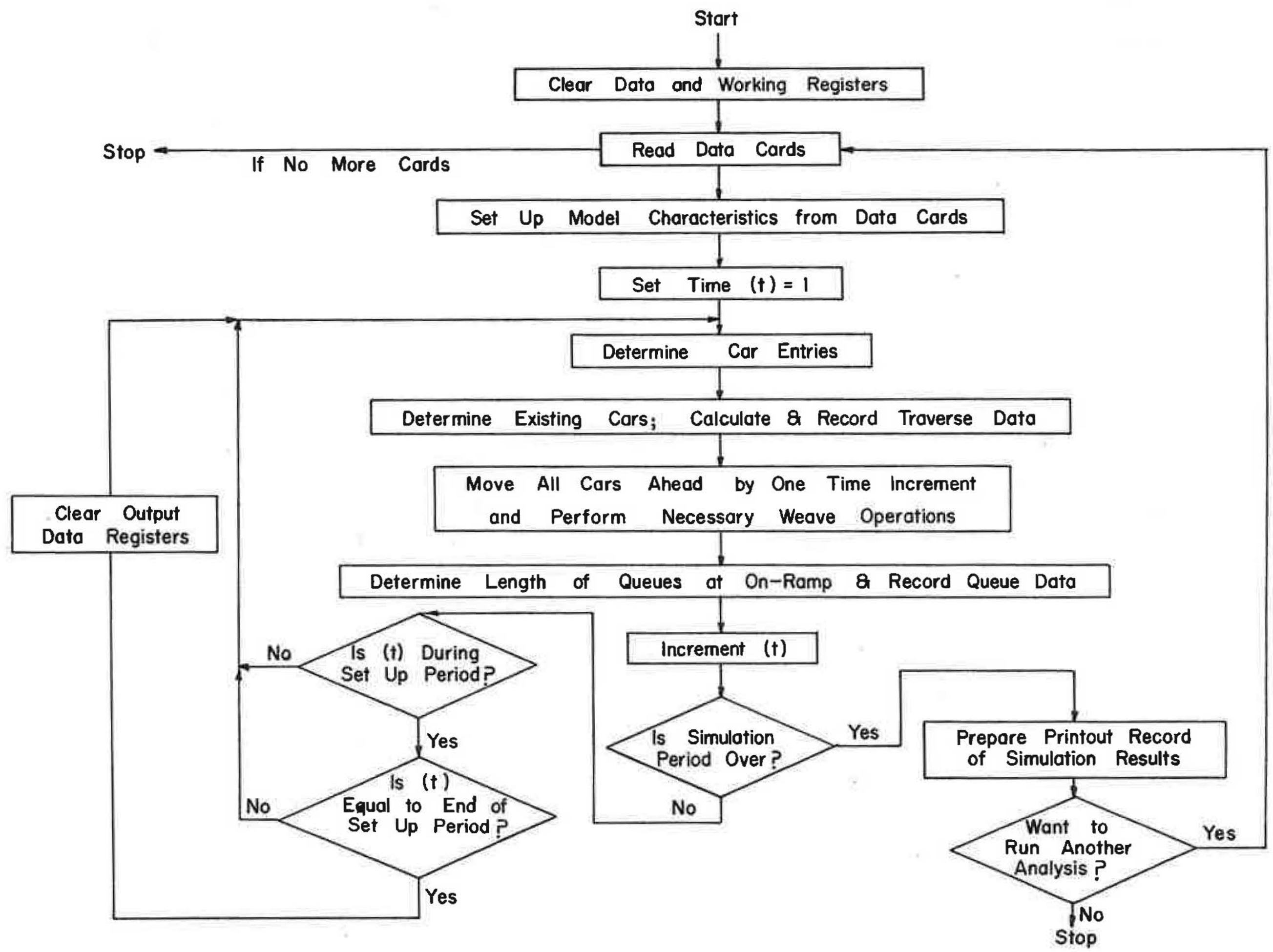



Figure 4.18. Flow diagram of over-all computer logic.

### 4.4.3 Freeway Ramp Spacing

Most of the work on simulation of automobile traffic flow in the United States has involved studies of techniques.

Levy and his associates (*18, 8, 9*), however, have published work which may advance techniques to the point of usefulness for design purposes. They have developed a simulation model which will produce traffic flow on a 17,000-ft section of freeway, including two on-ramps and two off-ramps. Alternate design criteria may be studied with this simulation model. It is also possible to examine the effect of various spacings between ramps and to investigate the effect on freeway operation of various parameters such as speed distribution, and weaving.

Figure 4.17 shows the organization of a roadway layout used in their studies. Figure 4.18 shows the over-all layout for the various portions of the computer program. The answers to many problems may be found through this simulation (*8, 9*).

### 4.4.4 Simulation of Tunnel Traffic

Helly (*15*) has considered the case of traffic moving through a tunnel in which no lane changing is permitted. This enabled him to concentrate on the relationship of vehicles in the same lane. Without lane changing, the simulation is simplified, and one vehicle can be followed through the tunnel, storing a complete record of its trajectory. The next vehicle is then followed through, with the trajectory of the previous vehicle being used as a reference.

Helly used the theoretical work on car following presented in Chapter 2 and assumed that the driver attempts to minimize the difference between his actual headway and his desired headway with respect to the car ahead. He was then able to develop an equation which gives the acceleration of the $n$th car at any time $t$. Using experimental data from other references, he established values for various driving parameters.

Helly's objective was to study the flow of vehicles through bottlenecks in a tunnel. He defined two types of bottlenecks. In one type, the increase of headway within the bottleneck is the same for all vehicles, regardless of position in the platoon.

In the other type, the increase of headway within the bottleneck varies from one vehicle to another according to their relative position within the platoon. Helly reasoned that bottlenecks of the first type occur when gradual changes in desired headways take place. Bottlenecks of the second type occur when there is a sudden change in the desired headway. He demonstrated this difference by simple simulation runs and found that bottlenecks of the first type are rare. He concentrated, therefore, on bottlenecks of the second type, with emphasis on situations in which there is a localized acceleration limit.

## REFERENCES

*Simulation of Traffic Flow*

1. Beckmann, M., McGuire, C. B., and Winsten, C. G., *Studies in the Economics of Transportation.* Yale Univ. Press (1956).
2. Benhard, F. G., "Simulation of a Traffic Intersection on a Digital Computer." M. S. Thesis, Univ. of California, Los Angeles (June 1959).
3. Dunzer, J. B., "Simulation of Freeway Traffic on an IBM 704." Report submitted in Engineering 104D, Univ. of California, Los Angeles (May 27, 1958).
4. Gerlough, D. L., "Analogs and Simulators for the Study of Traffic Problems." Proc. Sixth California Street and Highway Conf., pp. 82-83 (1954).
5. Gerlough, D. L., "Simulation of Freeway Traffic by an Electronic Computer." *HRB Proc.,* 35: 543-547 (1956).
6. Gerlough, D. L., and Mathewson, J. H., "Approaches to Operational Problems in Street and Highway Traffic." *Oper. Res.,* 4: 1. 32-41 (Feb. 1956).
7. Gerlough, D. L. "Simulation of Freeway Traffic by Digital Computers." Proc. of Conf. on Increasing Highway Engineering Productivity, Georgia Inst. of Technology, July 9-11, 1956, U. S. Bureau of Public Roads (1957).
8. Glickstein, A., Findley, L. D., and Levy, S. L., "A Study of the Application of Computer Simulation Techniques to Interchange Design Problems." HRB Bull. 291, pp. 139-162 (1962).
9. Glickstein, A., and Levy, S. L., "Application of Digital Simulation Techniques to Highway Design Prob-

lems." Proc. of Western Joint Computer Conf., pp. 39-50 (1961).
10. Goode, H. H., "Simulation—Its Place in System Design." *Proc.*, Inst. Radio Eng., 39: 12, 1501-1506 (Dec. 1951).
11. Goode, H. H., Pollmar, C. H., and Wright, J. B., "The Use of a Digital Computer to Model a Signalized Intersection." *HRB Proc.*, 35: 548-557 (1956).
12. Goode, H. H., "The Application of a High-Speed Computer to the Definition and Solution of the Vehicular Traffic Problem." *Oper. Res.*, 5: 6, 775-793 (Dec. 1957).
13. Goode, H. H., and True, W. C., "Vehicular Traffic Intersections." Paper presented at 13th National Meeting, Assoc. of Computing Machinery (June 11-13, 1958).
14. Helly, W., *Dynamics of Single-Lane Vehicular Traffic Flow.* Res. Rpt. No. 2, Center for Oper. Res., Mass. Inst. of Tech. (Oct. 1959).
15. Helly, W., "Simulation of Bottlenecks in Single Lane Traffic Flow." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 207-238 (1961).
16. Mathewson, J. H., Trautman, D. L., and Gerlough, D. L., "Study of Traffic Flow by Simulation." *HRB Proc.*, 34: 522-530 (1955).
17. Naar, J., "Simulation of Vehicular Flow." Thesis for degree of Civil Eng., Mass. Inst. of Tech. (June 1958).
18. Perchonok, P. A., and Levy, S. L., "Application of Digital Simulation Techniques to Freeway On-Ramp Traffic Operations." Final Report to Bur. of Public Roads, Midwest Research Inst. (Nov. 1959); *HRB Proc.*, 39: 506-523 (1960).
19. Trautman, D. L., Davis, H., Heilfron, J., Ho, E. C., Mathewson, J. H., and Rosenbloom, A., *Analysis and Simulation of Vehicular Traffic Flow.* Res. Rpt. No. 20, Inst. of Transp. and Traffic Eng., Univ. of California (Dec. 1954).
20. Wohl, M., "Simulation—Its Application to Traffic Engineering." Part I, *Traffic Eng.*, 30: 11, 13-17, 29 (Aug. 1960); Part II, *Traffic Eng.*, 31: 1, 14-25, 56 (Oct. 1960).
21. Wong, S. Y., "Traffic Simulator With a Digital Computer." Proc. of Western Joint Computer Conf., pp. 92-94 (1956).

*Simulation of Flow of Discrete Objects*

22. Gerlough, D. L., "A Comparison of Techniques for Simulating the Flow of Discrete Objects." Paper presented at National Simulation Conf., Dallas, Tex. (Oct. 23-25, 1958).
23. Moore, C. J., and Lewis, T. S., "Digital Simulation of Discrete Flow Systems." Commun. Assoc. for Computing Machines, 3: 12, 569-660, 662 (Dec. 1960).

*Generation of Traffic Inputs*

24. Gerlough, D. L., "The Use of the Poisson Distribution in Problems of Street and Highway Traffic." *Poisson and Traffic,* The Eno Foundation for Highway Traffic Control, pp. 1-58 (1955).
25. Gerlough, D. L., "Traffic Inputs for Simulation on a Digital Computer." *HRB Proc.*, 38: 480-492 (1959).
26. Greenshields, B. D., Shapiro, D., and Ericksen, E. L., *Traffic Performance at Urban Street Intersections.* Tech. Rpt. No. 1, Yale Bureau of Highway Traffic (1947).
27. Greenshields, B. D., and Weida, F. M., *Statistics Included in Application to Highway Traffic Analyses.* The Eno Foundation for Highway Traffic Control, p. 238 (1952).
28. 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).
29. Kell, J. H., "A Theory of Traffic Flow on Urban Streets." Proc. 13th Ann. Meeting, Western Section, Inst. of Traffic Eng., pp. 66-70 (1960).
30. Schuhl, A., "The Probability Theory Applied to Distributions of Vehicles on Two-Lane Highways." *Poisson and Traffic,* The Eno Foundation for Highway Traffic Control, pp. 59-75 (1955).
31. Hastings, C., *Approximations for Digital Computers.* Princeton Univ. Press (1955).

32. Taussky, O., and Todd, J., "Generation and Testing of Pseudo-Random Numbers." Symposium on Monte Carlo Methods held at Univ. of Florida, Mar. 16-17, 1954. Edited by H. A. Meyer, Wiley and Sons (1956).
33. Von Neumann, J., "Various Techniques Used in Connection with Random Digits." *Monte Carlo Method*, Nat. Bur. of Standards, Applied Mathematics Series, No. 12 (June 11, 1951).

*Results of Simulation*

34. Webster, F. V., *Traffic Signal Settings*. Road Res. Tech. Paper No. 39, H. M. Stat. Office, London (1958).

Chapter 5

# SOME EXPERIMENTS AND APPLICATIONS

R. S. Foote
The Port of New York Authority

## Contents

*Section*

# Chapter 5

# SOME EXPERIMENTS AND APPLICATIONS

## 5.1 INTRODUCTION

In contrast to the extensive theoretical study of road traffic reviewed in preceding chapters, relatively little scientific experimentation has been performed. Experimentation to evaluate a formal theoretical model has been limited, and still fewer experiments have been made to test a prediction based on such theory by altering the conditions of traffic flow.

The qualifications of "scientific," "theoretical model," and "prediction based on such theory," are necessary to distinguish the experimentation which is the subject of this chapter from the experimentation which has provided most of the present knowledge in the field of traffic engineering. As it exists today, the knowledge which traffic engineers draw upon to solve problems and to design roads is largely empirical knowledge, built up through observations, measurements, and statistical analyses. There is a large body of such experimental knowledge, but that experimentation is beyond the scope of this chapter.

### 5.1.1 Background

The basically empirical approach, relying on collection of large amounts of data and statistical analyses, has not been able to answer satisfactorily many of the most important questions about road traffic: Why does one traffic lane regularly carry twice as much traffic as another before congestion occurs? What causes traffic to move in an accordion-like fashion? Will this bridge be a bottleneck if shoulder widths are reduced to 5 ft? Even the more powerful tools of scientific analysis, which rely on an interaction between theory and experiment, are challenged by the complexity of the road traffic problem.

It is likely that the answers to these and similar questions will be developed and refined through a high degree of interaction between theoretical and experimental work. Experiments alone have not provided the answers; theoretical work by itself is also unlikely to produce detailed workable answers. Both theory and experiment are needed for the development of a scientific understanding of road traffic. Collaboration and understanding among theoreticians, traffic engineers and roadway operators are essential to advance the solution of road traffic problems.

There are three immediate factors which undoubtedly tend to limit the use of traffic experimentation for scientific purposes. First is the inaccessibility of experimental situations for most traffic theorists. Most of the authors of papers on road traffic theory are mathematicians, physicists and other academicians and researchers who do not have ready contact with road traffic operating agencies.

A second difficulty has been the expense of instrumentation for traffic experiments. Although the expense of instruments to measure and understand road traffic is small compared with the expense of roads to carry traffic, the cost of adequate instruments is excessive in light of funds available for basic scientific research. The third factor which has limited the extent of traffic experimentation for scientific purposes is the time usually needed to reduce data for analysis. A shortage of manpower has made it impossible to obtain full value from experiments conducted for scientific purposes.

Important gains are being made to meet these three major needs: accessibility to an experimental situation, better instrumentation, efficient techniques for data reduction. In addition to increasing interest in traffic theory (this publication is one example), applications of new operating systems based on road traffic theory can provide theoreticians with permanent laboratories. Great strides are being made in the general field of instrumentation, and improved devices are now becoming available to the traffic researcher. Instruments which furnish data in a form suitable for automatic conversion and processing on electronic computers make possible analysis which would be excessively burdensome if manpower were needed for data reduction.

With these gains, and, more importantly, with increasingly better theoretical work, experimentation is likely to become a more important source of scientific knowledge about road traffic.

### 5.1.2 Comparison of Experiments and Applications

Traffic engineering experiments are conducted in a search for knowledge. However, the primary purpose of an application is to improve a traffic operation.

Due to limited traffic knowledge today, it is probable that all straightforward repetitive applications of road traffic theory will produce important contributions to traffic engineering techniques. Any application will be in large measure a continuing experiment which will at first produce considerable new knowledge.

Although there have been few applications of road traffic theory in the past, those made to date are an important source of knowledge for traffic engineers and roadway operators. Applications of road traffic theory are increasing, and it is likely that still more will be made in the next few years.

With both experimentation and application contributing to traffic knowledge, it is clear that the principal difference between the two is merely the length of time involved.

## 5.2 INSTRUMENTATION

In recent years there has been considerable progress in the development of traffic instrumentation systems for measuring driver behavior and vehicular movement. These systems, which utilize improved equipment and new techniques, have become essential in the study of complex traffic flow problems. This section discusses some of the techniques which have been developed.

### 5.2.1 Photographic

Movie cameras offer a relatively inexpensive way to record a large amount of data, but the manpower needed to convert these qualitative data to numerical form for further analysis is often excessive. Cameras are probably best suited when complex traffic movements must be analyzed by using the speeds and paths of individual vehicles. In such a case, securing this information with other equipment (such as transducers and recorders) would be highly expensive. An excellent camera application for this purpose was made by Capelle and Pinnell (*4*), who analyzed the capacity of signalized diamond interchanges. They used a 16-mm camera at 10 frames per second. Pictures of interchanges were taken from a truck with a tower 35 ft high. A special projector was used to observe vehicle movements and identify frame numbers.

Greenshields (*22*), who recognized early the usefulness of photography in traffic study, devised special assemblies which adapted a 35-mm movie camera for pulse operation and included reference data in the field of view.

Pulse operation was achieved by a solenoid-operated clutch that advanced the spring-wound camera one frame per pulse. Thus, used as a time-lapse camera, less film was required and the camera could be synchronized with the operation under study. For example, shooting at 88 frames per minute, each foot of advance along the roadway made by a vehicle in succeeding frames was equivalent to a speed of 1 mph; that is, a vehicle that advanced 20 ft in succeeding frames moved at 20 mph. The addition of reference data (such as time, frame number, event) greatly increased usefulness of the camera. However, the photographic equipment available at the time of Greenshields' early experiments required a relatively crude system of external lenses and was limited to use with camera lenses of short focal length.

In 1957, Forbes *et al.* (*16*) relied heavily on the ability of a camera to record reference data directly on each exposed frame. They used a 35-mm movie camera specially adapted by a series of internal lenses to display on each frame a variety of information. In their experiment, which is described in Section 5.3.1, they displayed a watch, speedometer, and six lights indicating various driver actions on the accelerator and brake pedals. The camera, operated at a rate of 24 frames per second inside the third car of a three-car platoon, viewed the position of the first two cars in relation to the third car. Positions of the first two cars were determined by parallax. Films were projected on a grid, and the distance of the first and second cars from the third was determined by comparing the apparent distance between marker lights on each car with the known actual distance. Reduction of data by this means required an excessive amount of labor.

The increasing use of cameras as a study instrument has resulted in the development of highly flexible and precise equipment. A 35-mm camera driven by an electric motor is now available. It can be pulse operated and used for time-lapse photography. The camera also contains an internal data chamber for placing reference information on part of each exposed frame. This equipment has been used for the study of wave action in traffic streams moving over the George Washington Bridge (*39*).

When mounted in a tower of the bridge and aimed straight down at traffic passing on the roadway below, the camera facilitated studying a number of traffic relationships.

Continuous-strip aerial photography (*48, 42, 38*) enables the greatest coverage of ground area for traffic surveys. This technique makes it possible to inventory individual vehicle speeds and densities existing at the time of the flight over a length of roadway. The technique uses a continuous strip of film moving over an aperture at a speed which is synchronized with the ground speed of the plane. Thus the view is "painted" onto the film. Focused onto the film are two views (one on each half of the film), one slightly leading the plane and the other slightly following it. Each vehicle is thereby photographed twice. The offset between the position of the vehicle on half the film and its position on the other half is proportional to the speed of the vehicle. When developed, these two film halves are superimposed through a stero viewer, the offset in vehicle position appearing as a vertical, rather than horizontal, displacement. The vertical displacement can be measured by optical techniques to 0.001 in., and thus the speed of the vehicle can be determined.

Another aerial photographic technique uses time-lapse photography to determine traffic speeds, flows, densities, etc. An increasing number of applications of this technique is being made, and the principle is being extended for use with high-altitude photography (10,000 ft) capable of covering wide areas with a minimum of distortion.

### 5.2.2 Car Following

Instrumentation developed by Chandler, Herman and Montroll (*6*) provides the most detailed and precise measure of car following now obtainable. As discussed from the theoretical standpoint in a previous chapter, and from the experimental standpoint in Section 5.4.2, data are obtained by connecting two cars with a wire. The second car is equipped with a reel of fine wire. The end of the wire is fastened to the rear bumper of the lead car, and constant tension is maintained by a slipping friction clutch. The distance at any instant between the two vehicles is measured by the position of the reel through a potentiometer geared to the reel shaft. The relative velocity of the two vehicles is measured through a DC generator tachometer operating off the reel shaft. This generates a signal proportional to the rate at which the reel is supplying or taking up wire. Absolute velocity and acceleration of the following car are measured directly in the following car. These parameters, together with a time pulse and an event pulse, are recorded on a six-channel oscillograph carried in the following car. Analysis of the oscillograph charts requires measuring the displacement of each of the six pens and is time consuming. Although it would be possible to automatically convert these data to digital form for direct analysis on an electronic computer, the equipment to do this is costly.

An excellent application of commercially available equipment for traffic study has been made by Jones and Potts (*26*) using tachographs. These devices are installed in vehicles and, when driven by speedometer cables, inscribe vehicle speed, distance and engine operation on a wax-coated disc. The disc, marked for time, revolves at a uniform rate. In the Jones and Potts application, the speed trace was used to determine the frequency of changes in acceleration. This parameter is a useful index of the driving quality of a road and of car-following behavior on various roads.

Although not intended primarily for car-following studies, a device developed recently by Greenshields (*21*) provides extensive data on driving behavior by recording the driver's actions in steering, accelerating, and braking. When records are made of the input stimuli to the driver by means of photography or other measurements, studies of driver responses to road and traffic events can be made.

### 5.2.3 Transducers

The ideal detector will be accurate, inexpensive, reliable, easy to install, offer a long and trouble-free life, not drift or require other adjustment, and have no effect on traffic. It does not yet exist; but there is now a great variety of detectors, each of which satisfies some of the mentioned requirements.

Before briefly reviewing the various detectors now available, it might be helpful to consider manual data collection. For short-term surveys, the ease of using manpower to collect data is a great advantage. However, the accuracy of the data, the duration of time over which a person can operate continuously, and the influence a person might have on the traffic being measured should all be considered. Where manpower is required to operate recorders, the fastest response time requires more than 0.1 sec (*45*).

The detectors most frequently used in traffic experimentation fall into three classes—axle detectors, vehicle detectors, and speed detectors (*25*). Among axle detectors, the pneumatic tube has received widespread use and is one of the least expensive transducers. Its main drawbacks are low reliability, high maintenance, and influence on traffic behavior. A surface-mounted strip treadle that closes an electrical circuit under the pressure of passing wheels is another inexpensive detector. More rugged detectors operating on this same principle, but enclosed in a frame and embedded in the roadway, are used in toll lanes and are highly accurate. A third unit now being developed, which should have a long life, uses strain gages to measure a minute deflection in a metal plate flush on the roadway surface as wheels pass over.

Particularly where commercial traffic is present, it is difficult to convert axle counts to vehicle counts. As a result, vehicle detectors are being used increasingly. Magnetic detectors are relatively simple in operation, but are not as accurate as the other detectors discussed here. Probably the most accurate and least expensive vehicle detectors are photoelectric cells. However, these are often difficult to install and are not feasible in many environments. A detector which senses a vehicle when a direct beam of ultrasonic energy is broken may be less susceptible to interference by dirt buildup than photocells, but the need to locate transmitter and receiver on opposite sides of the vehicle poses similar installation problems. This resonant ultrasonic detector should not be confused with another detector which also operates on ultrasonic energy. It measures the time required for a pulse of sound from a transmitter to be reflected back to a receiver located in the same housing. The latter detector is simpler to install and is highly accurate, although more expensive. Several other vehicle detectors are available in the same price range as the pulsed ultrasonic unit. Radar detectors are in widespread use, and present models are quite accurate. A highly accurate vehicle detector, which can be installed relatively easily in pavements, detects vehicles through current flow induced in a loop of wire. The induction loops are also proving to be relatively free of the need for maintenance and adjustment.

Speed detectors are not available in such variety. Radar speed detectors are accurate and reliable, although they require associated logic circuitry. These units measure speed by sensing a Doppler shift in the radar frequency. Ultrasonic units are also available to measure speeds by use of the Doppler principle. A simplified and less accurate approximation of speed has been derived. This method uses ultrasonic detectors which convert the duration of time a vehicle is under the detector into a speed reading by assuming the vehicle is of a particular length. Although vehicle lengths vary considerably, ultrasonic units are available to distinguish autos from trucks and buses by differences in vehicle height. The use of one assumed length for low vehicles and a longer assumed length for high vehicles brings the accuracy of this speed measure to an operationally useful level.

### 5.2.4 Recorders

The multi-pen recorder used extensively in traffic studies offers a low-cost method of recording many events simultaneously. For experiments in which it is difficult to work with ink, traces can be made on wax-coated or electro-sensitive paper. Reduction of event data generally involves scaling the distance between successive displacements on the trace from each pen and/or relating the displacements with base time through reference to chart speed. Sweep pen recorders are also available with several channels. These record an analog of an input signal, tracing the magnitude of flow, speed, density or some other parameter over a period of time.

To provide a direct numerical record for ease in data reduction, a printing time clock has been used on some applications (*12*). This device prints time to 0.1 sec, plus a six-letter code, on standard adding machine tape. Hours, minutes and seconds are printed in figures; tenths of a second are printed on a vernier scale. The machine is actuated by switches which key the printing mechanism and select the assigned code for the particular event. This device does not incorporate any means of temporarily storing simultaneous pulses, thus is limited in the number of points which can be recorded at any one time. Even though the direct digital recorder is easier to work with than the multi-pen graph, there is still a large amount of labor required to analyze the record from the time clock device.

An extensive system for recording traffic data has been assembled by the Bureau of Public Roads (*43*) to analyze vehicle speeds and lateral positions in traffic lanes. Inputs are from two pneumatic or strip treadles laid across the traffic lanes to measure vehicle speeds, and from a third treadle which is segmented to measure vehicle placement. These events are associated with a 100-cycle-per-second time base, and through equipment mounted in a specially designed truck are punched on a five-channel paper tape for subsequent conversion to punched cards and analysis by electronic computer.

A further breakthrough has been made recently by General Motors with a device which functions similarly to the time clock recorder previously discussed. It provides a punched paper tape output with times measured to the nearest 0.02 sec for automatic conversion to IBM cards (*11*). Driving this device with a trap of photocells or other vehicle presence-sensitive transducers, it is possible to determine the time headways, speeds, space headways, relative velocities, accelerations and lengths of successive vehicles passing a point. The device is now being augmented with magnetic tape recorders to permit gathering these data simultaneously at many points on a traffic lane and in adjacent lanes. A simplified version of this clock-tape punch unit has recently been purchased from a data systems manufacturer by The Port of New York Authority and is available for experimentation.

In addition to these graphical and digital recorders, many other recorders have been developed in recent years for all types of experimentation. As with transducers, the traffic experimenter can select among a broad range of instruments to find the one best suited for this purpose.

### 5.2.5 Computers

Increasing attention is being given to devices for traffic experimentation which accept a number of inputs and generate particular output signals depending on the pattern and timing of the inputs. A complete range of analog computers is available to indicate directly the volume, speed and time-based density of vehicles flowing past a point (*41*). Another manufacturer markets equipment to compute volume and lane occupancy ratio. This equipment is installed for experiments on the John Lodge Expressway in Detroit and on the Congress Street Expressway in Chicago. By immediately assigning values to the principal characteristics of a traffic stream, these devices enable direct experimentation without intermediate data reduction and analysis.

More detailed studies can be performed on more general purpose digital computers using punched paper tape inputs from field recorders or IBM cards punched from manually prepared records. In addition to being a powerful analytical tool permitting the rapid handling of vast quantities of traffic data in complex analyses, the general purpose digital computer also offers great promise as a traffic control component (*2*, *5*). The potential of general purpose computers for contributing to the understanding and more effective control of road traffic is largely untapped at the present time.

## 5.3 DRIVER EXPERIMENTS

It is important to measure driver behavior in response to traffic events when developing a theoretical understanding of road traffic. The driver is one of the key elements in road traffic, together with vehicle, traffic stream, roadway, and roadway environment characteristics. Although considerable attention has been given, particularly in recent years, to measuring driver behavior from the standpoints of safety and accident prevention, relatively little experimentation has been performed for the purpose of understanding how driver behavior determines and affects road traffic.

### 5.3.1 Forbes' Measurements of Driver Reaction Time

One of the most significant experimental studies in the field was conducted by Forbes *et al.* in 1956 and 1957 (*16*). Objectives of their research were to determine effects of lighting and other operating conditions on driver reactions, to measure driver perception response relationships which determine acceleration and deceleration rates of successive vehicles, and to examine these factors in relation to other factors affecting driver behavior under various operating conditions.

Their experiment measured the reaction of a driver under various environmental conditions to a set of standard changes in the behavior of cars ahead of him. A platoon of three cars was formed with the driver under observation located in the third car. Control of the experiment was exercised by a passenger in the third car who could signal to the driver of the first car in a manner undetected by the driver of the third car. On signal, the driver of the first car would accelerate or decelerate, and drivers of the second and third cars would respond as though they were commuters on the way home. This experiment was conducted for many third-car drivers and in many different environments.

The speed and driver actions on the accelerator and brake pedals of the third car, together with the field of view observed by the third-car driver, were recorded on film by a specially-equipped camera mounted in the third car (see Section 5.2.1). One of the most interesting findings of the research was that any marked deceleration and acceleration in the platoon increased time headways between the vehicles. This suggested that traffic flows would be highest when speed changes are at a minimum, an indication which has subsequently been confirmed (*45*).

The experiment also showed that there is a difference in the reaction time of drivers, depending on the environment in which they must react. On a roadway curving to the right, drivers took longer to react than on a roadway curving to the left. Similarly, downgrades, low illumination, and even subjective constrictions, tended to lengthen reaction times.

The third important finding from these experiments was that the time required for the driver of the third car to react to a deceleration was shorter than that for an acceleration. The theoretical implications of this finding have yet to be fully evaluated.

### 5.3.2 Driving Simulation

It is likely that the Forbes findings concerning driver reaction times can be understood better when it is possible to quantify the information presented to the driver in each of the experimental situations. Presumably, the better informed a driver is, at least up to some point, the better he will be able to anticipate the proper reaction. This area of research should produce much important information for the designers and operators of roadways, and can probably best be undertaken in simulated controlled environments.

Work has been under way in the last few years at the Institute of Transportation and Traffic Engineering in California (*24*) on the construction of a driving simulator that uses motion pictures to provide visual inputs. Progressive steps are being taken to increase the realism of the projected films, including shooting both front and rear views and projecting them on wide-angle curved screens. The vehicle which the driver operates is mounted on dynamometers and is being equipped to roll and pitch in response to accelerations and projected "road" characteristics.

Another approach has been outlined by the U. S. Public Health Service, using scale-model terrains over which television cameras would be "driven" by an experimental subject viewing the television monitor. This technique would enable complete control of the environment presented to the experimental drivers, but it apparently is not as close to operation as the filmed input simulator at ITTE.

Another approach has been developed at Ohio State University (*44*) to investigate the acceleration patterns followed by drivers responding to changes in relative velocity with a car ahead. These researchers used a television camera viewing a model car ahead and, beyond the model, a representation of landscape passing at a controlled velocity.

Distance between the model car and the television camera is controlled by moving the model car in response to acceleration changes initiated by the experimental driver.

An even simpler simulation has been used recently by Michaels and Stephens (*32*) to study the driver's ability to perform simultaneously a guidance control function while recognizing other information inputs. Such rudimentary simulators are of definite use in studying limited aspects of driver behavior.

Simulators are an especially powerful tool for developing traffic theory because they enable control and precise measurement of nearly all variables in the driving process. The progress being made at ITTE is promising, and further attention in this area will assist greatly in understanding road traffic.

### 5.3.3 Galvanic Skin Reflex Studies

Significant research on human factors is being conducted by the Bureau of Public Roads. Michaels (*31*) is measuring galvanic skin reflexes of drivers moving in the traffic stream through various environments. His experiments show that driver tension decreases as predictability of interferences increases. A corollary is that driver tension increases as the driving situation becomes more complex.

Measurements of galvanic skin reflex have also been made by Cleveland (*7*), who found that illumination of a "Y" intersection reduced the tension in drivers traversing the intersection.

## 5.4 CAR FOLLOWING

The theory of car following describes the manner in which one vehicle follows another. This has been used to describe the discontinuity which exists in a stream of traffic when it suddenly becomes congested. Data derived from many car-following experiments have substantiated the theory, which is essentially a measure of driver behavior.

### 5.4.1 Forbes' Tunnel Experiments

As part of his research to measure driver reactions as a function of environment, Forbes (*16*) conducted car-following experiments in the Holland, Lincoln and the Queens Midtown Tunnels. These experiments consisted of having the camera-equipped car drive in the regular traffic stream. The driver of the experimental car was instructed to drive in a manner duplicating as much as possible the behavior of the driver in the car in front of him. Thus, when the driver of the preceding car would accelerate, so would the driver in the experimental car.

Inasmuch as the lead car in these experiments was not equipped with reference lights enabling subsequent parallax measurements to determine the difference between lead and experimental cars, these measurements have considerably more error than the measurements taken within the three-car platoons mentioned in Section 5.3.1. Nevertheless, they provide an interesting pattern of driver behavior which is now being analyzed.

### 5.4.2 General Motors Car-Following Experiments

Extensive experimentation has been conducted by Chandler, Herman, Montroll, Potts, Gazis and Rothery (*6, 23, 17*). Their work is of major theoretical importance and is included in Chapter 2. It will be mentioned here only briefly, but because the work does include experimentation, it is cited as a matter of experimental as well as theoretical interest.

Their experiment consisted of measuring the distance and relative velocity between a leading car and a following car by means of a wire stretched between the two cars (see Section 5.2.2). The take-up reel for the wire was mounted on the front bumper of the following car, and motions of the reel were converted to electrical signals by potentiometer and tachometer. These signals were recorded on a six-channel oscillograph, which was also recording the speed and acceleration of the following car plus the time and special reference points.

Analysis of the oscillograph records presents a difficult problem in data reduction. Only part of the data collected in Detroit and in vehicular tunnels in the New York area has been analyzed. However, even the limited data available shed light on the processes by which one car follows another.

More recently Jones and Potts (*26*) have used tachographs to measure the dispersion of acceleration noise in car-following studies. These records, although much less precise, are considerably easier to analyze. Section 2.3, on car following, presents the theoretical results of this work which have already been published.

## 5.5 PLATOON STUDIES

Many studies have been made of traffic behavior within platoons to evaluate this effect on traffic flow. Although this work has not been conclusive, it has shown that platoon behavior is of major consequence in the application of traffic flow theory.

### 5.5.1 Forbes' Platoon Studies in Pasadena

Forbes (*15*) reported in 1951 that traffic behavior within platoons was not adequately described by the behavior of the over-all traffic stream. He based his work on speed, headway and volume studies of traffic on the Pasadena Freeway. Although previous studies of mass traffic behavior reported decreases in speed as traffic volumes increased (and average headway times decreased), Forbes observed that there was little relation between speeds and headways (either time or distance) of vehicles within platoons. These observations identified platoon behavior as an important element in traffic theory, although Forbes did not consider the theoretical consequences of the data in that report.

Defining a platoon member was a major difficulty encountered by Forbes and others who subsequently studied platoon behavior. Forbes reviewed his data in 5-min classes and considered a platoon member as one whose time headway was less than the mean for that class. At least three successive vehicles with below-average time headways were necessary to constitute a platoon.

### 5.5.2 Port Authority Platoon Experiments

The problem of defining who is a platoon leader and who is a member of a platoon was avoided in experiments conducted by The Port of New York Authority (*10*) in the Holland Tunnel South Tube in 1959. In this experiment, the platoon leader was an experimental car traveling at a fixed speed. These experiments to evaluate platoon behavior and measure road capacity were conducted during off-peak periods when normal speed of the traffic stream was higher than optimum (in off-peak periods the speeds in the Holland Tunnel are approximately 35 mph). Optimum speeds for this tube lie between 20 and 25 mph—that is, at those speeds traffic flows are highest. The hypothesis was that if a platoon of drivers were assembled at random and required to drive behind a platoon leader traveling at a speed close to optimum, the length of the platoon at any point would indicate roadway capacity. In traversing roads of high capacity at this optimum speed, drivers would drive closer together on the average than in traversing a road of lower capacity.

Following the platoon leader, nine cars taken at random entered the tunnel. Ten headways after the first experimental car, a second experimental car was introduced. The time for the two experimental cars to pass any given point in the tunnel (reconstructed from records maintained in the experimental cars) was, therefore, 10 times the average time headway being maintained by the members of the platoon at that point. It was found that the capacity profile of the roadway determined by repeated experiments of this type was similar in pattern to the profile measured in other experiments (see Section 5.6.3). However, the capacity values for this platoon analysis were much higher than those obtained in other measurements (1,550 versus 1,250 veh/hr).

In other analyses of Holland Tunnel data Greenberg and Daou (*20*) observed the tendency for flow to be higher following a gap. This observation was consistent with the results of the optimum speed platoon experiments and suggested that over-all flows might be improved by the periodic introduction of gaps in the traffic stream (see Section 5.7.1).

Characteristics of platoons in the traffic stream passing through the bottleneck in the Holland Tunnel South Tube were examined through extensive measurements made in 1960 by using the General Motors data acquisition system discussed in Section

5.2.4. This system collects data in a form suitable for direct processing by electronic computer, enabling use of quite complex definitions to identify platoons. The definitions used in the published report (*11*) considered both the extent of space headway available to a vehicle (the particular headway varied with speed) and whether a change in the type of flow had occurred. The latter was manifested by excessive relative velocity, or by a change in acceleration sign (positive to negative or vice versa) for a group of vehicles.

Behavior of the various types of platoons was examined, and it was clear that the highest flows were attained with the highest speed platoon leaders. There did not appear to be a consistent relationship between speed reduction and platoon length or platoon type. However, further analysis of platoon behavior is needed, and theoretical work is under way.

### 5.5.3 Platoon Flow Through Intersection Traffic Signal Systems

Apart from the formation of traffic platoons in continuous flow, the action of signalized intersections in forming, releasing and passing platoons of traffic makes platoon behavior of major consequence in traffic theory. Experimental studies of platoon behavior through signalized intersections have been made by Lewis (*28*), Pacey (*36*), Gerlough (*19*), and Newell (*34*).

As an example, measurements of platoon dispersal as the vehicles move downstream from a traffic signal were reported by Lewis (*28*) to compare vehicle-actuated traffic signals with coordinated signals. In the Lewis study the signals in question were 0.81 mi apart on a road with virtually no traffic interruptions between the signals. The analysis showed that coordination of signals at a distance up to 0.65 mi apart would yield better results, because of the platoon behavior, than would be yielded by random operation of the downstream signal.

Applications of platoon studies made by Von Stein (*46*) and Morrison (*33*) are discussed in Section 5.8.

## 5.6 CONTINUOUS-STREAM MODELS

For many years researchers have compared the characteristics of traffic flow to those of a stream of water. Studies based on this comparison have been both statistical and theoretical, describing the wavelike action by which changes in the flow of a few vehicles are transmitted throughout the traffic stream.

### 5.6.1 Early Work by Greenshields

Greenshields (*22*) reported his observations of vehicle traffic flow in 1934, and suggested a linear relationship between speed and concentration. Thus, on flow-concentration coordinates as concentration increased from zero to jam, flow would rise to a maximum and then fall back to zero on a parabolic path. The form of this relationship is particularly important because it permits inferring the maximum flow obtainable over a short section of roadway from the speed and space headway relationships maintained by drivers at lesser flows.

### 5.6.2 Speed Headway Measures by Olcott

The capacity of a roadway section as determined from speeds and headways of traffic passing through that section at less than capacity flows is of basic importance in locating bottlenecks on uninterrupted roadways and determining the margin of capacity available at non-bottleneck sections. This problem is particularly noticeable in the case of vehicular tunnels which, except for sections of vertical and horizontal curves, are uniform throughout their length. Despite this uniformity, congestion occurs in the traffic stream at characteristic points in vehicular tunnels.

To determine whether the speed concentration relationship proposed by Greenshields would be applicable to tunnel traffic flow, Olcott (*35*) measured the speeds and headways of vehicles in various points in Queens, Midtown and Lincoln Tunnels. Using 5-min time slices of traffic to compute mean space speeds and mean densities, Olcott ran a linear regression analysis and found that 88 to 97 percent of the variability in traffic speeds was related to change in the density of the stream. The estimates of capacity derived from his analysis were somewhat higher than the maximum flows generally obtained in these tunnels, indicating that the linear speed-density relationship was not completely applicable.

### 5.6.3 Experiments in New York Tunnels Using Fluid Flow Models

The most significant developments in theoretical descriptions of road traffic as a continuous stream were made in 1955 and 1956 by Lighthill and Whitham (*29*) and by Richards (*40*). Their work on hydrodynamic models of traffic flow is discussed in detail in Chapter 1.

Edie and Foote (*12*) analyzed traffic flows in the Holland and Lincoln Tunnels to evaluate hydrodynamic models as a description of tunnel traffic flow. The hydrodynamic models have provided considerable insight into the behavior of traffic upstream and downstream from bottlenecks. This analogy produces insight particularly into the wave-like action by which changes in flow are transmitted in the traffic stream and is useful in relating flow at the observed point to a restriction at another point.

It is an assumption of the kinematic and similar models that at a given flow less than maximum, traffic would tend to vary around the low concentration or the high one given by the flow-concentration curve and would not fall in between. This assumption permits deductions to be made about flow behavior on each side of the bottleneck which are useful in determining the capacity of non-bottleneck sections. The empirical results showed, however, that traffic flow upstream and downstream does not follow exactly the patterns suggested by the fluid flow analogy. Upstream from the bottleneck, congested flows tended to stretch out over a range of concentrations at a fixed flow level. Downstream from the bottleneck flows tended to assume a small range of concentrations at a fixed speed.

Work by Palmer (*37*) independently evaluated the application of the Lighthill and Whitham kinematic flow theory for describing traffic behavior on the Merritt Parkway. It was found that the actual observed results conformed quite closely with the results obtained when using the Lighthill and Whitham model.

The hydrodynamic models have been most useful in describing the wave-like behavior with which changes in traffic flow, speed and concentration are propagated through the traffic stream. Experiments conducted in the Holland Tunnel (*35*) measured these waves and related them to the action of a bottleneck.

### 5.6.4 Experimental Work by Edie, Foote, Herman and Rothery

The Lighthill-Whitham model does not specify a particular form for the flow concentration relationship over the entire range of possible concentrations. Therefore, it is not directly useful in determining capacities from the speed headway relationships observed in subcapacity flows. Richards' model assumed the linear speed-density relationship proposed by Greenshields. Greenberg (*20*) has suggested a fluid flow model which would specify a continuous relationship over the entire range of densities. Although Greenberg's model appears to fit observed data reasonably well, the observations available at the time his model was formulated were not sufficient to determine whether his model provided more accurate description than previous models. A further refinement was suggested by Edie (*39*).

More recent experiments by Edie, Foote, Herman and Rothery (*11*), using an electronic time clock with punched tape output, have provided flow data on 24,000 vehicles through the Holland Tunnel South Tube. This sample provides an empirical basis for defining relationships among flow and concentration, speed and relative speed, and other relationships for both the total stream and for platoons of various types.

## 5.7 APPLYING TRAFFIC THEORY TO FLOW CONTROL

The most significant practical result stemming from application of traffic theory has been an awareness of the importance of maintaining speed to obtain maximum flow through bottlenecks. This awareness was brought about through the application of traffic theory flow control in locating bottlenecks and measuring their capacity.

### 5.7.1 Manual Traffic Spacing Experiments of Greenberg and Daou

The possibility of improving traffic flows by controlling the number of vehicles on the critical road section at any time has been of interest to traffic engineers for many years (*1, 3*). Attempts prior to 1956 to apply this concept to tunnel traffic flow were not successful. Working with the Holland Tunnel data discussed in Sections 5.5 and 5.6, Greenberg and Daou (*20*) of the Port

Authority observed that small gaps frequently occurred in the traffic stream when movement was at a high rate and fluid. This suggested that the forced introduction of small gaps might bring about the desired result of maintaining fluid flow and preventing congestion.

To test this possibility, experiments were conducted in 1959 whereby traffic flows entering the fast lane of the Holland Tunnel were limited to no more than 22 vehicles each minute. This was accomplished by an observer who counted the vehicles, observed a stop watch, and signaled a police officer to interrupt traffic for the time between the passage of the 22nd vehicle and the completion of one minute after the arrival of the first vehicle. Inasmuch as more than 22 vehicles per lane would frequently desire access to the tunnel in a minute, this control caused periodic introduction of spaces in the traffic stream.

Experiments were conducted on alternate days over a six-week period for one hour in the afternoon. The rate of traffic flow increased 6 percent when the control procedure was used.

This improvement was related to the tendency of each successive vehicle in a platoon to travel slower than the vehicle ahead of it as the platoon passed through a bottleneck. A study by Palmer (*37*) independently reported the same finding.

It has been suggested (*10*) that as more vehicles enter the critical road section, the space between them is necessarily less and hence platoons become longer. With longer platoons and the average decline in speeds as the platoon lengthens, speeds eventually drop so low that the time required by successive vehicles to pass a point starts to lengthen. When this occurs, vehicles approaching this point must delay arriving at that point, and their sudden slowing precipitates a shock wave back through the stream (as discussed in Section 5.6.3).

The frequent gaps inserted by Greenberg and Daou prevented formation of excessively long platoons and enabled speeds to remain slightly higher than critical.

### 5.7.2 Instrumented Traffic Spacing by Foote, Crowley and Gonseth

Because the manual traffic spacing experiments discussed in Section 5.7.1 were limited to a particular hour, with a flow limit of 22 vehicles per minute derived from study of traffic behavior in that lane at that time, a new series of experiments was undertaken by Foote, Crowley and Gonseth (*14*) to extend the application of this procedure.

As the first step, a simple device for spacing traffic automatically was constructed. It consisted of a stepping switch to count vehicles and a one-minute timer. Vehicles were detected entering the tunnel lane. When the pre-set number of vehicles (ranging from 15 to 24) had entered the fast lane of the tunnel in less than a certain time, the device would automatically turn entrance signal lights red, energize flashing "STOP" signs and sound a bell for the remainder of that minute.

Because traffic conditions in the Holland Tunnel vary widely during the week, it was not feasible to think in terms of a predetermined "optimum" setting for the traffic spacer. A means of setting the input control for the particular traffic conditions existing at any one time was considered essential for regular operation. Inasmuch as the immediate purpose of the control system was to prevent traffic speeds at the bottleneck from dropping below critical, it was decided to continuously measure bottleneck speeds and determine the proper setting for the input control based on that information. With slow speeds at the bottleneck, the input control was set to insert frequent spaces in entering traffic.

Experiments were conducted over a six-week period for three hours in the morning and three hours in the afternoon. An experimenter observing speeds inside the tunnel constantly determined the proper settings for the input control.

Several major improvements were gained when this system was in operation. Traffic production through both lanes of the tube was increased 5 percent during the critical hour, particularly in the afternoon. The increase was caused by both a 25 percent reduction in the occurrence of disabled vehicles and by an improvement in the speed-headway relationships maintained by motorists passing through the bottleneck in the optimal speed range. The modest but significant production increases had a marked effect on traffic congestion awaiting entrance to the tunnel. There was a reduction of 33 percent from three hours of congestion to two hours of congestion, when the flow of tunnel traffic was controlled. Also, there was a sharp reduction in the contamination of tunnel air by vehicle exhausts. When traffic moves at constant speeds, far less carbon monoxide is emitted than in stop-and-go driving.

### 5.7.3 Experiments with Completely Automatic Flow Control System

Another purpose of the instrumented traffic spacing experiments was to evaluate equipment for permanent operation of the flow control system. It was found that the test procedure, relying on an observer to evaluate constantly information on traffic speeds and flows from several sources and then to decide the best setting for the input control, was not fully effective. The continuous concentration required on the part of the observer, the extreme rapidity with which traffic conditions could change, and the number of points which the observer should consider in reaching his decision, all made his job highly demanding. For regular operation, it was concluded that a completely automatic system should be developed.

Based on these experiments, the Port Authority has now developed what might be described as the "first generation" prototype automatic system for controlling traffic flow (*13*). The prototype system measures speeds at two points in the fast lane of the Holland Tunnel South Tube by means of two sets of photocells, each set bounding a 13-ft zone. The amount of time required for vehicles to pass through each zone indicates whether they are going faster or slower than a pre-set speed value. The system also measures the number of vehicles passing through the bottleneck at the foot of the upgrade, where one of the sets of photocells is located. During each minute, the system considers the bottleneck flow and determines whether speeds at the bottleneck and approaching the bottleneck are high, medium, or low. Depending on this volume and speed information, the computer then sets the input control for spacing traffic at a certain level for the next minute.

If the speeds are low inside the tunnel, the computer will set the input control at a value lower than the number of vehicles that pass through the bottleneck in the preceding minute. As long as speeds inside the tunnel remain at a low level, fewer vehicles will be entering the tunnel than passing through the bottleneck. Eventually speeds will rise. When this occurs, the system will allow additional vehicles to enter the tunnel until speeds stabilize in the mid-range. Limited operating experience has been gained with this system, and it appears to be functioning as intended. However, analysis of its results shows that further improvements can be made, and it is expected that development will continue through several additional models. This system is also being extended to other tunnels operated by the Port Authority.

### 5.7.4 Traffic Surveillance on the John Lodge Expressway in Detroit

An extensive test of surveillance and control equipment on the John Lodge Expressway in Detroit is being conducted to evaluate equipment and improve traffic flow. Major steps are being taken to improve flow by the early detection of interruptions, use of lane control and changeable speed signals, and control of traffic entering the expressway.

Reports by Gervais (*18*) and others indicate that significant improvements in expressway traffic flow can be achieved.

Research is under way on a 3.2-mi section with traffic volumes of more than 160,000 veh/day. Despite overloading, traffic is being handled reasonably well.

Closed-circuit television is one of the surveillance methods. Fourteen cameras, each attached to its own monitor, are so spaced on bridges that pictures of almost all portions of the freeway can be obtained. Special equipment provides a usable night picture and excellent daylight pictures under varying weather conditions.

Research is intended to lead to development of a traffic control system which would utilize information gained from the TV monitoring network.

A recently-installed control system consists of lane signals and variable speed signs. The lane signals are a red "X," to indicate that a driver must leave his lane as soon as it is safe, and a green arrow, which indicates the lane is open for traffic. The variable speed signs permit speed messages in 5-mi increments from 20 to 60 mph. Information from sensing equipment along the freeway is instantly analyzed to determine appropriate speeds.

The system requires control from a central point with equipment circuitry which changes signals at remote points and then confirms that the signals are operating properly.

An ultrasonic vehicle volume detector also in use is able to detect with reasonable accuracy whether a vehicle is a truck or a passenger car.

Data obtained from this second surveillance system have enabled operators to compute average speed of the traffic stream, average time headway, average distance headway, and average distance spacing.

This project, using television monitoring in conjunction with traffic sensing devices, has produced valuable information for both traffic control and research purposes.

### 5.7.5 Congress Street Expressway Studies in Chicago

A series of vehicle detectors has been installed on the outbound lanes of the Congress Street Expressway in Chicago (*30*).

A 5-mi test section includes several on- and off-ramps, a transition of four to three lanes, and some apparent bottlenecks. Traffic volume counts and lane occupancy are measured at each detector location. The surveillance project began with a complete inventory of flow characteristics throughout the test section.

Plans are being made to experiment with control measures as possible methods for limiting congestion and maintaining traffic flows at maximum capacity level.

The project is designed to conduct operational studies, locate critical points, determine causes of congestion both qualitatively and quantitatively, investigate ways to improve flow, and measure the resulting benefits to traffic. The project staff is also investigating electronic techniques for surveillance of traffic behavior and the detection of stalled vehicles.

The system consists of traffic detectors on the ramps and on the expressway at selected locations along the study section. Also in use are analog computers, an interconnection network, map display, and various data recording devices, including a paper punched tape output.

Some of the initial traffic studies being made from the punched tape output are as follows:

(a) Interrelationships between volume, occupancy, and speed.
(b) Comparison of measured speed and speed calculated from volume and occupancy.
(c) Comparison of measured occupancy and density calculated from volume and speed.
(d) Comparison of lane traffic characteristics.
(e) Comparison of traffic characteristics between mainline stations.
(f) Changes in traffic characteristics just prior to congestion.
(g) Combination of shoulder lane volume and ramp volume resulting in maximum flow and satisfactory operation.
(h) Measurement of the effect of congestion on traffic flow and travel time.

The detection system is providing a comprehensive library of measurements which permit microscopic and macroscopic investigations, both qualitative and quantitative. The data logging subsystem is recording the measurements in a manner permitting full utilization of data processing equipment with a minimum of time and without loss of accuracy.

This extensive experimentation, like that at Detroit and New York, is providing a major testing ground for the application of traffic theories.

## 5.8 APPLYING TRAFFIC THEORY TO INTERSECTION CONTROL

The limited applications of road traffic theory have in general aimed at the same goal: to increase the proportion of time that traffic passing over a critical roadway is moving smoothly at mid-range speeds. The tunnel traffic flow control in New York accomplishes this aim by spacing traffic entering the tunnel so that vehicles will naturally move at mid-range speeds, at which flow is maximum.

Other experiments have been conducted to determine if traffic flow through a series of intersections can be increased by controlling the spacing of vehicles entering the critical area.

### 5.8.1 Dusseldorf's Signal Funnel

An application of this principle is reported from Dusseldorf (*46*) where, since 1954, traffic approaching certain intersections has been spaced. The "spacing" is accomplished by speed signals advising motorists to travel at 20, 25 or 30 mph in order to reach an intersection at a time when the intersection signal will be green for their direction. This "signal funnel" as it is called by its developer, significantly increases the proportion of vehicles able to pass through an intersection without stopping. It is reported that this has the effect of increasing capacity by approximately 20 percent.

The signal funnel requires spacing the speed signals in advance of an intersection by an amount which varies according to the permissible difference in minimum and maximum speeds and the duration of green time. In many United States urban areas, it is doubtful that sufficient length of roadway is available in advance of critical intersections to effectively assemble vehicles in a moving platoon. This system also would not be as effective when traffic densities are so high that motorists are not able to drive at the higher speeds. However, the Dusseldorf experience does appear to merit serious consideration by United States traffic authorities for controlling boulevard traffic where grade intersections are spaced at intervals of 1,000 ft or more.

### 5.8.2 Experiments with Pacer System in Detroit

A major application of the signal funnel is being tested on Mound Road in Warren, Mich., by the General Motors Research Laboratories in collaboration with county and state officials. This test has involved development of variable speed signs and installation of pre-signals. The experiment consists of measuring such parameters as transit time through the test section, number of vehicles stopped at intersections, and frequency of stops through the test section under three types of traffic signal operation—non-interconnected, progressive, and pacer. Results reported to date by Morrison (*33*) indicate reductions in the number of stops required by motorists traversing the test section. The experiment is being continued, and a final evaluation has yet to be made.

### 5.8.3 Traffic Signal Control Experiments in Toronto

The first application of a general purpose computer to the control of a network of urban traffic signals has been reported from Toronto (*19*). Although the main findings of these experiments reported to date do not relate directly to road traffic theory, the use of a general purpose computer made it possible to study the detection of impending traffic congestion. It was also possible to consider new traffic signal operating strategies aimed at increasing the time during which fluid traffic conditions are maintained. Inasmuch as this work has not been reported in detail, the theoretical aspects of the work have yet to be evaluated.

However, the Toronto use of the general purpose computer has illustrated the challenges and opportunities facing traffic engineers and roadway operators. The ability of general purpose computers to handle large amounts of data at unbelievably rapid speeds indicates that the essential instrumentation is now available to apply more sensitive and detailed road traffic theories for improved traffic operations. The computer also is a powerful tool in evaluating traffic control experiments and, together with the interests of physicists and mathematicians in the theoretical aspects of road traffic flow, suggests that major improvements in traffic operations can be obtained.

## REFERENCES

1. Barnett, J., "Operation of Urban Expressways." *Proc.*, Amer. Soc. Civil Eng., 83: No. HW4, Paper 1374, 1-7 (Sept. 1957).
2. Bauer, W. F., Gerlough, D. L., and Granholm, J. W., "Advanced Computer Applications." *Proc.*, Inst. Radio Eng., 49: 296-304 (Jan. 1961).
3. *Highway Capacity Manual*, Bur. of Public Roads. U. S. Government Printing Office, Washington, D. C. (1950).
4. Capelle, D. G., and Pinnell, C., "Capacity Study of Signalized Diamond Interchanges." HRB Bull. 291, pp. 1-25 (1961).
5. Cass, S., and Casciato, L., "Centralized Traffic Signal Control by a General Purpose Computer." Proc. of 30th

Ann. Meeting, Inst. of Traffic Eng., pp. 203-211 (1960).

6. Chandler, R. E., Herman, R., and Montroll, E. W., "Traffic Dynamics: Studies in Car Following." *Oper. Res.*, 6: 2, 165-184 (Mar.-Apr. 1958).
7. Cleveland, D. E., "Driver Tension and Rural Intersection Illumination." Proc. 31st Ann. Meeting, Inst. of Traffic Eng., pp. 10-22 (1961).
8. Edie, L. C., "Car-Following and Steady-State Theory for Non-Congested Traffic." *Oper. Res.*, 9: No. 1 (Jan.-Feb. 1961).
9. Edie, L. C., and Foote, R. S., "Experiments on Single-Lane Flow in Tunnels." *Theory of Traffic Flow*, pp. 175-192, Elsevier Publ. Co. (1961).
10. Edie, L. C., and Foote, R. S., "Effect of Shock Waves on Tunnel Traffic Flow." *HRB Proc.*, 39: 492-505 (1960).
11. Edie, L. C., Foote, R. S., Herman, R., and Rothery, R., "Analysis of Single Lane Traffic Flow." *Traffic Eng.*, 33: 21-27 (Jan. 1963).
12. Edie, L. C., and Foote, R. S., "Traffic Flow in Tunnels." *HRB Proc.*, 37: 334-344 (1958).
13. Foote, R. S., "Development of an Automatic Traffic Flow Monitor and Control System." HRB Bull. 291, 79-103 (1961).
14. Foote, R. S., Crowley, K., and Gonseth, A. T., "Instrumentation for Improved Traffic Flow." Proc. 30th Ann. Meeting, Inst. of Traffic Eng., pp. 273-295 (1960).
15. Forbes, T. W., "Speed, Headway and Volume Relationships on a Freeway." Proc. 22nd Ann. Meeting, Inst. of Traffic Eng., pp. 103-126 (1951).
16. Forbes, T. W., Zagoraki, H. J., Holshouser, E. L. S., and Deterline, W. A., "Measurement of Driver Reactions to Tunnel Conditions." *HRB Proc.*, 37: 345-357 (1958).
17. Gazis, D. C., Herman, R., and Potts, R. B., "Car-Following Theory of Steady-State Traffic Flow." *Oper. Res.*, 7: No. 4 (July-Aug. 1959).
18. Gervais, E., DeRose, F., Dudek, C., Richard, C. and Roth, W., "The Development and Evaluation of John C. Lodge Freeway Surveillance and Control Research Project." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
19. Gerlough, D. L., "Some Problems in Intersection Traffic Control." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 10-27 (1961).
20. Greenberg, H., and Daou, A., "The Control of Traffic Flow to Increase the Flow." *Oper. Res.*, 8: No. 4 (July-Aug. 1960).
21. Greenshields, B. D., "Driving Behavior and Related Problems." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
22. Greenshields, B. D., "A Study of Traffic Capacity." *HRB Proc.*, 14: 448-481 (1934).
23. Herman, R., Montroll, E., Potts, R. B., Rothery, R. W., "Traffic Dynamics: Analysis of Stability in Car Following." *Oper. Res.*, 7: 1, 86-106 (Jan.-Feb. 1959).
24. Hulbert, S. F., and Wojcik, C., "Simulating Inertial Forces Upon the Driver With and Without Visual Clues." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
25. Inst. of Traffic Eng., Tech. Comm. 7-G, "Volume Survey Devices." *Traffic Eng.*, 31: 44-51 (Mar. 1961).
26. Jones, T. R., and Potts, R. B., "The Measurement of Acceleration Noise—A Traffic Parameter." *Oper. Res.*, 10: 6, 745-763 (Nov.-Dec. 1962).
27. Jordan, T. D., "Project Sky Count." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
28. Lewis, B. J., "Platoon Movement of Traffic from an Isolated Signalized Intersection." HRB Bull. 178, pp. 1-11 (1959).
29. Lighthill, M. S., and Whitham, G. B., "On Kinematic Waves: A Theory of Traffic Flow on Long Crowded Roads." *Proc.*, Royal Soc., Series A, 229: 1178, 317-345 (May 1955).
30. May, A., Athol, P., and Parker, W., "Development and Evaluation of Congress Street Expressway Pilot Detection System." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
31. Michaels, R. M., "Tension Responses of

Drivers Generated on Urban Streets." HRB Bull. 271, pp. 29-44 (1960).

32. Michaels, R. M., and Stephens, B. W., "Part-Task Simulation in Driving Research." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
33. Morrison, H. M., Underwood, A. F., and Bierley, R. L., "Traffic Pacer." HRB Bull. 338, pp. 40-68 (1962).
34. Newell, G. F., "The Flow of Highway Traffic Through a Sequence of Synchronized Traffic Signals." *Oper. Res.*, 8: 3, 390-405 (May-June 1960).
35. Olcott, E. S., "The Influence of Vehicular Speed and Spacing on Tunnel Capacity." *Oper. Res.*, 3: 2, 147-167 (1955).
36. Pacey, G. M., "The Progress of Vehicles Released from a Traffic Signal." Res. Note No. RN 126651 GMP, British Road Res. Lab. (Jan. 1956).
37. Palmer, R. M., "The Development of Traffic Congestion." Thesis submitted to Yale Bur. of Highway Traffic (May 1959).
38. Port of New York Authority, "Traffic Data Report, Western Approaches to the Lincoln Tunnel." Continuous-strip aerial photography by Chicago Aerial Survey (1960).
39. Port of New York Authority, Unpublished data on wave action in traffic streams on George Washington Bridge obtained by use of 35-mm camera. Tunnels and Bridges Dept. (1962).
40. Richards, P. I., "Shock Waves on the Highways." *Oper. Res.*, 4: 1, 42-51 (Feb. 1956).
41. Ricker, E. R., "Monitoring Traffic Speed and Volume." *Traffic Quart.*, 13: 1, 128-140 (Jan. 1959).
42. Sickle, S. M., "Continuous-Strip Photography—An Approach to Traffic Studies." *Traffic Eng.*, 29: 10, 11-12, 59 (July 1959).
43. Taragin, A., and Hopkins, R. C., "A Traffic Analyzer: Its Development and Application." Proc. 30th Ann. Meeting, Inst. of Traffic Eng., pp. 263-273 (1960).
44. Todosiev, E. P., "Applications of the Automobile Simulator." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
45. Tufts College of Applied Experimental Psychology, *Handbook of Human Engineering Data* (Nov. 1952).
46. Von Stein, W., "Traffic Flow with Pre-Signals and the Signal Funnel." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 28-56 (1961).
47. Wagner, F., and May, A., "The Use of Aerial Photography for Traffic Operation Studies." Paper presented at 42nd Ann. Meeting, Highway Research Board (Jan. 1963).
48. Wohl, M., "Vehicle Speeds and Volumes Using Sonne Stereo Continuous-Strip Photography." *Traffic Eng.*, 29: 14-19, 49 (Jan. 1959).

# APPENDICES

## Contents

# Appendix A

# BIBLIOGRAPHY ON THEORY OF TRAFFIC FLOW AND RELATED SUBJECTS

There has been an increasing interest within the past few years in the analysis of traffic flow by various mathematical models. Theoreticians in several fields have found that traffic flow problems constitute challenging areas for research.

The following bibliography was prepared by the Highway Research Board Committee on Theory of Traffic Flow as an aid in the search for knowledge in this emerging area of highway and traffic engineering.

## HYDRODYNAMIC ANALOGIES AND KINEMATIC WAVES


1. Bick, J. H., and Newell, G. F., "A Continuum Model for Two-Directional Traffic Flow." *Quart. Appl. Math.*, 18: 2, 191-204 (July 1960).
2. De, S. C., "Kinematic Wave Theory of Bottlenecks of Varying Capacity." *Proc.*, Cambridge Phil. Soc., 52: Pt. 3, 564-572 (July 1956).
3. Greenberg, H., "An Analysis of Traffic Flow." *Oper. Res.*, 7: 1, 79-85 (Jan.-Feb. 1959).
4. Lighthill, M. J., "Dynamic Instability of Transport Systems: The Hydrodynamic Analogy." *Oper. Res. Quart.*, 8: 3, 109-114 (Sept. 1957).
5. Lighthill, M. J., and Whitham, G. B., "On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads." *Proc.*, Royal Soc., London, Series A, 229: 1178, 317-345 (May 10, 1955).
6. Richards, P. I., "Shock Waves on the Highway." *Oper. Res.*, 4: 1, 42-51 (Feb. 1956).


## TRAFFIC DYNAMICS (CAR FOLLOWING)


1. Chandler, R. E., Herman, R., and Montroll, E. W., "Traffic Dynamics: Studies in Car-Following." *Oper. Res.*, 6, 2, 165-184 (Mar.-Apr. 1958).
2. Chow, T. S., "Operational Analysis of a Traffic-Dynamics Problem." *Oper. Res.*, 6: 6, 827-834 (Nov.-Dec. 1958).
3. Edie, L. C., "Car-Following and Steady-State Theory of Noncongested Traffic." *Oper. Res.*, 9: 1, 66-67 (Jan.-Feb. 1961).
4. Franklin, R. E., "The Propagation of Disturbances in a Single Lane of Traffic." *Civil Eng. Pub. Works Rev.*, London, 57: 668, 344-346 (Mar. 1962); No. 669, 487-489 (Apr. 1962).
5. Gazis, D. C., Herman, R., and Potts, R. B., "Car-Following Theory of Steady-State Traffic Flow." *Oper. Res.*, 7: 4, 499-505 (July-Aug. 1959).
6. Gazis, D., Herman, R., and Maradudin, A., "The Problem of the Amber Signal Light in Traffic Flow." *Oper. Res.*, 8: 1, 112-132 (Jan.-Feb. 1960); Reprinted in *Traffic Eng.*, 30: 10, 19-26, 53 (July 1960).
7. Gazis, D. C., Herman, R., and Rothery, R. W., "Nonlinear Follow-the-Leader Models of Traffic Flow." *Oper. Res.*, 9: 4, 546-567 (July-Aug. 1961).

8. Gazis, D. C., and Weiss, G. H., "Density Oscillation Between Lanes of a Multilane Highway." *Oper. Res.*, 10: 5, 658-667 (Sept.-Oct. 1962).
9. Herman, R., Montroll, E. W., Potts, R. B., and Rothery, R. W., "Traffic Dynamics: Analysis of Stability in Car Following." *Oper. Res.*, 7: 1, 86-106 (Jan.-Feb. 1959).
10. Herman, R., and Potts, R. B., "Single-Lane Traffic Theory and Experiment." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 120-146 (1961).
11. Herrey, E. M. J., and Herrey, H., "Principles of Physics Applied to Traffic Movements and Road Conditions." *Amer. Jour. Physics*, 13: 1, 1-14 (1945).
12. Jones, T. R., and Potts, R. B., "The Measurement of Acceleration Noise—A Traffic Parameter." *Oper. Res.*, 10: 6, 745-763 (Nov.-Dec. 1962).
13. Kometani, E., and Sasaki, T., "On the Stability of Traffic Flow—Report I." *Oper. Res.* (Japan), 2: 11-26 (1958).
14. Kometani, E., and Sasaki, T., "A Safety Index for Traffic With Linear Spacing." *Oper. Res.*, 7: 6, 704-720 (Nov.-Dec. 1959).
15. Kometani, E., and Sasaki, T., "Dynamic Behavior of Traffic with a Nonlinear Spacing-Speed Relationship." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 105-119 (1961).
16. Kometani, E., and Sasaki, T., "Traffic Dynamics." Fifth Japan Road Conf., 1959 (in press).
17. Montroll, E. W., "Acceleration Noise and Vehicular Clustering on Highways." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 147-157 (1961).
18. Newell, G. F., "Nonlinear Effects in the Dynamics of Car-Following." *Oper. Res.*, 9: 2, 209-229 (Mar.-Apr. 1961).
19. Pipes, L. A., "A Proposed Dynamic Analogy of Traffic." Inst. of Transportation and Traffic Eng., Univ. of California, Special Study (July 11, 1950).
20. Pipes, L. A., "An Operational Analysis of Traffic Dynamics." *Jour. Appl. Physics*, 24: 3, 274-281 (Mar.-Apr. 1958).
21. Reuschel, A., "Fahrzeugbewegungen in der Kolonne." *Österreichische Ingenieuren Archiv*, 4: 193-215 (1950).
22. Reuschel, A., "Fahrzeugbewegungen in der Kolonne bei gleichförmig beschleunigtem oder verzögertem Leitfahrzeug." *Zeit. d. österreichischen Ing. u. Arch. Vereins*, 95: 59-62, 73-77 (1950).
23. Sasaki, T., "On the Stability of Traffic Flow — Report II." *Oper. Res.* (Japan), 2: 60-79 (1959).
24. Tuck, E., "Stability of Following in Two Dimension." *Oper. Res.*, 9: 4, 479-495 (July-Aug. 1961).
25. Woods, W. A., "Dynamics of Halting a Line of Traffic." *Eng.* (London), 191: 4960, 655-656 (May 1961).

## PROBABILISTIC FLOW MODELS

1. Anderson, R., Herman, R., and Prigogine, I., "On the Statistical Distribution Function Theory of Traffic Flow." *Oper. Res.*, 10: 2, 180-196 (Mar.-Apr. 1962).
2. Bartlett, M. S., "Some Problems Associated with Random Velocities." *Publ. de l'Inst. de Statis.*, Univ. de Paris, 6: 261-270 (1957).
3. Carleson, L., "En Matematisk Modell fór Landsvágstrafik." *Nord. Matemat. Tidskr.*, 5: 176-180, 213 (in Swedish with English summary) (1957).
4. Gluss, B., "Four Streams of Traffic Converging on a Cross-Road." *Ann. Math. Statis.*, 27: 215-216 (1956).
5. Haight, F. A., "Towards a Unified Theory for Road Traffic." *Oper. Res.*, 6: 6, 813-826 (Nov.-Dec. 1958).
6. Haight, F. A., "Overflow at a Traffic Light." *Biometrika*, 46: Pts. 3 and 4, 420-424 (Dec. 1959).
7. Herman, R., and Weiss, G., "Comments on the Highway-Crossing Problem." *Oper. Res.*, 9: 6, 828-840 (Nov.-Dec. 1961).
8. Kometani, E., "On the Theoretical Solution of Highway Traffic Capacity Under Mixed Traffic." *Memoirs Faculty of Kyoto Univ.*, 17: 79-88 (1955).
9. Miller, A. J., "Traffic Flow Treated as a Stochastic Process." *Theory of*

*Traffic Flow,* Elsevier Publ. Co., pp. 165-174 (1961).
10. Miller, A. J., "Road Traffic Flow Considered as a Stochastic Process." *Proc.,* Cambridge Phil. Soc., 58: Pt. 2 (Apr. 1962).
11. Miller, A. J., "A Queueing Model Road Traffic." *Jour.,* Royal Statis. Soc., Series B, 23: 64-75 (1961).
12. Newell, G. F., "Mathematical Models for Freely-Flowing Highway Traffic." *Oper. Res.,* 3: 2, 176-186 (May 1955).
13. Newell, G. F., "A Theory of Platoon Formation in Tunnel Traffic." *Oper. Res.,* 7: 5, 589-598 (Sept.-Oct. 1959).
14. Newell, G. F., "A Theory of Traffic Flow in Tunnels." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 191-206 (1961).
15. Oliver, R. M., "Distribution of Gaps and Blocks in a Traffic Stream." *Oper. Res.,* 10: 2, 197-217 (Mar.-Apr. 1962).
16. Prigogine, I., "A Boltzmann-Like Approach to the Statistical Theory of Traffic Flow." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 158-164 (1961).
17. Prigogine, I., and Andrews, F. C., "A Boltzmann-Like Approach for Traffic Flow." *Oper. Res.,* 8: 6, 789-797 (Nov.-Dec. 1960).
18. Tanner, J. C., "Delays on a Two-Lane Road." *Jour.,* Royal Statis. Soc., Series B, 23: 38-63 (1961).
19. Tanner, J. C., "A Simplified Model for Delays in Overtaking on a Two-Lane Road." *Jour.,* Royal Statis. Soc., Series B, 20: 2, 408-414 (1958).

## SIMULATION OF TRAFFIC FLOW

1. Beckman, M., McGuire, C. B., and Winsten, C. B., *Studies in the Economics of Transportation,* Yale Univ. Press (1956).
2. Benhard, F. G., "Simulation of a Traffic Intersection on a Digital Computer." M. S. Thesis, Univ. of California, Los Angeles (June 1959).
3. Gerlough, D. L., "Analogs and Simulators for the Study of Traffic Problems." Proc. Sixth California Street and Highway Conf., pp. 82-83 (1954).
4. Gerlough, D. L., "Simulation of Freeway Traffic by an Electronic Computer." *HRB Proc.,* 35: 543-547 (1956).
5. Gerlough, D. L., and Mathewson, J. H., "Approaches to Operational Problems in Street and Highway Traffic." *Oper. Res.,* 4: 1, 32-41 (Feb. 1956).
6. Gerlough, D. L., "Simulation of Freeway Traffic by Digital Computers." Proc. of Conf. on Increasing Highway Engineering Productivity, held at Georgia Inst. of Technology, July 9-11, 1956, U. S. Bureau of Public Roads (1957).
7. Gerlough, D. L., "A Comparison of Techniques for Simulating the Flow of Discrete Objects." Paper presented at National Simulation Conf., Dallas, Tex. (Oct. 23-25, 1958).
8. Gerlough, D. L., "Traffic Inputs for Simulation on a Digital Computer." *HRB Proc.,* 38: 480-492 (1959).
9. Glickstein, A., Findley, L. D., and Levy, S. L., "A Study of the Application of Computer Simulation Techniques to Interchange Design Problems." HRB Bull. 291, pp. 139-162 (1962).
10. Goode, H. H., "Simulation—Its Place in System Design." *Proc.,* Inst. Radio Eng., 39: 12, 1501-1506 (Dec. 1951).
11. Goode, H. H., Pollmar, C. H., and Wright, J. B., "The Use of a Digital Computer to Model a Signalized Intersection." *HRB Proc.,* 35: 548-557 (1956).
12. Goode, H. H., "The Application of a High-Speed Computer to the Definition and Solution of the Vehicular Traffic Problem." *Oper. Res.,* 5: 6, 775-793 (Dec. 1957).
13. Goode, H. H., and True, W. C., "Vehicular Traffic Intersections." Paper presented at 13th Nat. Meeting, Assoc. of Computing Machinery (June 11-13, 1958).
14. Helly, W., "Dynamics of Single-Lane Vehicular Traffic Flow." *Research Report No. 2,* Center for Operations Research, Massachusetts Inst. of Tech., adapted from Ph.D. Thesis, Massachusetts Inst. of Tech. (Oct. 1959).
15. Helly, W., "Simulation of Bottlenecks in Single-Lane Traffic Flow." *Theory*

of *Traffic Flow,* Elsevier Publ. Co., pp. 205-238 (1961).

16. Jorgensen, N. O., "Determination of the Capacity of Road Intersections by Model Testing." *Ingenioren* (Internat. Ed.), 5: 3, 99-101 (1961).
17. Kell, J. H. "Analyzing Vehicular Delay at Intersections Through Simulation." HRB Bull. 356, pp. 28-39 (1962).
18. Lewis, E., "A Digital Model for Vehicular Traffic." Proc. of Symposium on Digital Simulation Techniques for Predicting the Performance of Large-Scale Systems, held at Univ. of Michigan, pp. 287-293 (May 1960).
19. Mathewson, J. H., Trautman, D. L., and Gerlough, D. L., "Study of Traffic Flow by Simulation." *HRB Proc.,* 34: 522-530 (1955).
20. Perchonok, P., and Levy, S. L., "Application of Digital Simulation Techniques to Freeway On-Ramp Traffic Operations." Final Report to Bureau of Public Roads, Midwest Res. Inst. (Nov. 24, 1959); *HRB Proc.,* 39: 506-523 (1960).
21. Stark, M. C., "Computer Simulation of Street Traffic." Nat. Bur. of Standards Tech. Note 119 (Nov. 1961).
22. Stark, M. C., "Computer Simulation of Traffic on Nine Blocks of a City Street." HRB Bull. 356, pp. 40-47 (1962).
23. Trautman, D. L., Davis, H., Heilfron, J., Ho, E. C., Mathewson, J. H., and Rosenbloom, A., "Analysis and Simulation of Vehicular Traffic Flow." *Res. Rpt. No. 20,* Inst. of Transportation and Traffic Eng., Univ. of California (Dec. 1954).
24. Walton, J. R., and Douglas, R. A., "A LaGrangian Approach to Traffic Simulation on a Digital Computer." HRB Bull. 356, 48-50 (1962).
25. Wong, S. Y., "Traffic Simulator With a Digital Computer." Proc. Western Joint Computer Conf., pp. 92-94 (1956).

## DISTRIBUTION OF TRAFFIC OVER A NETWORK

1. Beckman, M., McGuire, C. B., and Winsten, C. B., *Studies in the Economics of Transportation.* Yale University Press (1956).
2. Charnes, A., and Cooper, W. W., "Extremal Principles for Simulating Traffic Flow in a Network." *Proc.,* Nat. Acad. of Sciences, 44: 2, 201-204 (Feb. 1958).
3. Charnes, A., and Cooper, W. W., "Theories of Traffic Network." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 85-96 (1961).
4. Hoffman, W., and Pavley, R., "Method for Solution of the *N*th Best Path." *Jour.,* Assoc. Computing Machinery, 6: 4, 506-514 (Oct. 1959).
5. Hoffman, W., and Pavley, R., "Applications of Digital Computers to Problems in the Study of Vehicular Traffic." Proc. Western Joint Computer Conf., pp. 159-161 (1958).
6. Moore, E. F., "The Shortest Path Through a Maze." Proc. Internat. Symposium on Theory of Switching. Pt. II. April 2-5, 1957, Harvard Univ. Press, pp. 285-292 (1959).
7. Pandit, S. M. N., "The Shortest-Route Problem — An Addendum." *Oper. Res.,* 9: 1, 129-132 (Jan.-Feb. 1961).
8. Pollack, M., "The Maximum Capacity Route Through a Network." *Oper. Res.,* 8: 5, 733-736 (Sept.-Oct. 1960).
9. Prager, W., "On the Design of Communication and Transportation Networks." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 97-104 (1961).
10. Prager, W., "On the Role of Congestion in Transportation Problems." *Zeits. f. angewandte Mathematik und Mechanik,* 35: 264-268 (1955).
11. Prager, W., "Problems in Traffic and Transportation." Proc. of Symposium on Operations Research in Business and Industry, Midwest Research Inst., Kansas City, Mo. (Apr. 1954).
12. Voorhees, A. M., "A General Theory of Traffic Movement." *Proc.,* Inst. Traffic Eng., 25: 46-56 (1955).
13. Wardrop, J. G., "Some Theoretical Aspects of Road Traffic Research." *Proc.,* Inst. Civil Eng., Pt. II, 1: 2, 325-378 (June 1952).
14. Wardrop, J. G., "The Distribution of

Traffic on a Road System." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 57-78 (1961).
15. Whiting, P. D., and Hillier, J. A., "A Method for Finding the Shortest Route Through a Road Network." *Oper. Res. Quart.*, 11: 1-2, 37-40 (Mar.-June 1960).

## MISCELLANEOUS PROBABILISTIC AND STATISTICAL STUDIES

1. Adams, W. F., "Road Traffic Considered as a Random Series." *Jour.*, Inst. Civil Eng., 4: 121-130 (Nov. 1936).
2. Beckman, M., McGuire, C. B., and Winsten, C. B., *Studies in the Economics of Transportation*, Yale Univ. Press (1956).
3. Egert, P., "The Mathematical Principles of Traffic Statistics." Technische und volkswirtschaftliche Berichte des Wirtschafts- und Verkehrsministerium, Nordrhein-Westfalen, 34 pp. (1954).
4. Garwood, F., "An Application of the Theory of Probability to the Operation of Vehicular-Controlled Traffic Signals." *Jour.*, Royal Statis. Soc., Supplement, 7: 1, 65-77 (1940).
5. 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).
6. Greenshields, B. D., Shapiro, D., and Ericksen, E. L., "Traffic Performance at Urban Street Intersections." Tech. Rpt. No. 1, Yale Bur. of Highway Traffic, pp. 73-109 (1947).
7. Haight, F. A., "The Generalized Poisson Distribution." *Ann.*, Inst. Statis. Math. (Tokyo), 11: 2, 101-105 (1959).
8. Haight, F. A., Jacobson, A. S., "Some Mathematical Aspects of the Parking Problem." *HRB Proc.*, 41: 363-374 (1962).
9. Kinzer, J. P., "Application of the Theory of Probability to Problems of Highway Traffic." Thesis submitted for degree of B. C. E., Polytechnic Inst. of Brooklyn, June 1, 1933. Abstr., *Proc.*, Inst. Traffic Eng., 5: 118-124 (1934).
10. Klamkin, M. S., Newman, D. J., "A Parking Problem." *Rev. Soc. Indus. and Appl. Math.*, 4: 3, 257-258 (July 1962).
11. Little, J. D. C., "Approximate Expected Delays for Several Maneuvers by a Driver in Poisson Traffic." *Oper. Res.*, 9: 1, 39-52 (Jan.-Feb. 1961).
12. Mayne, A. J., "Some Further Results in the Theory of Pedestrians and Road Traffic." *Biometrika*, 41: 375-389 (1954).
13. Moskowitz, K., "Waiting for a Gap in a Traffic Stream." *HRB Proc.*, 33: 385-395 (1954).
14. Oliver, R. M., "A Traffic Counting Distribution." *Oper. Res.*, 9: 6, 802-810 (Nov.-Dec. 1961).
15. Oliver, R. M., and Bisbee, E. F., "Queueing for Gaps in High-Flow Traffic Streams." *Oper. Res.*, 10: 1, 105-114 (Jan.-Feb. 1962).
16. Oliver, R. M., "A Note on a Traffic Counting Distribution." *Oper. Res. Quart.*, 13: 2, 171-178 (June 1962).
17. Oliver, R. M., "Delays to Aircraft Serviced by the Glide-Path." *Oper. Res. Quart.*, 13: 2, 201-209 (June 1962).
18. Raff, M. S., "The Distribution of Blocks in an Uncongested Stream of Automobile Traffic." *Jour.*, Amer. Statis. Assoc., 46: 62-89 (Mar. 1951).
19. Schuhl, A., "The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highways." *Poisson and Traffic*, The Eno Foundation for Highway Traffic Control, pp. 59-72 (1955).
20. Schuhl, A., "Le Calcul des Probabilites et la Circulation des Vehicules sur une Chausse a Deux Voies." *Annales des Ponts et Chaussees*, 125: 631-663 (1955).
21. Stanly, A. L., "A Study of the Variables Involved and the Statistical Techniques Utilized in Conducting and Evaluating a Traffic Experiment on Traffic Delays at Overloaded Intersection." Ph.D. Thesis, Univ. of California, Los Angeles (Aug. 1949).
22. Tanner, J. C., "The Delay to Pedestrians Crossing a Road." *Biometrika*, 38: Pts. 3 and 4, 383-392 (Dec. 1951).

23. Tanner, J. C., "Problems in the Interference of Two Queues." *Biometrika*, 40: Pts. 1 and 2, 58-69 (June 1953).
24. Wardrop, J. G., "Some Theoretical Aspects of Road Traffic Research." *Proc.*, Inst. Civil Eng., Pt. II, 1: 2, 325-378 (June 1952).
25. Weiss, G. H., and Maradudin, A. A., "Some Problems in Traffic Delay." *Oper. Res.*, 10: 1, 74-104 (Jan.-Feb. 1962).

## STUDIES OF MECHANICS AND GEOMETRY

1. Clayton, A. J. H., "Road Traffic Calculations." *Jour.*, Inst. Civil Eng., 16: 7, 247-284 (June 1941); 16: 8, 588-594 (Oct. 1941).
2. Gazis, D., Herman, R., and Maradudin, A., "The Problem of the Amber Signal in Traffic Flow." *Oper. Res.*, 8: 1, 112-132 (Jan.-Feb. 1961).
3. Goodrich, E. P., "The Application of Mathematics to Traffic Engineering." *Proc.*, Inst. Traffic Eng., pp. 1-24 (1931-1934).
4. Herwald, N., "Some Theoretical Aspects of Control of Road Traffic." *Jour. Scient. Instr.*, 14: 12, 393-401 (Dec. 1937).
5. Mori, M., "Traffic Characteristics of Roads Under Mixed Traffic Conditions." *Traffic Eng.*, 30: 1, 23-28 (Oct. 1959).
6. Peleg, M., "Encounter of Vehicles at Intersections." *Bull.*, Res. Council (Israel), C7: 1, 55-60 (Apr. 1959).
7. Rashevsky, N., "Some Remarks on the Mathematical Aspects of Automobile Driving." *Bull. Math. Biophysics*, 21: 299-308 (1959).

## EXPERIMENTAL STUDIES OF CAPACITY, SPEED, AND FLOW-CONCENTRATION RELATIONSHIPS

1. Berry, D. S., and Belmont, D. M., "Distribution of Vehicle Speeds and Travel Times." Proc. Second Berkeley Symp. on Mathematical Statistics and Probability, Univ. of California Press, pp. 589-562 (1951).
2. Crawford, A., and Taylor, D. H., "Driver Behavior at Traffic Lights: Critical Amber Period." *Traffic Eng. and Control*, London, 3: 8, 473-478 (Dec. 1961).
3. Creighton, R. L., "Speed-Volume Relationships on Signalized Roads." *Research News*, Chicago Area Transp. Study, 1: 11, 6-11 (June 21, 1957).
4. Edie, L. C., and Foote, R. S., "Traffic Flow in Tunnels." *HRB Proc.*, 37: 334-344 (1958).
5. Edie, L. C., and Foote, R. S., "Experiments on Single-Lane Flow in Tunnels." *Theory of Traffic Flow*, Elsevier Publ. Co., pp. 175-192 (1959).
6. Edie, L. C., and Foote, R. S., "Effect of Shock Waves on Tunnel Traffic." *HRB Proc.*, 39: 492-505 (1960).
7. Forbes, T. W., "Speed, Headway, and Volume Relationships on a Freeway." *Proc.*, Inst. Traffic Eng., 21: 103-126 (1951).
8. Graham, E. F., and Chenu, D. C., "A Study of Unrestricted Platoon Movement of Traffic." *Traffic Eng.*, 32: 7, 11-13 (Apr. 1962).
9. Greenberg, H., and Daou, A., "The Control of Traffic Flow to Increase the Flow." *Oper. Res.*, 8: 4, 524-532 (July-Aug. 1960).
10. May, A. D., Jr., and Wagner, F. A., "Headway Characteristics and Interrelationships of the Fundamental Characteristics of Traffic Flow." *HRB Proc.*, 39: 524-547 (1960).
11. Normann, O. K., and Walker, W. P., *Highway Capacity Manual*, U. S. Bureau of Public Roads, 147 pp. (1950).
12. Olcott, E. S., "The Influence of Vehicular Speed and Spacing on Tunnel Capacity." *Oper. Res.* 3: 2, 147-167 (May 1955).
13. Olson, P. L., and Rothery, R. W., "Driver Response to the Amber Phase of Traffic Signals." *Oper. Res.*, 9: 5, 650-663 (Sept.-Oct. 1961).
14. Wardrop, J. G., "Some Theoretical Aspects of Road Traffic Research." *Proc.*, Inst. Civil Eng., Pt. II, 1: 2, 325-378 (June 1952).
15. Wardrop, J. G., "The Capacity of Roads." *Oper. Res. Quart.*, 5: 1, 14-24 (Mar. 1954).

## INTERSECTION SITUATIONS


1. Castoldi, L., "Queue Alternance and Traffic Flow at a Crossroad." *Boll. del Centro per La Ricerca Operative,* No. 5-6, pp. 1-13 (1957).
2. Gerlough, D. L., "Some Problems in Intersection Traffic Control." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 10-27 (1959).
3. LaVallee, R. S., "Scheduling of Traffic Signals by Linear Programming." *HRB Proc.,* 35: 534-542 (1956).
4. Newell, G. F., "Statistical Analysis of the Flow of Highway Traffic Through a Signalized Intersection." *Quart. Appl. Math.,* 13: 4, 353-369 (Jan. 1956).
5. Newell, G. F., "The Effect of Left Turns on the Capacity of a Traffic Intersection." *Quart. Appl. Math.,* 17: 1, 67-76 (Apr. 1959).
6. Pacey, G. M., "The Progress of a Bunch of Vehicles Released from a Traffic Signal." Brit. Road Research Lab., unpubl. Res. Note RN/2665/GMP (Jan. 1956).
7. Tanner, J. C., "A Theoretical Analysis of Delays at an Uncontrolled Intersection." *Biometrika,* 49: 1 and 2, 163-170 (1962).
8. Uematu, T., "On Traffic Control at an Intersection Controlled by Repeated Fixed-Cycle Traffic Light." *Ann.,* Inst. Statis. Math. (Tokyo), 9: 87-107 (1958).
9. Von Stein, W., "Traffic Flow with Pre-Signals and the Signal Funnel." *Theory of Traffic Flow,* Elsevier Publ. Co., pp. 28-56 (1959).
10. Webster, F. V., "Traffic Signal Settings." Road Res. Tech. Paper No. 39, Brit. Road Research Lab. (1958).


## MISCELLANEOUS


1. Edie, L. C., "Traffic Delays at Toll Booths." *Oper. Res.,* 2: 2, 107-138 (May 1954).
2. Edie, L. C., "Expectancy of Multiple Vehicular Breakdowns in a Tunnel." *Oper. Res.,* 3: 4, 513-522 (Nov. 1955); 4: 5, 609-619 (Oct. 1956).
3. Hankin, B. D., and Wright, R. A., "Passenger Flow in Subways." *Oper. Res. Quart.,* 9: 2, 81-88 (1958).
4. Jensen, A., "Traffic Theory as an Aid in the Planning and Operation of the Road Grid." *Ingenioren* (Internat. Ed.), 1: 2, 48-61 (1957).
5. Kometani, E., and Kato, A., "On the Theoretical Capacity of Off-Street Parking Space." *Memoirs,* Faculty of Eng., Kyoto Univ., 18: 4, 315-328 (Oct. 1956).
6. Muranyi, T., "Scientific Foundations of Highway System Planning." *Acta Tech.,* Academiae Scientiarum Hungaricae (Hungary), 16: 13 (1956).
7. Newell, G. F., "The Flow of Traffic Through a Sequence of Signalized Traffic Signals." *Oper. Res.,* 8: 3, 390-405 (May-June 1960).
8. Pacey, G. M., "Some Problems in the Study of Road Traffic Flow." M.Sc. Thesis, Univ. of Manchester.
9. Platt, F. M., "Operations Research on Driver Behavior." *Traffic Eng.,* 31: 12, 11-16 (Sept. 1961).
10. Rashevsky, N., "Contributions to the Mathematical Biophysics of Automobile Driving." *Bull. Math. Biophysics,* 23: 1, 19-29 (Mar. 1961).
11. Smeed, R. J., "Theoretical Studies and Operational Research on Traffic and Traffic Congestion." *Bull. de L'Inst. Internat. de Statis.* (Stockholm), 36: Pt. 4, 347-375 (1958).
12. Turner, J. K., and Wardrop, J. G., "The Variation of Journey Time in Central London." Brit. Road Res. Lab., unpublished Res. Note No. RN/1511/JKT (Feb. 1951).
13. Turner, J. K., and Wardrop, J. G., "The Variation in the Time Stopped at Controlled Intersections in Central London." Brit. Road Res. Lab., unpublished Research Note No. RN/2158/JKT. JGW (Feb. 1954).
14. Welding, P. I., and Stringer, J., "A Problem in Vehicle Fuel Consumption." *Oper. Res. Quart.* (London), 11: 4, 197-204 (Dec. 1960).

# Appendix B
## ADDRESSES OF PUBLISHERS

Academiae Scientiarum Hungaricae
Akademiai Kiado
Publishing House of the Hungarian Academy of Sciences
Alkotmany U. 21,
Budapest V, Hungary

*American Journal of Physics*
American Institute of Physics
335 East 45th Street
New York 17, New York

American Statistical Assoc.
1757 K Street, N. W.
Washington 6, D. C.

*Annals of Mathematical Statistics*
Charles E. Tuttle Co.
Tokyo, Japan

*Annales des Ponts et Chaussees*
5 Rue Jules-Lefebore
Paris, France

Assoc. for Computing Machinery
2 East 63rd Street
New York 21, New York

*Biometrika*
Biometrika Office
University College
Gower Street
London W. C.1, England

*Bulletin of Math. Biophysics*
University of Chicago Press
5750 Ellis Avenue
Chicago 37, Illinois

Bureau of Public Roads
1717 H Street, N. W.
Washington 25, D. C.

California Street & Highway Conf.
Institute of Transportation and Traffic Engineering
University of California
Berkeley, California

Cambridge Philosophical Society
Cambridge University Press
200 Euston Road
London N.W. 1, England

*C.A.T.S. Research News*
Chicago Area Transportation Study
4812 West Madison Street
Chicago 44, Illinois

Center for Operations Research
Massachusetts Inst. of Technology
Cambridge 39, Massachusetts

*Civil Engineering and Public Works Review*
8 Buckingham Street
London W.C. 2, England

*Communication on Pure and Applied Mathematics*
Inst. for Mathematical Sciences
New York University
New York, New York

*Engineering*
36 Bedford Street
Strand, London W.C.2, England

Elsevier Publishing Company
D. Van Nostrand Company
210 Alexander Street
Princeton, New Jersey

Harvard University Press
Cambridge, Massachusetts

Highway Research Board
2101 Constitution Avenue
Washington 25, D. C.

*Ingenioren*
Vester Farimagsgade 31
Copenhagen 5, Denmark

Inst. of Mathematical Statistics
Department of Statistics
Stanford University
Stanford, California

Institute of Radio Engineers
1 East 79th Street
New York 21, New York

Inst. of Statistical Mathematics
*(Annals)*
1 Azabu Huzimi-Tyo
Minatoku, Tokyo, Japan

Institute of Traffic Engineers
*(Traffic Engineering)*
Suite 506, 1725 DeSales St., N. W.
Washington 6, D. C.

Institute of Transport
80 Portland Place
London S.W. 1, England

Inst. of Transportation and Traffic Engineering
University of California
Berkeley 4, California
(or Los Angeles 24, Calif.)

Inst. Internat. de Statistique
2 Oostduinlaan
The Hague, Netherlands

Institution of Civil Engineers
Great George Street
Westminster, London S.W.1, England

Interscience Publishers
250 Fifth Avenue
New York 1, New York

*Journal of Applied Physics*
American Institute of Physics
335 East 45th Street
New York 17, New York

*Jour. of Scientific Instruments*
Inst. of Physics and the Physical Society
47 Belgrave Square
London S.W.1, England

Kyoto University
Faculty of Engineering
Kyoto, Japan

Midwest Research Institute
425 Volker Boulevard
Kansas City 10, Missouri

National Simulation Conference
Institute of Radio Engineers
1 East 79th Street
New York 21, New York

National Bureau of Standards
Connecticut Avenue, N. W.
Washington, D. C.

*Operational Research Quarterly*
Operational Research Society
64 Cannon Street
London E.C.4, England

Operations Research Center
University of California
Berkeley, California

Operations Research Soc. of America
*(Operations Research)*
Mount Royal & Guilford Avenues
Baltimore 2, Maryland

Operations Research Society of Japan
Japan Railway Association
4-Otemachi 2-Chome
Chiyoda-Ku, Tokyo, Japan

Osaka University
Faculty of Engineering
Higashinada
Osaka, Japan

Permanent International Assoc. of Road Congresses
43 Avenue President Wilson
Paris 16 E, France

Proceedings of the Meetings of Japan Road Conference
Japan Road Association
c/o Kensetsusho Kanto Chiho Kensetsukyoku
1 Sannen-Cho
Chiyoda-Ku, Tokyo, Japan

*Public Roads*
Superintendent of Documents
Washington 25, D. C.

*Quarterly of Applied Mathematics*
Brown University
Providence 12, Rhode Island

Research Council of Israel
Weizmann Science Press of Israel
P. O. Box 801
Jerusalem, Israel

*Roads and Road Construction*
Carriers Publishing Company
147 Victoria Street
London S.W.1, England

British Road Research Lab.
Harmondsworth, Middlesex, England

Royal Society of London
Burlington House
Piccadilly, London W.1, England

Royal Society of Arts
6 John Adam Street
Adelphi, London W.C.2, England

Royal Statistical Society
21 Bentinck Street
London W.1, England

Society for Industrial and Applied Mathematics
SIAM Publications
Box 7541
Philadelphia, Pennsylvania

The Eno Foundation for Highway Traffic Control
Saugatuck, Connecticut

*Traffic Engineering and Control*
Ludgate Hill
London E.C.4, England

University of Manchester
Manchester, England

Western Joint Computer Conf.
American Institute of Electrical Engineers
345 East 47th
New York 17, New York

# Appendix C
# ILLUSTRATION CREDITS

Grateful acknowledgment is made to the respective publishers for permission to use the following items as the basis of certain of the illustrations prepared for use in this publication:

| *Figure* | *Data Source* |
|---|---|
| 2.1 | *Oper. Res.*, Vol. 10, No. 6, pp. 745-763 (Jones and Potts) |
| 2.3 | *Oper. Res.*, Vol. 7, No. 4, pp. 499-505 (Gazis, Herman and Potts) |
| 2.4 | Private communication (R. W. Rothery) |
| 2.5 | *Oper. Res.*, Vol. 7, No. 1, p. 105, Fig. 7 (Herman, Montroll, Potts and Rothery) |
| 3.5 | Paper presented before Western Section, I.T.E., 1960 (Kell) |
| 3.8 | "A Volume Warrant for Urban Stop Signs" (Raff) The Eno Foundation, Saugatuck, Conn. (1950) |
| 3.9 | *Oper. Res. Quart.*, Vol. 6, No. 3, pp. 120-128 (Cohen, Dearnaley and Hansel) |
| 3.11 | *Oper. Res.*, Vol. 9, No. 6, pp. 828-840 (Herman and Weiss) |
| 3.12 | *Oper. Res.*, Vol. 5, No. 2, pp. 161-200 (Saaty) |
| 3.20 | *Biometrika*, Vol. 38, pp. 383-392 (Tanner) |
| 3.26 | *Jour. Roy. Stat. Soc.*, Series B, No. 23, pp. 64-75 (Miller) |
| 3.27 | "Studies in Economics of Transportation" (Beckmann, McGuire and Winsten) Yale Univ. Press (1956) |
| 3.32, 3.33, 3.34 | *Jour. Roy. Stat. Soc.*, Suppl., Vol. 7, No. 1, pp. 65-77 (Garwood) |
| 3.38, 3.39 | *Jour. Roy. Stat. Soc.*, Series B, No. 23, pp. 38-63 (Tanner) |
| 3.41 | *Oper. Res.*, Vol. 11, No. 2, pp. 236-247 (Miller) |
| 3.42, 3.43 | *Memoirs, Faculty of Engg.*, Kyoto Univ., Vol. 17, pp. 79-88 (Kometani) |
| 3.50, 3.51 | *Oper Res.*, Vol. 2, No. 2, pp. 107-138 (Edie) |
| 3.52 | *Oper. Res.*, Vol. 8, No. 3, pp. 390-405 (Newell) |
| 4.17, 4.18 | Report from Midwest Res. Inst. to U.S. Bur. Pub. Roads (Perchonok and Levy) (Nov. 24, 1959) |