A Digital Simulation Program of a Section of Freeway With Entrance and Exit Ramps

JOHANN H. BUHR, THOMAS C. MESEROLE, and DONALD R. DREW
Texas Transportation Institute, Texas A&M University

This paper describes a computer program developed for the simulation of a section of freeway, including several exit and entrance ramps. The program allows for the simulation of the traffic operation under different modes of entrance ramp control: fixed-rate metering, demand-capacity metering, gap-acceptance control and, of course, no control. The computer logic and simulation technique are discussed in detail. Limited output of the program is presented as evidence of the feasibility and realism of the simulation model.

•A penny is tossed until it comes down heads. If this happens on the first toss the player receives one dollar from the bank. If heads appears for the first time on the second toss, the player receives two dollars; on the third toss four dollars, etc., doubling each time. What should the player pay the bank for the privilege of playing this game if the game is fair?

If one has a coin handy, the simplest way to get some insight into how much a player should pay is to play, say, a thousand games and determine the average winnings per game. This seemingly unscientific approach to the Gambler's Ruin or St. Petersburg Paradox, as it is also called, represents a simple illustration of simulation.

SIMULATION

Simulation is essentially a working analogy. It involves the construction of a working model presenting similarity of properties or relationships to the real problem under study. Simulation is a technique which permits the study of a complex traffic system in the laboratory rather than in the field.

In a more general sense, simulation may be defined as a dynamic representation of some part of the real world achieved by building a computer model and moving it through time. The term, computer model, is used to denote a special kind of formal mathematical model, namely a model which is not intended to be solved analytically but rather to be simulated on an electronic computer. Thus, simulation consists of using a digital or analog computer to trace the time paths; the distinction being that the digital device counts and the analog device measures. This distinction is actually a fundamental one, being essentially the mathematical distinction between the discrete variable (digital) and the continuous variable (analog). The differences in capabilities between the digital and analog computers are manifest in the mathematical distinctions between summation and integration, or between difference equations and differential equations.

Why Simulate?

Simulation is resorted to when the system under consideration cannot be analyzed using direct or formal analytical methods. There are a few additional reasons for simulation, most of which have been found to pertain to the simulation of traffic systems.

Paper sponsored by Committee on Characteristics of Traffic Flow and presented at the 47th Annual Meeting.

1. The task of laying out and operating a simulation is a good way to gather pertinent data systematically. It makes for a broad education in traffic characteristics and operation.

2. Simulation gives an intuitive feel for the traffic system being studied, and is therefore instructive.

3. Simulation of complex traffic operations may provide an indication of which variables are important and how they relate since control can be exercised over the various input parameters. This may lead to eventual successful analytic formulations.

4. In some problems, information on the probability distribution of the outcome of a process is desired, rather than only means and variances such as obtained in queueing. Where traffic interaction is involved, the Monte Carlo technique is about the only tool which can give the complete distribution.

5. A simulation can be performed to check an uncertain analytic solution.

6. Simulation is cheaper than many forms of experiment. Imagine the cost savings in simulating to find the optimum spacing of freeway interchanges.

7. Simulation gives a control over time. Real time can be compressed so that the results of a long time period can be observed in a few minutes of computer time. On the other hand, real time can be expanded and run slower than real time so that all the manifestations of the complex interactions of freeway movement can be comprehended.

8. Simulation is safe. It provides a means for studying the effect of traffic control measures on existing highways. The effect of signals, speed limits, signs, and access control all can be studied in detail without confusing or alarming drivers. Simulation offers the ability to determine in advance the effect of increased traffic flow on existing facilities. Probable congestion points and accident locations can be anticipated and changes in the physical design of the highway can be effected before the need is demonstrated through accident and congestion experience.

As a form of model, a simulation model should be compared with analysis on the one hand which involves the use of analytical, rigid, and probabilistic models and trial and error on the other, which involves devising some kind of trial solution and then taking it into actual traffic and trying it out. The relative merits of analysis, simulation, and trial and error, can best be discussed with the aid of Table 1, prepared by Goode (1).

The traffic problem has of course been attacked in the past with the tools of both analysis and trial. Simulation is actually a combination of both methods, but unlike analysis, it allows attack on the most complicated of processes. On the other hand, it does not affect traffic until the solution has been reached. Simulation is almost always midway between analysis and trial (Table 1). But as the situation being studied becomes more complex, such as in the case of traffic systems the differences between methods in terms of cost, time, etc., become more pronounced, until finally neither of the extremes can be tolerated and simulation becomes the only feasible method.

Simulation is a powerful tool and like all powerful tools it can be dangerous in the wrong hands. The increased emphasis on simulation studies and the corresponding lack of experience on the part of some people who attempt to apply the method can lead to a sort of pseudo-simulation. Pitfalls exist in simulation as in every human attempt to abstract and idealize. Some rules to follow in avoiding these pitfalls are (a) no assumption should be made before its effects are clearly defined, (b) no variables should be combined into a working system unless each one is properly explained and its relationships to the other variables are set and understood, and (c) simplification is desirable, but oversimplification can destroy the realism of the model.

TABLE 1

RELATIVE MERITS

Criterion	Analysis	Simulation	Trial
Cost	least	medium	most
Time	least	medium	most
Reproducibility	most	medium	least
Realism	least	medium	most
Generality of results	most	medium	least

For the most part, it can be said that the goals achievable by simulation in the traffic process are clear-cut and offer a profound payoff. Simulation is an ideal technique for traffic research. The simulation model is not just another means for accomplishing

what we can do today but is a tool for solving problems which cannot be solved today.

The Freeway Merging Process

One traffic problem that has generally defied analysis is the process of entrance ramp vehicles merging into a freeway stream. In the summer of 1965, the Bureau of Public Roads undertook research to furnish detailed criteria on the merging of ramp vehicles into the freeway stream. To this end, a contract, "Gap Acceptance and Traffic Interaction in the Freeway Merging Process," was awarded the Texas Transportation Institute. The general aim was the conception of relationships between the many variables associated with the interaction of vehicles traversing a ramp and merging onto a freeway so as to determine the effect of traffic and geometric characteristics on merging operation and level of service.

One objective of this project was to develop a computer program to simulate the operation in a freeway merging area. The purpose of the simulation study was primarily to show the feasibility of simulation as a tool in the study of freeway control for the eventual development of simulation programs to study optimum ramp metering and control techniques and equipment, rather than a reproduction of the minute details of vehicular behavior.

LIST OF SYMBOLS

$D_{o,i}$	Distance from the zero reference point of vehicle i
$D_{t,i}$	Tentative distance from the zero reference point of vehicle i
$D_{f,ij}$	Following distance between vehicles i and j
$D_{d,i}$	Desired following distance factor of vehicle i
$D_{a,ij}$	Acceptable minimum following distance between vehicles i and j
$D_{u,ij}$	Tentative updated following distance between vehicles i and j at the present time
$D_{t,ij}$	Tentative updated following distance between vehicles i and j during the next update
$S_{c,i}$	Current speed of vehicle i
$S_{d,i}$	Desired speed of vehicle i
$S_{w,i}$	Lane changing speed of vehicle i
$A_{s,i}$	Smoothing acceleration of vehicle i
$A_{x,i}$	Maximum acceleration of vehicle i
$A_{b,i}$	Normal acceleration of vehicle i
$B_{s,i}$	Smoothing deceleration of vehicle i
$B_{x,i}$	Maximum deceleration of vehicle i
$B_{b,i}$	Normal deceleration of vehicle i
$B_{n,i}$	Minimum deceleration of vehicle i
$L_{c,i}$	Calculated lag time for vehicle i
$L_{a,i}$	Minimum acceptable lag time for vehicle i
$V_{s,i}$	Length of vehicle i
T	Time increment between updates

THE SIMULATION PROGRAM

The Basic Model

The program allows for the simulation of a freeway segment with a number of entrance and/or exit ramps, having a geometric configuration that can be changed at will. The number of exit ramps is limited to 2, whereas the number of entrance ramps plus freeway lanes is limited to 6. Thus, for example, if the freeway segment to be simulated has 3 lanes, then the number of entrance ramps is limited to 3.

The program initially places a number of vehicles on the freeway segment and then proceeds to process each vehicle in the system according to the programmed flow logic, starting with the vehicle most distant from the beginning of the simulated section, regardless of the lane it is in. During simulation, new vehicles are generated on the freeway lanes and entrance ramps according to a Poisson distribution of arrivals.

Each vehicle is assigned a number of characteristics such as length, current speed, desired speed, distance from the zero reference point (beginning of the simulated section), etc., and the program, using a periodic scan technique, updates these characteristics for each vehicle during the scan interval. The program therefore has available, at any time, all the information concerning each vehicle in the system. The general logic organization is shown in Figure 1. Essentially, each vehicle attempts to travel at its desired speed, subject to certain restrictions. If the vehicle being processed is traveling at its desired speed and a safe headway exists between it and the vehicle ahead, it is simply allowed to proceed down the freeway. If it encounters a slower vehicle ahead, it will, under certain conditions of following distance, attempt to change lanes, first to the left and then to the right. If unsuccessful, the vehicle will decelerate.

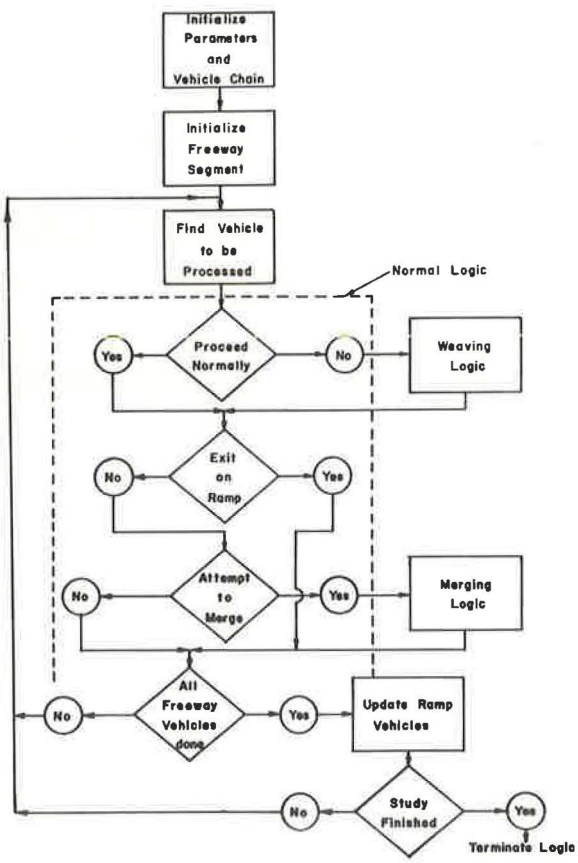

Figure 1. Logic flow diagram of the system.

The program was written in the Fortran IV computer language and is generally run on an IBM 7094 Model I computer. However, it can be run on other computers with a Fortran IV compiler. The simulation program consists of 1 monitor routine and 16 subroutines. Each subroutine is completely modular so that any logic changes in any subroutine will not affect the remainder of the program.

The program also provides the option of simulating the merging operation under various modes of entrance ramp control: (a) no control, (b) fixed-time metering, (c) demand-capacity metering, and (d) gap-acceptance control.

Input Parameters

Inherent in the formulation of a simulation model is the determination of the significant input and output variables. Inputs may generally be divided into four categories: geometrics, traffic characteristics, driver policy and vehicle performance. A fifth category of input serves to control the program directly.

The inputs to this simulation program are listed below by category. Each input parameter has a built-in default value which is used if no other value is specified. The default values are shown in parentheses behind each input parameter.

(1) Geometric Characteristics
- (a) Length of freeway section (6000 ft)
- (b) Number of through lanes (3)
- (c) Number of entrance ramps (1)
- (d) Number of exit ramps (1)
- (e) Location of entrance ramps (4000 ft)
- (f) Location of exit ramps (1000 ft)
- (g) Length of each on-ramp (500 ft)
- (h) Length of each acceleration lane (300 ft)
- (i) Location of ramp signal
- (j) Grade (0)
- (k) Start of grade (4000 ft)
- (l) End of grade (5000 ft)

(2) Traffic Characteristics
- (a) Lane volumes in vehicles per hour*

  Default: Lane 1 = 1430
  Lane 2 = 1730
  Lane 3 = 1840
  Lane 4 = 600
  Lane 5 = 0
  Lane 6 = 0
- (b) Proportion of vehicles on each lane that exit on first ramp

  Default: Lane 1 = .30
  Lane 2 = .10
  Lane 3 = .05
  Lane 4 = 0
  Lane 5 = 0
  Lane 6 = 0
- (c) Proportion of vehicles on each lane that exit on second exit ramp

  Default: Lane 1 = .80
  Lane 2 = .15
  Lane 3 = .05
  Lane 4 = .20
  Lane 5 = 0
  Lane 6 = 0
- (d) Proportion of commercial vehicles in each lane

  Default: Lane 1 = .06
  Lane 2 = .03
  Lane 3 = .01
  Lane 4 = .03
  Lane 5 = 0
  Lane 6 = 0

(3) Driver Policy
- (a) Average acceleration (3 ft/sec^2)
- (b) Average deceleration (6 ft/sec^2)
- (c) Minimum deceleration (2 ft/sec^2)
- (d) Maximum speed (60 mph)
- (e) Crowding factor (0.7)
- (f) Gap acceptance characteristics, specified by two probit equation coefficients (0.5, 2.0)
- (g) Location of exit decision stations (710 and 0)

*The freeway shoulder lane is designated "lane 1." The adjacent freeway lanes and then the entrance ramps are numbered consecutively.

(4) Vehicle Performance
  (a) Maximum acceleration (11 ft/sec^2)
  (b) Maximum deceleration (15 ft/sec^2)

(5) Program Control
  (a) Scan interval (1 sec)
  (b) Analysis time (4 min)
  (c) Warm-up time (1 min)
  (d) Length of warm-up section (500 ft)
  (e) Print option (no)
  (f) Plot option (no)
  (g) Number of plot stations (9)
  (h) Size of plot increments (200 ft)
  (i) Location of last plot station (4300 ft)
  (j) Number of check stations (12)
  (k) Location of each check station

Most of the input parameters given above are self-explanatory. The purpose and use of those that are not immediately evident, will become clear later.

Internal Bookkeeping

The internal bookkeeping procedure used in representing the flow of vehicles within the computer and in keeping track of the characteristics of each unit is perhaps the most complex aspect of simulation requiring the highest degree of programming skill. It is of vital importance to the efficient and successful operation of the program. To implement any practical simulation program, the method of bookkeeping must keep the core storage requirements at a minimum and lend itself to fast sequential processing.

In this simulation program, each vehicle is assigned a subscript number between 1 and 500, allowing no other vehicle to have the same subscript. Each vehicle characteristic is stored in an array so that any particular characteristic, such as current speed for example, of any particular vehicle, can be found by addressing the appropriate array with the subscript number assigned to the vehicle of which the characteristic is desired. If the simulation is allowed to operate over a long period of time, a considerable quantity of data is collected. To conserve storage in the computer, only the characteristics of the vehicles presently in the system are stored in memory. The characteristics of a vehicle which has passed the end of the study section are no longer stored. This is accomplished by reassigning the subscript number of the vehicle using a chaining technique.

The chaining technique is a method of logically organizing the relative position of a vehicle to all other vehicles in the system. There is assigned to each vehicle a value in each of two characteristic arrays which contain the subscript numbers of the vehicles immediately ahead of and behind it and in line with its movement. The arrays of these characteristics are named LAST and NEXT, respectively. This gives the processing program access to the characteristics of the vehicles between which a vehicle being processed is situated and allows determination of those characteristics of the vehicles which must be changed.

TABLE 2

ARRAY ORGANIZATION OF A CHAIN

Subscript Number	Last	Next
2	1	4
4	2	8
8	4	10
10	8	23
23	10	49
49	23	75
75	49	76

Since each vehicle has characteristics in arrays LAST and NEXT, the processing program has an overall picture of the placement of vehicles in the system. These two arrays can be thought of as being a chain with each link being one subscript of a vehicle. For example a chain array might appear as in Table 2.

A chain from the arrays NEXT and LAST containing the subscripts of vehicles shows the organization of one lane of traffic. From Table 2, the lane represented by this chain

would physically consist of vehicle 1, followed consecutively by vehicles 2, 4, 8, 10, 23, 49, 75, and 76.

At the zero reference point a location called the generation pool is constructed for each lane of the system, always containing one vehicle already assigned its characteristics. When the logic determines that a vehicle will enter the freeway, a new vehicle and its characteristics are generated. The vehicle in the generation pool of that lane enters the freeway, and the newly generated vehicle is placed into the generation pool and remains there until another vehicle is generated. This method of always having one vehicle in each generation pool allows the chain associated with each lane to have an end link at the zero reference point, facilitating processing.

The vehicle most distant from the zero reference point in each lane has a value in the LAST array as does every other vehicle in the system. Since this vehicle does not have a vehicle between it and the end of the freeway segment, a fictitious vehicle is placed outside the freeway segment to represent the vehicle ahead of the most distant vehicles. Only one fictitious vehicle is generated for all lanes and is assigned a complete set of characteristics and a unique subscript number. This allows all chains to begin with the same unique vehicle logically terminating the chains. As a vehicle passes through the freeway segment its characteristics are stored, but once it passes off the end of the freeway segment its characteristics are no longer saved and its subscript number is removed from the chain of that lane. The vehicle directly behind the vehicle just processed off in the same lane considers the fictitious vehicle as the vehicle now ahead of it.

As a vehicle is introduced into a generation pool, it is assigned a subscript number. To prevent it from accidentally acquiring the subscript of some vehicle already in the system, the new vehicle obtains its subscript number from a free links pool. Initially one chain is formed which contains all the subscript numbers available for the system linked together using the NEXT and LAST arrays. When the system is initially set up, the subscript numbers not assigned to vehicles placed on the freeway segment, form the free links pool. This pool subsequently contains every available subscript number not being used in the system. As a vehicle passes off the end of the freeway segment its subscript number is placed in the free links pool, and when a vehicle enters a generation pool it obtains its subscript number from the free links pool.

The purpose of the chaining logic is to permit an organized handling of vehicle characteristics while using a method which is easily adapted for digital computation. In the digital computer this process reduces the number of internal storage locations required. If this process were not used and specific locations were required for the characteristics of each vehicle, both in and out of the system, the number of vehicles that can pass through the system would be quite limited. Using the chaining technique, the restriction imposed on the system is that not more than 500 vehicles be on the freeway segment at any one time.

Initial Setup

The first step after reading in all the input parameters, is the geometric arrangement of the simulated section in the computer and the placement of vehicles on the freeway lanes and ramps.

Initially, one vehicle is placed in each freeway lane at the end of the study section and one vehicle on each entrance ramp, 10 ft upstream of the ramp nose. Each of these vehicles, as well as subsequent vehicles, is assigned a desired speed and its current speed is set equal to its desired speed. These speeds are pseudorandom numbers generated so as to be normally distributed with parameters:

$$\text{Mean} = 0.85 \times \text{maximum freeway speed}$$
$$\text{Standard deviation} = 0.07 \times \text{mean} + 1.375$$

Any speed generated to be higher than the maximum freeway speed, is set equal to the maximum freeway speed.

Based on their assigned speeds, vehicles are then placed at a distance behind the previous vehicle so as to correspond to an exponential speed-density curve. This relationship, first proposed by Greenberg (2), is given by

$$u = c \ln (k_j/k)$$

where

u = speed,
k = density,
k_j = jam density = 175 veh/mile, and
c = constant of proportionality = 35 mph.

Each vehicle is further assigned a "desired following distance factor" which is simply a random number, uniformly distributed between 0.2 and 1.0 and a length of either 16 or 32 ft based on whether it is a passenger car or truck. The decision as to the type of vehicle is made by comparing a uniformly distributed random number to the proportion of commercial vehicles specified as an input parameter. Similarly, each vehicle is tagged as to whether it desires to exit or not, and if so, which ramp it will exit on.

During simulation, new vehicles are generated from a Poisson distribution of arrivals by comparing a uniformly distributed random number to the volume in vehicles per scan interval, the volume in each lane being an input parameter. Vehicles so generated are assigned a current speed, which equals its desired speed, and a length, in the same manner as during initial setup, but it is placed at the beginning of the simulated section.

The program runs for a period, specified as the "warm up time" input parameter, before any vehicle data are collected and stored.

Simulation Logic

The simulation logic for stepping vehicles through the system, once the inputs are known, may be divided into three classifications: (a) flow logic for unimpeded vehicles, (b) car-following logic for platooned vehicles, and (c) maneuvering logic for vehicles executing maneuvers involving more than a single stream of traffic. In reality, all drivers of vehicles within the roadway system are continually and simultaneously making decisions and modifying their behavior. In the course of a simulation, the classification of most vehicles—unimpeded, following or maneuvering—will change many times. The computer, however, can make only one simple logical choice at a time. To control all the occurrences at any given instant, it must process all decisions sequentially. In other words, it must process each decision for every vehicle, for each vehicle in every lane, and for each lane within the system. It must do this in accordance with a prescribed sequence for each instant of time to be considered.

The program is organized into independent logical divisions with one monitor division to direct the control among the other divisions. By using separate divisions, experimentation can be carried out in one division without changing any other logic.

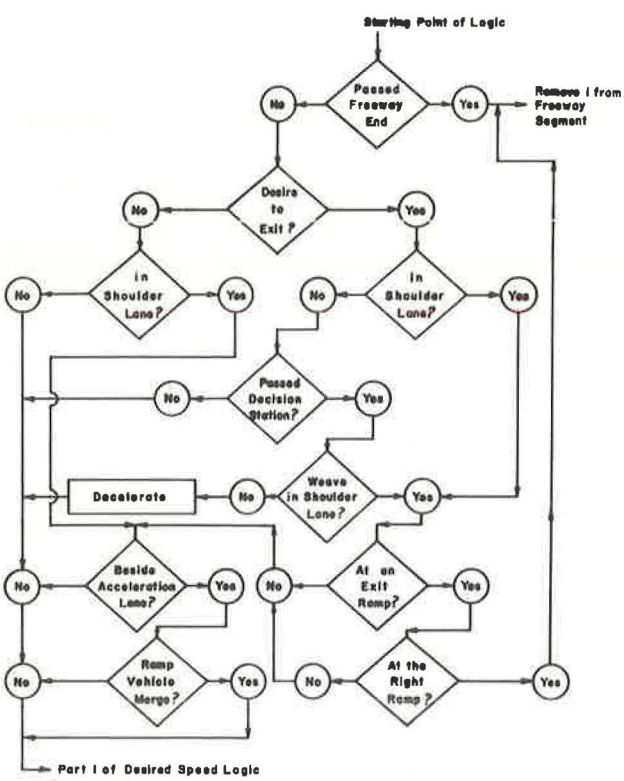

Figure 2. Distance logic flow diagram.

The monitor division incorporates the tasks of initializing all parameters including the chains, initializing the freeway segment for the first time period, and handling the normal or through flow of traffic. This section determines which of the vehicles on the freeway segment is the most distant nonprocessed vehicle and perform a series of tests on it. The tests are in the form of the following questions.

1. Will this vehicle travel past the end of the freeway segment during the next scan interval?
2. Does this vehicle exit on a ramp?
3. Does a vehicle merge in front of this vehicle from the acceleration lane of an entrance ramp?
4. If this vehicle is not in the shoulder lane and desires to exit soon, can it weave into the shoulder lane?
5. Is this vehicle traveling as fast as it desires?

Once the monitor division has operated on all freeway vehicles and has updated them, the ramp vehicles are updated.

Normal Flow—The normal flow section determines the behavior of vehicles in the same lane and is divided into two segments: the distance logic and the speed logic. A flow diagram of the distance logic is shown in Figure 2.

The distance logic stems from the fact that certain points on the freeway segment are of particular interest in the processing of vehicles. These are the end of the freeway segment, the beginning of the exit ramps, the exit decision stations, and the area next to the acceleration lanes. The number of these points and their distances from the zero reference point are input parameters which can be varied to give the scheme flexibility. This section determines whether a vehicle has passed one of the points during a given time period by calculating a tentative distance $D_{t,i}$.

$$D_{t,i} = D_{o,i} + (S_{c,i})\,T$$

$D_{t,i}$ represents the position of vehicle i if no speed change occurs. If this tentative distance is greater than the distance to one of the significant points, and the original distance $D_{o,i}$ of vehicle i is less, vehicle i will pass this point for the first time. Depending on which point it is, different algorithms are employed.

Vehicle i has passed the end of the freeway segment if its $D_{t,i}$ is greater than the distance to the end of the freeway segment. A sequence is then initiated to remove i from the chain of that lane and place its subscript into the free links pool. The exit decision stations are located upstream of the exit ramps so as to give a vehicle desiring to exit time to weave into the shoulder lane. After an exiting vehicle has passed such a station, it will attempt to weave until it is successful, even after it has passed the desired ramp. When a vehicle is in the shoulder lane and it desires to exit, its $D_{t,i}$ and $D_{o,i}$ are compared with the distance to each exit ramp on the freeway segment to determine if this vehicle is now at its desired exit ramp, and if so, it is processed out of the system.

The distance logic is based on the assumption that there is no change in the speed of a vehicle during a time interval. This is, of course, not necessarily true. After the position of a vehicle on the freeway segment has been compared with all the important points and this vehicle does not leave the freeway, control is transferred to the first part of the three-part speed logic.

The desired speed logic is broken down into three parts based on the relationship between the vehicle being processed and the vehicle ahead of it in the same lane. A following distance and an acceptable following distance are calculated for a pair of vehicles, i and j, where i is the vehicle being processed and j is its lead vehicle. The following distances are given by

$$D_{f,ij} = D_{o,j} - D_{o,i} - V_{s,j}$$

$$D_{a,ij} = V_{s,i} + S_{c,i} + \left[\frac{0.5\,(S_{c,i} - S_{c,j})^2}{B_{b,i}} + D_{d,i}\,(S_{c,i})\right] K$$

where

$K = 0$ if $S_{c,i} \leq S_{c,i}$, and
$K =$ crowding factor if $S_{c,i} > S_{c,j}$.

The factor, K, allows for a smaller acceptable following distance between two vehicles if the lead vehicle is traveling faster than the following vehicle.

The three parts of the speed logic are shown by the flow charts in Figures 3, 4, and 5. The first part of the speed logic deals with the case where the following distance is less than acceptable, the second part with the case where the following distance is exactly equal to the acceptable following distance and the third part with the case where the following distance is bigger than the acceptable following distance.

If the following distance is less than acceptable (Fig. 3) but i is traveling slower than j, then the logic is terminated by simply updating vehicle i at its current speed. However, if i is faster than j it will decelerate by either the normal deceleration, $B_{b,j}$, or the maximum deceleration, $B_{x,i}$, depending on how close i is to j.

If the following distance between i and j is equal to the acceptable following distance (Fig. 4) and i is traveling slower than j, then, if maximum acceleration would put it above its desired speed, it simply accelerates to its desired speed and the logic is terminated. Otherwise, it accelerates by its maximum acceleration and then attempts to change lanes. On the other hand, if i is traveling faster than j, then it either decelerates by the maximum amount or it decelerates to the same speed as j. If, in the latter case, its speed is now less than desirable, it will attempt to change lanes, otherwise, it will decelerate further to its desired speed.

The last part of the speed logic (Fig. 5) deals with the case where the following distance is greater than the minimum acceptable following distance. If, under this condition, i is traveling slower than j, it is accelerated by an amount $A_{s,i}$ given by

$$A_{s,i} = A_{x,i}\,(S_{d,i}/S_{c,i}) - A_{x,i}$$

Therefore, if it is already traveling at its desired speed, its speed will not change, but otherwise, it will adjust its speed so as to approach its desired speed asymptotically.

However, if under these conditions, vehicle i is traveling faster than vehicle j, a tentative following distance is calculated to inspect what will happen if i does not change its speed. The tentative following distance is given by

$$D_{u,ij} = D_{o,j} - D_{o,i} - (S_{c,i})T - V_{s,j}$$

If this tentative following distance is less than or equal to the minimum acceptable following distance, vehicle i decelerates by the normal deceleration $B_{b,i}$. Otherwise, the distance between i and j will remain safe after this scan interval and the vehicle is processed

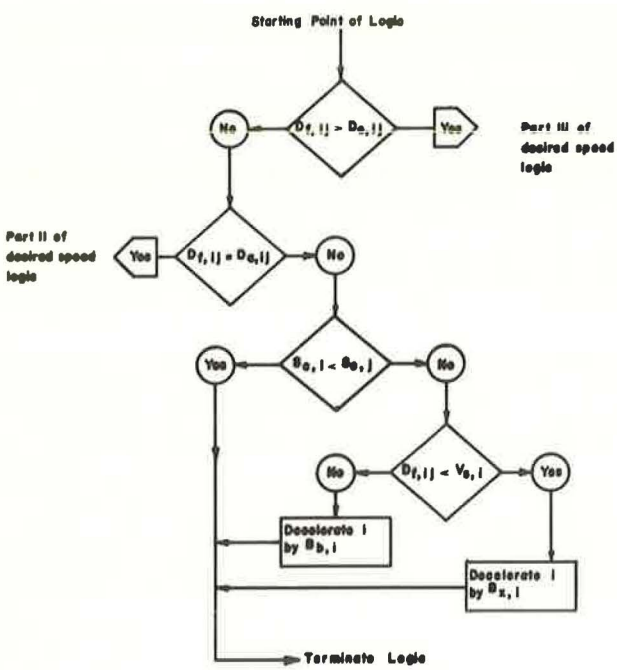

Figure 3. Speed logic flow diagram—Part I.

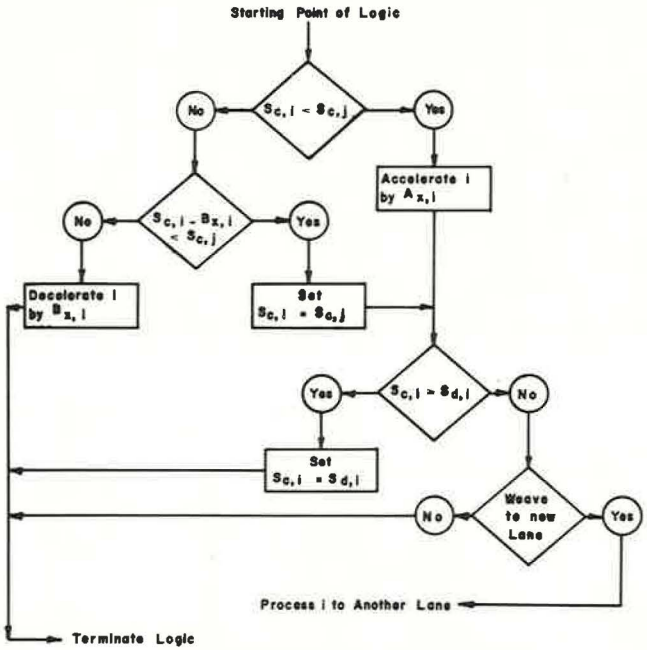

Figure 4. Speed logic flow diagram—Part II.

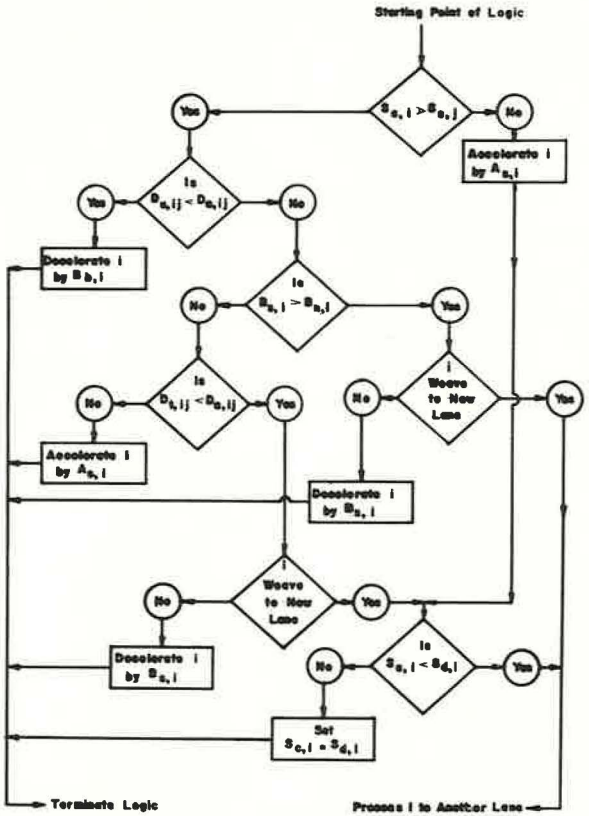

Figure 5. Speed logic flow diagram—Part III.

so as to eliminate erratic movement, by using a smoothing deceleration as expressed by

$$B_{s,i} = \frac{(S_{c,j} - S_{c,i})^2}{2(D_{u,ij} - D_{a,ij})}$$

This gives the deceleration that i must take for it to be moving at the same velocity as j when the following distance between them equals the acceptable following distance. If this smoothing deceleration is greater than the minimum deceleration $B_{n,i}$, vehicle i will try to change lanes. If vehicle i cannot change lanes, it is decelerated by $B_{s,i}$. If, on the other hand, $B_{s,i}$ is less than $B_{n,i}$, the possibility of accelerating vehicle i is explored and a new tentative following distance, being the projected following distance two scan intervals hence, is calculated, assuming $S_{c,i}$ and $S_{c,j}$ constant during the following time period.

$$D_{t,ij} = D_{o,j} + (S_{c,j})T - D_{o,i} - 2(S_{c,i})T - V_{s,j}$$

If this projected following distance is less than the minimum acceptable following distance, i will either decelerate by the amount $B_{s,i}$ or will change to another lane, if its speed is less than desirable or will increase its speed to its desired speed. On the other hand, if the projected following distance is greater than the minimum acceptable, vehicle i will accelerate by $A_{s,i}$.

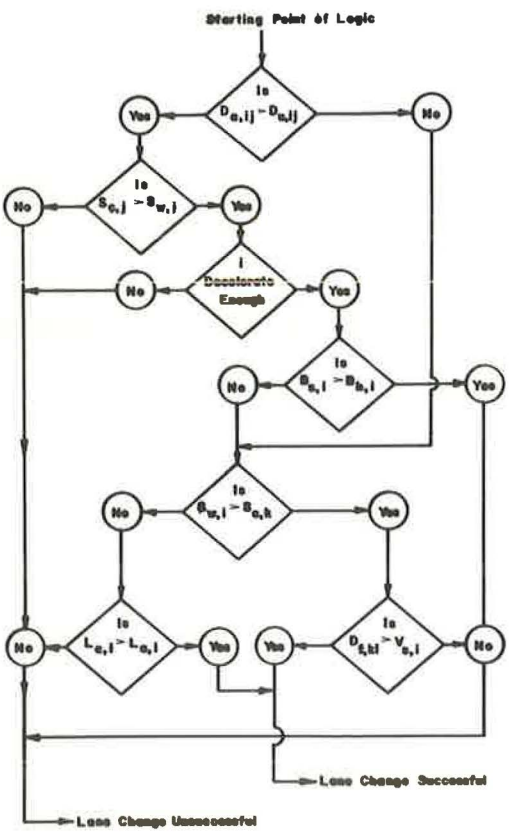

Figure 6. Lane change logic flow diagram.

Lane Changing—When it is determined that a vehicle desires to change to another lane, the possibility of changing to the left is explored and then to the right. If a vehicle is on either the shoulder lane or the median lane, only one direction is tried. If all lane change attempts are unsuccessful, the vehicle remains in its present lane. If successful, the vehicle is removed from its old lane to the new lane by rearranging the appropriate chains. After a lane has been chosen into which an attempted entry will be made, the changing vehicle is accelerated by the maximum amount $A_{x,i}$. This new speed is called the lane changing speed and is designated $S_{w,i}$. The vehicle will change lanes if it will not come hazardously close to the vehicles ahead of or behind it in the new lane. If the gap is inadequate the attempt is unsuccessful. The lane changing logic is diagrammed in Figure 6, assuming that vehicle i attempts to change lanes into a space between vehicles j and k.

First, the relationship between the lane changing vehicle i and the lead vehicle, j, in the new lane is investigated by calculating a tentative following distance $D_{u,ij}$ and a minimum acceptable following distance $D_{a,ij}$.

$$D_{u,ij} = D_{o,j} - D_{o,i} - V_{s,j} - (S_{w,i})T$$

$$D_{a,ij} = V_{s,i} + [S_{w,i} + D_{d,i}(S_{w,i})]K$$

where

$K = 0$ if $S_{w,i} \leq S_{c,j}$, and
$K =$ crowding factor if $S_{w,i} > S_{c,j}$.

If the following distance is bigger than the minimum acceptable following distance, the relationship between the lane changing vehicle and the following vehicle, k, in the new lane will be investigated. However, if the following distance is less than acceptable then the relative speed between i and j is investigated. If i is traveling faster than j, the lane change attempt is unsuccessful but if i is slower, a test is made to see if i can slow down so that a safe distance will exist between it and vehicle j, if i uses its maximum deceleration. If i cannot decelerate fast enough, the lane change attempt is unsuccessful, otherwise, a smoothing deceleration, $B_{s,i}$, is calculated to see if i is closing in on j too fast.

$$B_{s,i} = \frac{(S_{c,j} - S_{w,i})^2}{2(D_{o,j} - D_{o,i} - (S_{w,i})\,T)}$$

If this deceleration that would be required is greater than the normal deceleration, vehicle i does not change lanes, otherwise, it is assumed that the relationship between i and j is safe and the relationship between i and k is investigated.

If vehicle i is traveling faster than vehicle k and there is at least a vehicle length between them, a lane change occurs. Otherwise, the attempt is unsuccessful.

If vehicle i is slower than k, the time relationship between vehicles i and k is investigated. First, a minimum acceptable lag time, $L_{a,i}$, is generated as a random variable, normally distributed with a mean of 0.5 sec and a standard deviation of 0.1 sec. However, if the generated variable is less than 0.1 sec, it is set equal to 0.5 sec. This minimum acceptable lag time, is compared to a lag time, $L_{c,i}$, defined as the time required by vehicle k to come within the minimum safe distance of i and is given by

$$L_{c,i} = \frac{(D_{o,i} - D_{o,k} - V_{s,i} - D_{a,ik})}{(S_{w,i} - S_{c,k})}$$

If this lag time is less than acceptable, the lane change does not occur. Otherwise, vehicle i changes lanes and the chains are appropriately rearranged.

Ramp Discipline—The behavior of vehicles on an entrance ramp is treated in two parts: the movement of vehicles on the ramp and the merging of vehicles onto the freeway segment from the ramp.

The movement of vehicles on a ramp is handled similarly to the vehicles on the freeway segment, except that ramp vehicles cannot change lanes while on the ramp. The ramp vehicle closest to the end of the ramp is processed as a special case because this vehicle must stop at the end if it is unable to merge. This vehicle is gradually decelerated as it approaches the end of the acceleration lane. When it reaches the end, it stops. The remaining ramp vehicles are processed similarly to the freeway vehicles by calculating an acceptable following distance, given by

$$D_{a,ij} = V_{s,i}/2 + (S_{c,i}/2)\,K$$

where

$K = 0$ if $S_{c,i} \leq A_{b,i}\,T$, and
$K = 1$ if $S_{c,i} > A_{b,i}\,T$.

This acceptable following distance is shorter than those of through vehicles since ramp vehicles generally crowd closer together when attempting to merge. The procedure determines whether a vehicle will accelerate or decelerate by comparing $D_{a,ij}$ with its following distance during the last time period, the present time period, and the next

time period. Again smooth processing is the key criterion and when a vehicle approaches another vehicle it will do so gradually.

Gap Acceptance—The merging subroutine processes vehicles from the acceleration lane onto the shoulder lane of the freeway segment. Vehicles are processed starting with the one closest to the end of the acceleration lane, each vehicle will attempt to merge into the gap adjacent to its current position. The space relationships between the merging vehicle, the leader and the follower are first investigated and if these are found acceptable, the time gap is compared to an acceptable gap for the merging vehicle to determine if the merge will take place.

The criteria used to determine if the physical limitations will preclude a merge, depends on the relative speeds of the vehicles involved. If the speed of the leader is greater than the speed of the ramp vehicle, the acceptable distance between them is one vehicle length. The acceptable distance is reduced by one-half if the leader merged in the same scan interval. If the speed of the leader is less than or equal to the speed of the ramp vehicle, the acceptable distance between them is one vehicle length for each 10 mph of speed of the ramp vehicle.

The acceptable distance between the ramp vehicle and the follower also depends on the relative speed of the two vehicles. If the ramp vehicle is the faster of the two, the acceptable distance is one vehicle length. Otherwise, the acceptable distance is one vehicle length for each 10-mph ramp speed.

If any of the acceptable distances are less than the actual distances, the ramp vehicle does not merge. For vehicles which are forced to stop at the end of the acceleration lane, an "impatience factor" is introduced. This causes the acceptable following distances to be reduced by 5 percent for each scan interval that it waits, to a minimum of 60 percent of the original acceptable following distances.

After it has been determined that the physical conditions will permit a merge, an acceptable gap is generated for the ramp vehicle. Acceptable gaps are generated as random variables, distributed as a lognormal distribution. This is done by generating a normally distributed pseudorandom number, finding its probit value by means of a table look-up and comparing this probit value to the value given by the probit equation specified by the input parameters, solved for the available freeway gap (3).

The Output

The normal output of a simulation system is a table of average values of quantities such as travel times and volumes processed. This gives an overall view of the operation of a system, but it does not give any information about individual vehicle movement. A number of small mistakes can be present in the logic without being detected. A table of values indicating vehicle movement is cumbersome for a large study and, at best, is difficult to interpret. Part of the output of this simulation is such a table of average values but a more complete picture of the operation was developed as part of this research by displaying the movement of vehicles in a graphical form by means of a time-space diagram.

The abscissa of the time-space diagram represents time, the ordinate represents distance from the zero reference point, and one continuous line represents the movement of an individual vehicle in time and space. The plot therefore gives the position of each vehicle on the freeway segment for the entire period of the study plus a picture of relative speeds and positions. It therefore serves not only to give an overall view of the operation, but is also of tremendous value in debugging the program by displaying the occurrences of logic errors.

Figure 7 is an example of part of the time-space output of the simulation of an uncontrolled merging situation. It shows vehicles in the right-hand freeway lane plotted in continuous lines while vehicles on the ramp are represented by dashed lines. Lines that suddenly terminate, indicate that the vehicle has changed lanes. Different lanes and the ramp movement can be plotted using continuous lines and dashed lines or by using different colored lines to distinguish among them. A plot showing a large number of lanes can become quite cluttered and will lose its effectiveness.

Each curve of a plot is made up of a number of points with a line drawn between them. Each point represents the position and time of a vehicle in the system. Since the digital

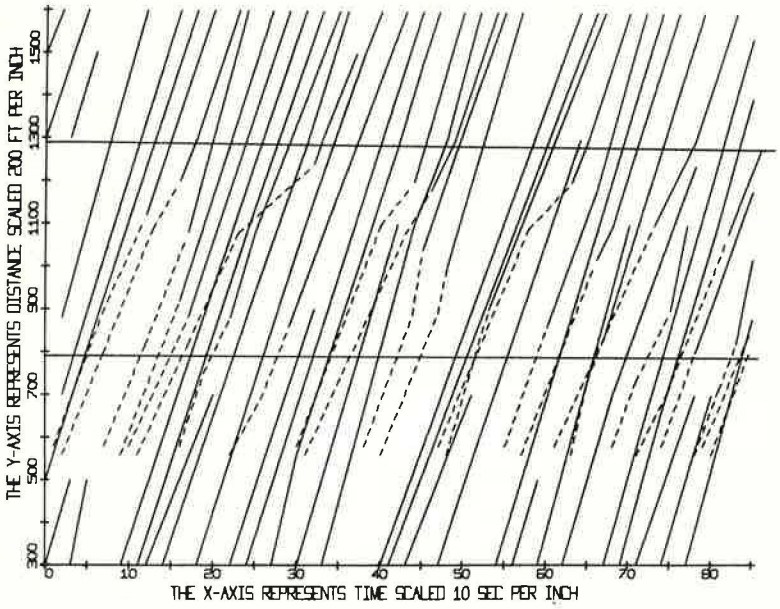

Figure 7. Time-space plot of uncontrolled ramp.

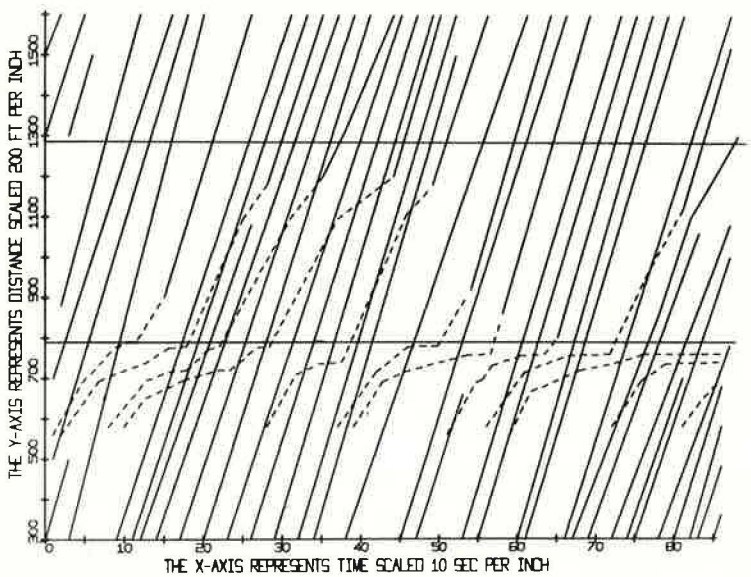

Figure 8. Time-space plot of controlled ramp.

computer program makes a periodic scan, it decides when data are to be saved and writes these items on magnetic tape, one record per block, indicating the name of the vehicle, its position on the freeway segment, and the time. After the study has been completed the data on the tape are sorted and rearranged into the proper form for plotting the time-space diagram.

Figure 8 is an example of part of the time-space output of a ramp under simulated gap-acceptance control. The behavior of vehicles on the freeway and on the ramp as

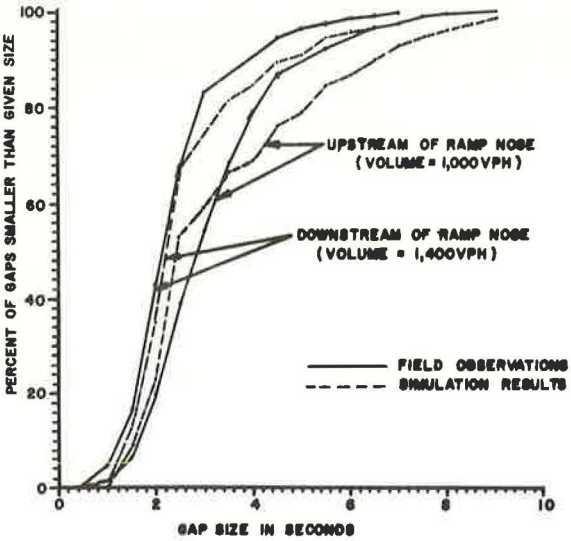

Figure 9. Gap distributions yielded by simulation program compared to field data.

demonstrated by these plots seems to represent real life conditions with a fair degree of realism.

A third type of output provided by this simulation is the time-space characteristics of each vehicle in tabular form, punched into computer cards. This output is in the same form as the input used on an extensive library of analysis programs developed by the Texas Transportation Institute and used for the analysis of the merging process (4).

Model Calibration and Validation

If programmed properly, the realism of the computer output is a function only of the realism of the system model and the inputs to the model. A simulation model is essentially a hypothesis and, therefore, must be tested before it can be accepted as fact. Such tests include its feasibility, realism and validity.

The feasibility and realism of this simulation study is illustrated by the time-space diagrams. Little work has been done on the calibration and refinement of the various models and assumptions that constitute the overall simulation program. Such studies are now in progress and indications are that the model can be satisfactorily calibrated.

Although it was not the intent of this simulation study to reproduce all the minute details associated with vehicular behavior, it is essential that the results be compared with known real world responses to the same inputs. This was done by running the simulation program using the geometrics of the outbound Cullen entrance ramp on the Gulf Freeway in Houston. A single 20-min period was simulated, using as input parameters the volume levels and gap acceptance characteristics observed at this ramp over a 20-min study period. Further analysis of the computer output revealed that simulated drivers generally behaved in a more uniform manner than real drivers. They generally maintained higher and more uniform speeds while following each other at greater distances. However, their speed-spacing relationships as revealed by a comparison of time headway distributions showed an encouraging agreement between simulated and actual conditions (Fig. 9). Through the calibration of the various models that constitute the simulation, any particular set of data for any particular ramp can probably be duplicated quite closely. However, the study was designed to simulate the operation in any merging area, with the result that the calibration and validation of this simulation program will involve a large number of runs under greatly varying conditions with associated compromises in adjustments.

Theoretically, the model must duplicate the characteristics which the highway engineer uses as design criteria, or the characteristics that the traffic engineer uses as operational criteria. As is usually the case, however, universally useful design and operational criteria cannot be precisely defined and each design application requires the selection of suitable criteria based on engineering judgment. One might adopt a microscopic philosophy in which attempts would be made to duplicate, in the computer, the specific details of field samples of moderate time length. Otherwise a macroscopic approach might be utilized in which computer runs seek to reproduce gross statistical properties of field samples accumulated over long periods of time. Since traffic is a stochastic process, valid arguments can be raised in support of either approach. In this study, greater emphasis is being put on the macroscopic results. The calibration, validation and further refinement of the model is presently in progress.

ACKNOWLEDGMENT

The simulation program described in this paper was developed under a research contract sponsored by the U.S. Department of Transportation, Federal Highway Administration, Bureau of Public Roads.

REFERENCES

1. Goode, H. H. Computers, Simulation, and Traffic. Unpublished.
2. Greenberg, H. An Analysis of Traffic Flow. Operations Research, Vol. 7, 1959.
3. Drew, D. R., LaMotte, L.R., Buhr, J. H., and Wattleworth, J. A. Gap Acceptance in the Freeway Merging Process. TTI Research Report 430-2, Dec. 1966.
4. Buhr, J. H., Drew, D. R., Wattleworth, J. A., and Williams, T. G. A Nationwide Study of Freeway Merging Operations. TTI Research Report 430-1, Dec. 1966.