Chapter 4

SIMULATION OF TRAFFIC FLOW

D. L. Gerlough
Planning Research Corporation

Contents

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

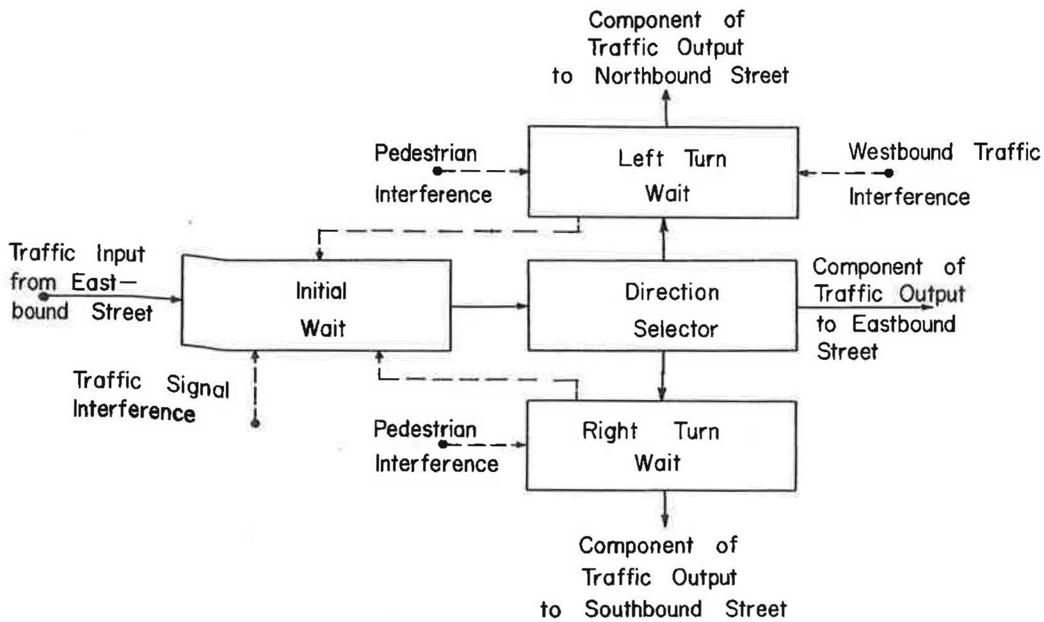


Figure 4.1. Block diagram of one approach of a generalized intersection having one lane in each direction.

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.

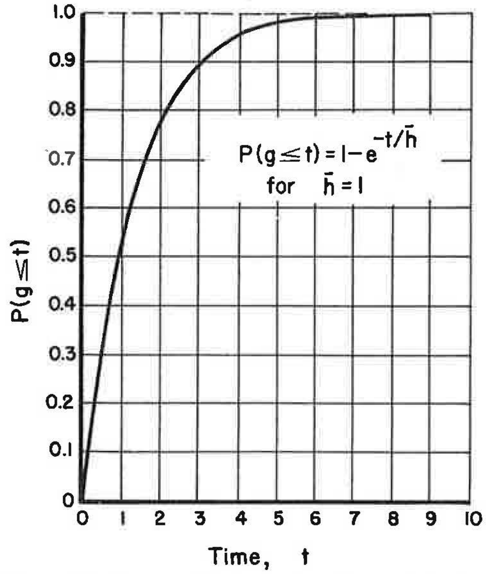


Figure 4.2. Exponential distribution of gaps (inter-arrival times). Ordinate is the probability of a gap t or less; $\bar{h}$ is the average inter-arrival time.

(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 $u_a \leq u_d$);

Δ = 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

$$\Delta = 1u_d - 1u_a = u_d - u_a \tag{4.1}$$

and

$$t = \frac{\Delta}{u_d} = \frac{u_d - u_a}{u_d} = 1 - \frac{u_a}{u_d} \tag{4.2}$$

If during the ith second n_i vehicles are traveling at reduced speeds, the total time lost, τ, in car-seconds, will be

$$\tau = \sum_i \sum_j n_{ij}\, t_{ij} \tag{4.3}$$

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:

$$B = \frac{\tau}{Tq} \tag{4.4}$$

(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

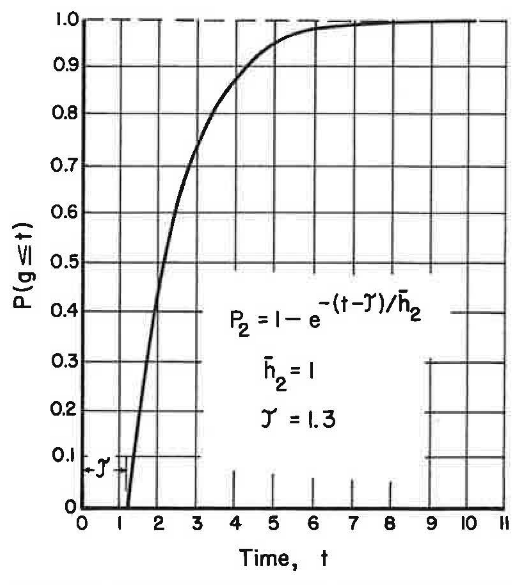


Figure 4.3. Exponential distribution of gaps shifted from origin. Ordinate is the probability of a gap $(t - \tau)$ or less. Average inter-arrival time is $(\bar{h}_2 + \tau)$.

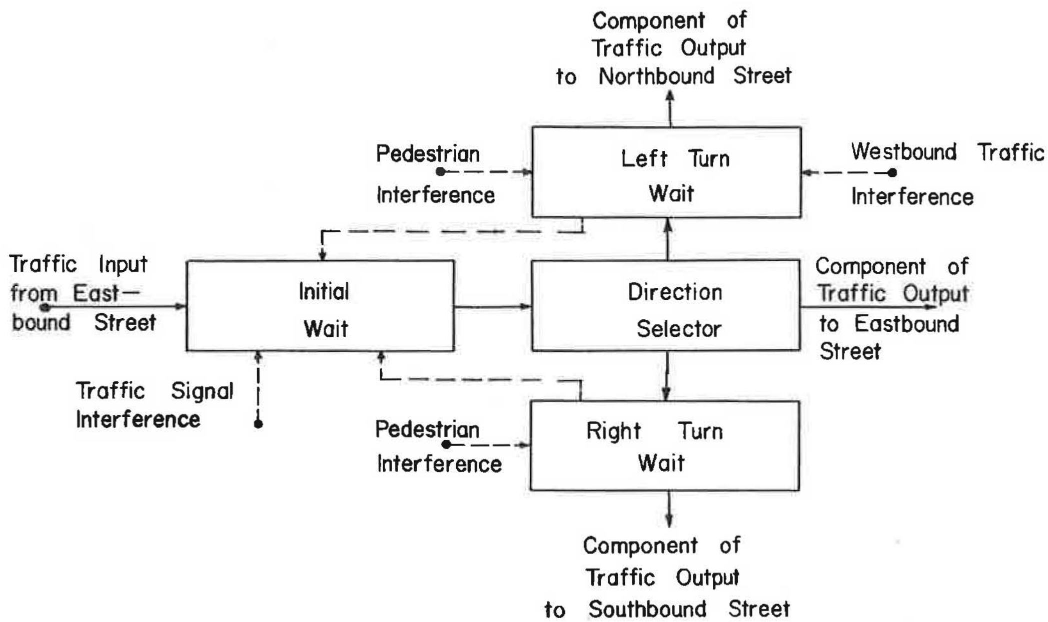


Figure 4.1. Block diagram of one approach of a generalized intersection having one lane in each direction.

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.

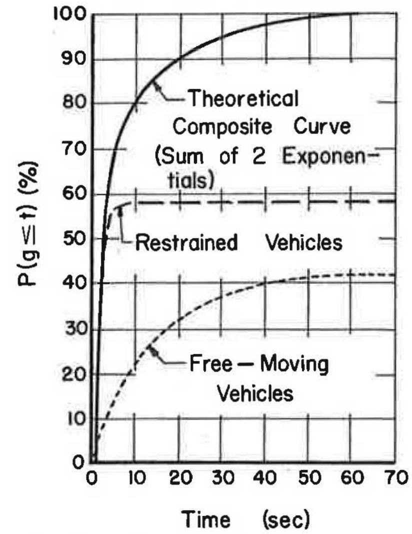


Figure 4.4. Composite exponential curve for inter-arrival times.

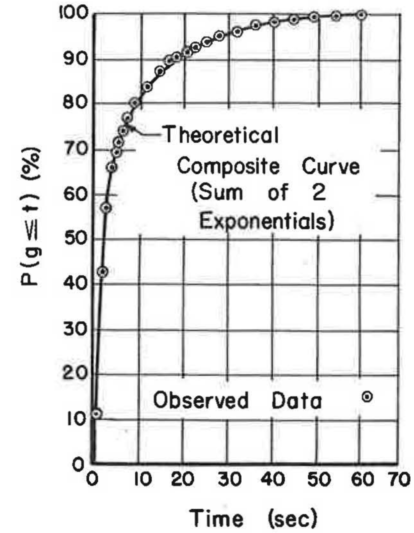


Figure 4.5. Composite exponential curve fitted to observed data.

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.

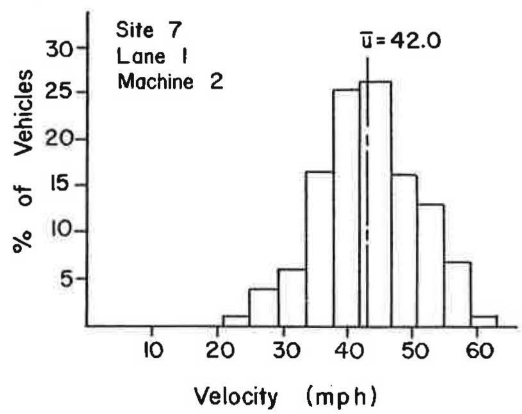


Figure 4.6. Typical speed distribution curve.

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-

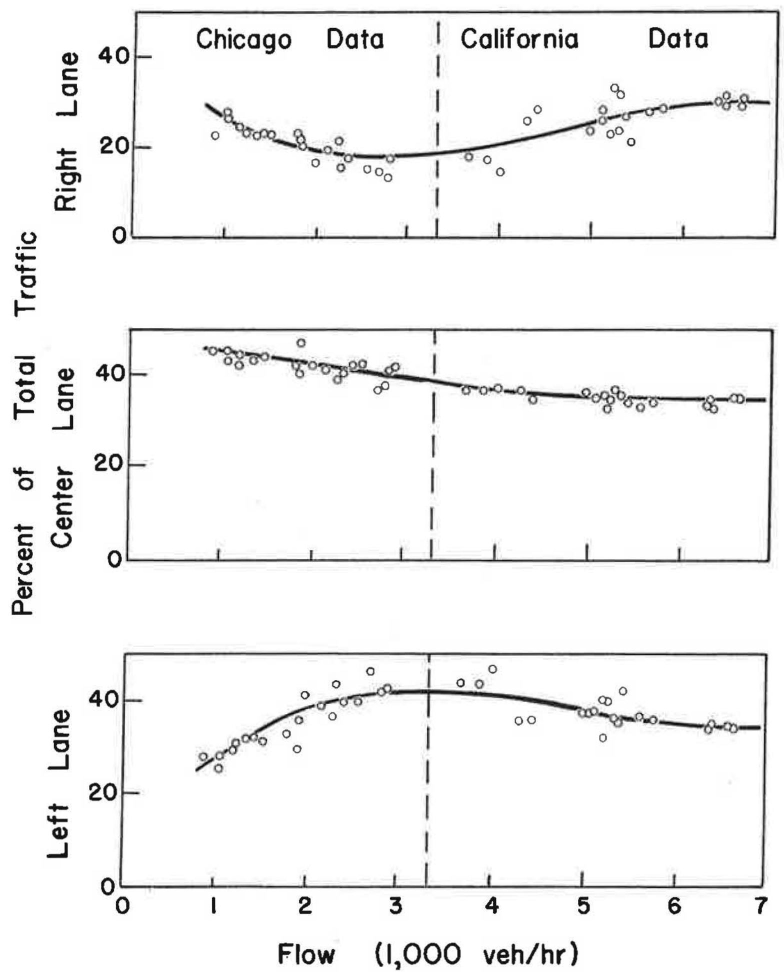


Figure 4.7 Distribution of traffic among lanes of a three-lane freeway.

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 mth random number;
ρ = 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 ρ 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.

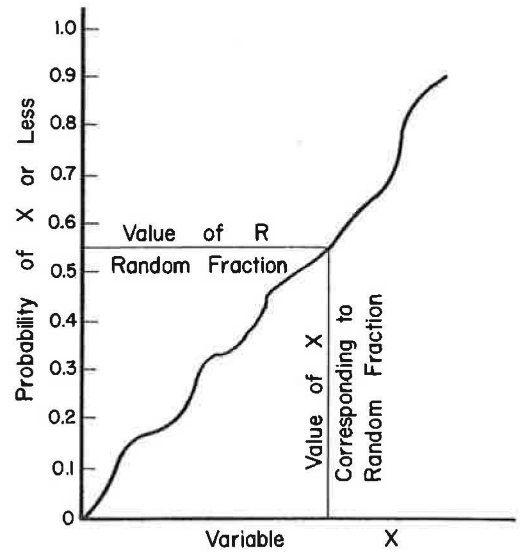


Figure 4.8. Generalized cumulative probability distribution.

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, ρ	Typical Computer with This Base
10^{10}	7^{11}	IBM 650
2^{35}	5^{13}	IBM 709
2^{36}	5^{13}	

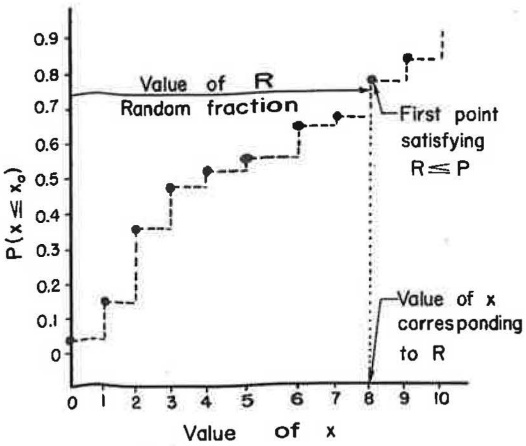


Figure 4.9. Cumulative probability distribution. Points represent tabular values; broken line, distribution for discrete values of x.

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 ith point. The value of x for the ith 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, τ, which can be maintained by vehicles. This may be stated: "The probability of a gap between successive vehicles of less than τ is zero." This phenomenon may be represented by an exponential curve shifted to the right by an amount τ, or

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

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

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

Solving Eq. 4.11 for t gives

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

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

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

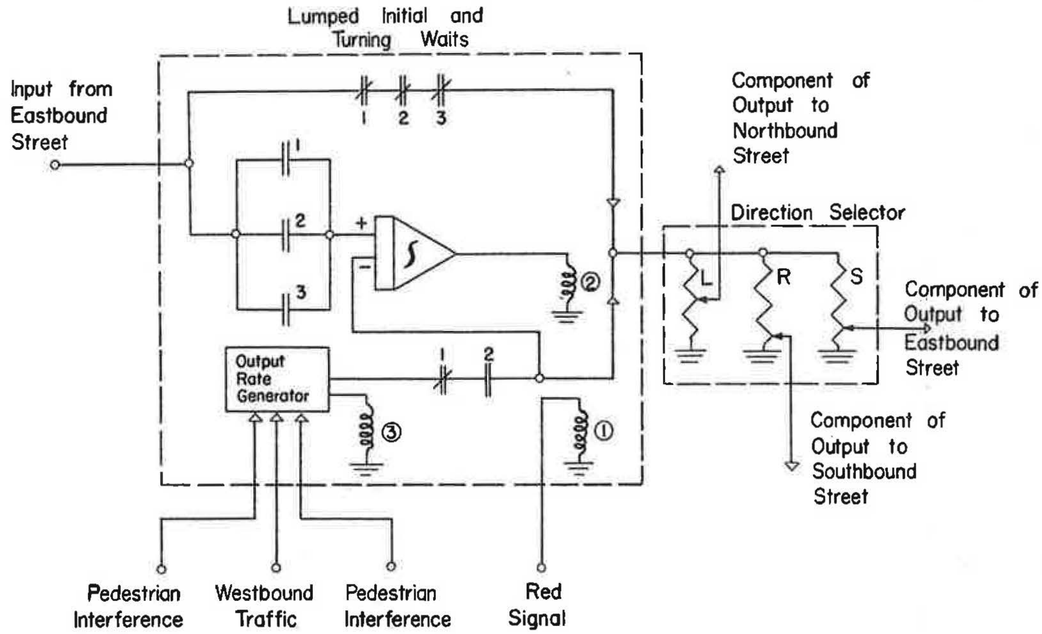


Figure 4.10. Portion of continuous-variable simulator, showing elements for representation of traffic entering from the west (see Fig. 4.11 for notation).

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."

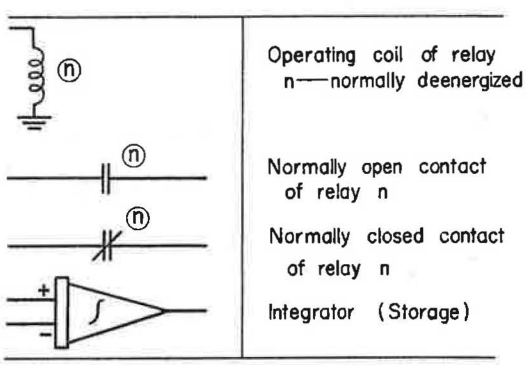


Figure 4.11. Symbols for continuous-variable simulator of Figure 4.10.

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 ρ, α, λ 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 λ (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.)

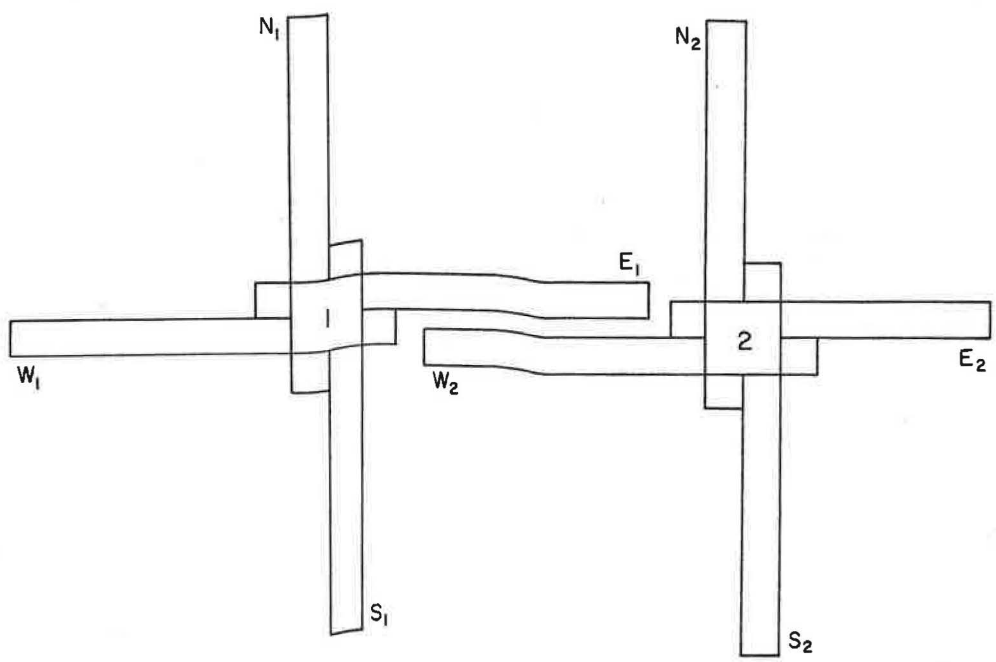


Figure 4.12. Crossblock representation of two intersections.

"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.

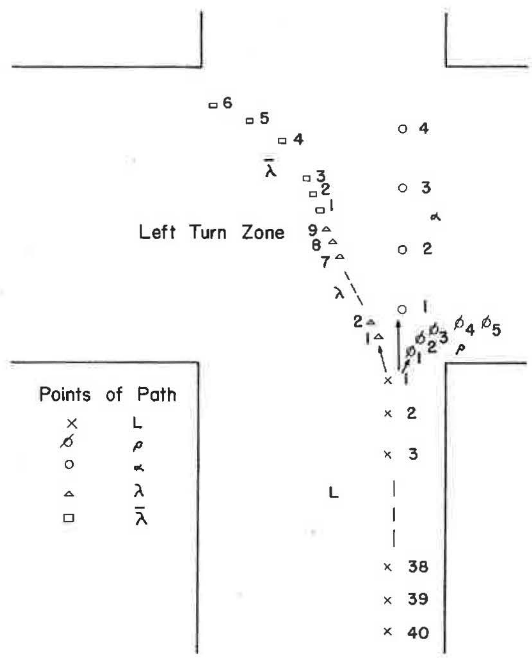


Figure 4.13. Computer "crossblock."

"The cars on α, ρ, 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.

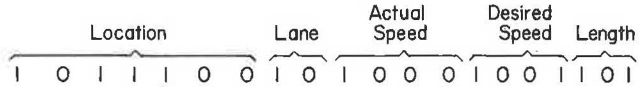
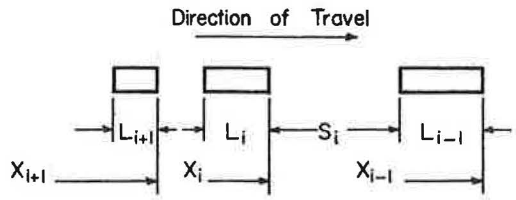


Figure 4.14. Typical arrangement of information in code word.



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

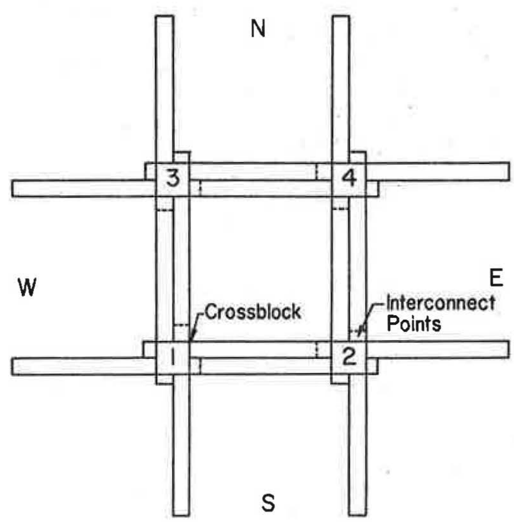


Figure 4.16. Intersection network formed from four crossblocks.

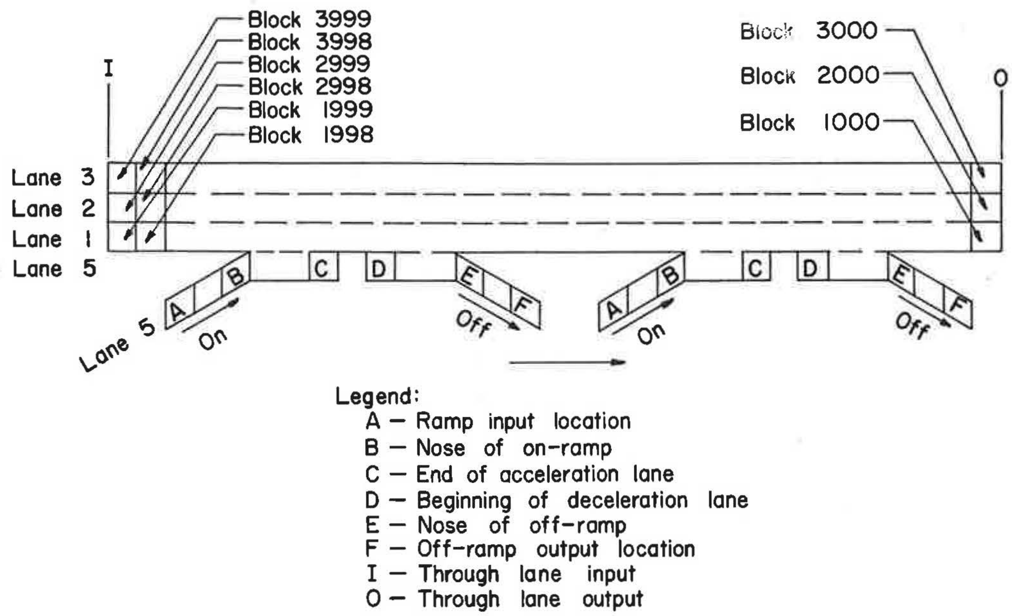


Figure 4.17. Organization of roadway layout for programming of computer simulation.

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.

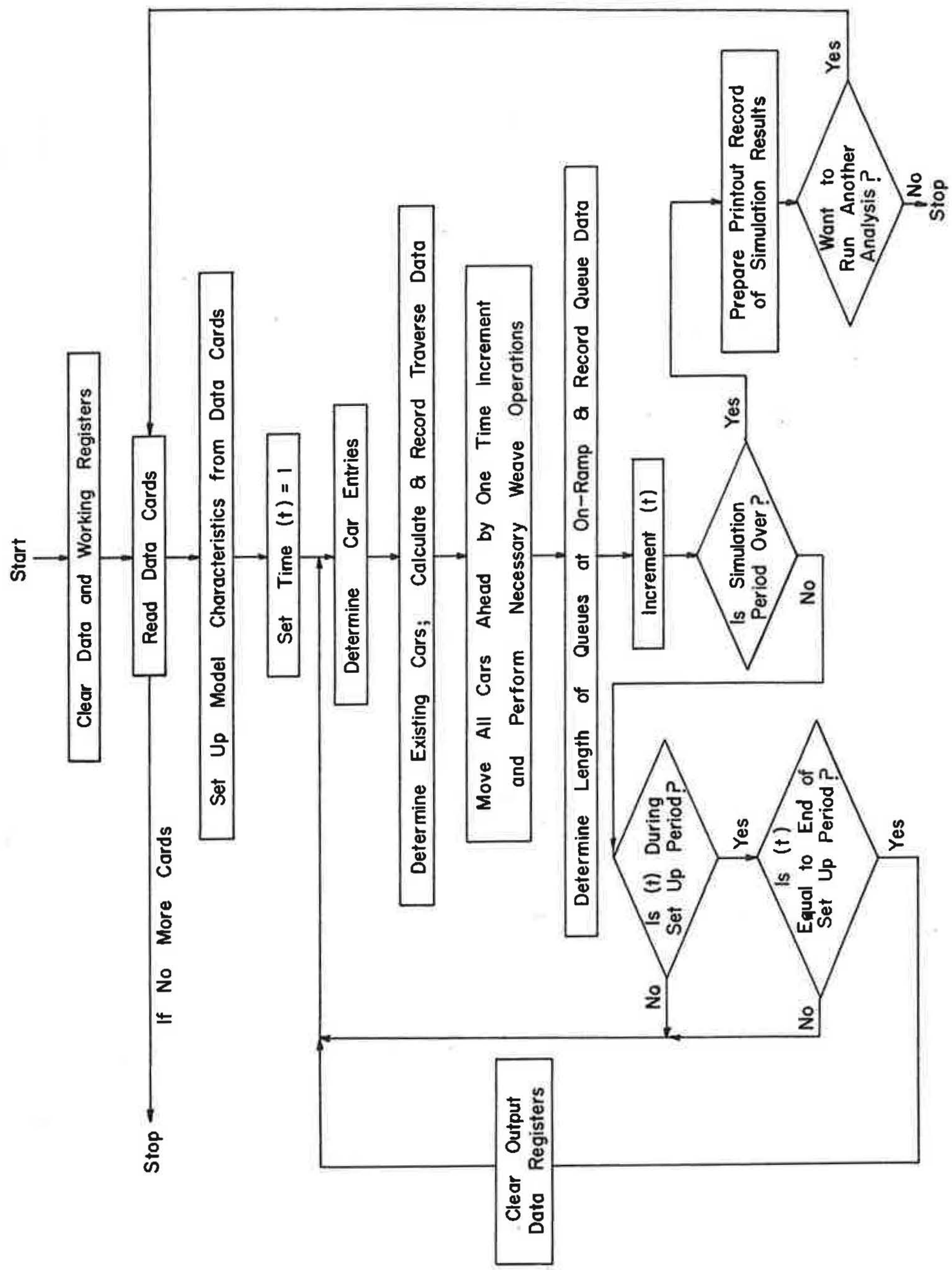


Figure 4.18. Flow diagram of over-all computer logic.

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 nth 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).