URBAN TRAFFIC SYSTEM SIMULATION AND CONTROL

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

WAYNE DAVID PANYAN

1969

This is to certify that the

thesis entitled

URBAN TRAFFIC SYSTEM
SIMULATION AND CONTROL

presented by

Wayne David Panyan

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in E.E.

John B. Kreer
Major professor

Date August 18, 1969

O-169

ABSTRACT

URBAN [illegible] SIMULATION [illegible]

[illegible]

A simulation [illegible] vehicular traffic systems is [illegible] this thesis. The model is applicable [illegible] signal controlled intersections [illegible] described as light to medium. Interpreting [illegible] interconnection of smaller components, [illegible] the same phenomena as the whole, is the key [illegible]. In each of these components the [illegible] of platoons and queues (the smallest vehicular units [illegible]) are described by a set of state equations. The [illegible] equations are position and vehicular density. Only one queue can be present on the component and it can be described wholly by its position. However, since more than one platoon may exist on the component and since each [illegible] two variables in its description, the number of equations [illegible] is a variable, $2 p_i + 1$, where p_i is the number of platoons at any instance.

ABSTRACT

URBAN TRAFFIC SYSTEM SIMULATION AND CONTROL

by Wayne David Panyan

A simulation model for large urban vehicular traffic systems is developed in the first part of this thesis. The model is applicable to systems having signal controlled intersections and vehicular densities described as light to medium. Interpreting the system as an interconnection of smaller components, each exhibiting the same phenomena as the whole, is the keystone of the development. On each of these components the behavior of platoons and queues (the smallest vehicular units considered) are described by a set of state equations. The variables in these equations are position and vehicular density. Only one queue can be present on the component and it can be described wholly by its position. However, since more than one platoon may exist on the component and since each requires two variables in its description, the number of equations required is a variable, $2p_i + 1$, where p_i is the number of platoons at any instance.

A complete simulation model comprises an interconnection of several such components and a set of 2p + q equations, where p and q are the total number of platoons and queues, respectively. The structure of the system is described by a connection matrix which is analogous to the incidence matrix of graph theory. The inclusion of acceleration phenomena, random inputs and turning movements results in a model which is general enough to simulate most traffic structures and behavior. A Fortran program based on the equations was written and used to simulate the traffic behavior of the central business area of Lansing, Michigan. Results of this simulation are included in the thesis as an example.

In the latter parts of the thesis the control problem is considered. If the vehicular densities are sufficiently low, the steady state control of an urban traffic system can be effected by a synchronization of the traffic signals. Such a synchronization, called a progression, allows vehicles to travel the length of an artery without having to stop for a traffic signal. Synchronizing the signals so that progressions are established in the two directions of a two-way street is simple enough in theory. However, certain auxiliary strategies can also be applied to discourage the queuing of vehicles. Further considerations are required when an overall control strategy is to be instituted on a traffic grid.

Wayne David Panyan

The most efficient use of an artery can be achieved when the progression design is selected in an optimal manner. An important innovation is the inclusion of the demands that exist on every part of the artery in a cost function which is proportional to the total vehicle travel time. Minimizing this function while satisfying the physical realizability constraints imposed by the arterial geometry, fixed signal parameters, and upper and lower velocity bounds results in the optimal design. The nonlinearities inherent in the problem and the nonconvexity resulting from the constraints require that an iterative solution technique be used. A Fortran program to obtain this optimal design was written and used to find the design for a typical street. These results are included in the thesis.

URBAN TRAFFIC SYSTEM SIMULATION
AND CONTROL

By

Wayne David Panyan

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering

1969

G61763
4-27-70

ACKNOWLEDGMENTS

The author [illegible] express his gratitude to his major professor, Dr. [illegible], who in the course of many fruitful [illegible] suggested many useful ideas and spurred this [illegible].

The author [illegible] the Division of Engineering Research of [illegible] State University under whose sponsorship this research was done.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to his major professor, Dr. John B. Kreer, who in the course of many fruitful discussions suggested many useful ideas and spurred this work to completion.

The author also wishes to thank the Division of Engineering Research of Michigan State University under whose sponsorship this research was done.

TABLE OF CONTENTS

LIST OF FIGURES

Figure		Page

Figure		Page

CHAPTER I

INTRODUCTION

The problem of urban congestion that accompanied the increased use of the automobile has become so great in recent times that the simple remedies developed in the past no longer are effective. Observing traffic, placing a signal here, and posting a speed limit there are insufficient. Effective utilization of today's traffic systems demands the use of sophisticated traffic controls. Such controls can be developed through modern control techniques. However, a prerequisite is a good model of the traffic system behavior.

In the past ten or fifteen years some efforts have been made to explain traffic flow mechanisms and particular traffic phenomena, but no complete model has been produced. In part this failure resulted from inadequate data. Gathering data is an enormous task because traffic systems are physically large, and to observe the propagation of variables within the system requires many expensive vehicle detectors. More importantly, a traffic system almost defies macroscopic analysis. The behavior it exhibits is the result of disparate phenomena, some of which are little understood in isolation and even less understood within the context of the system. The system is non-linear, not completely predictable, and susceptible to small changes of many factors.

This thesis is concerned with the modeling and control of an urban traffic system having signal controlled intersections.

In Chapter II a model of signal-controlled streets with medium traffic density conditions is developed. These conditions are often encountered during morning and evening rush hours, periods when improved control is definitely needed. After examining a set of possible traffic flow variables, average velocity and density are selected as most appropriate to describe vehicular movements. Since the densities encountered are assumed to be great enough, groups of vehicles, called platoons and queues, are the smallest vehicular units considered. It is reasonable to attempt the study of a complete urban area if platoons and queues are considered, but the problem becomes too complex and inefficient to solve if individual vehicles are considered.

An urban street system can be looked upon as an interconnection of basic components each displaying the characteristics of the whole system. The existence of such components is postulated, and for each component a set of equations is derived which describes the platoon and queue behavior in terms of the density and velocity variables.

The problem of simulation is investigated in Chapter III. It is demonstrated how an interconnection of a number of basic components can be used to simulate a variety of traffic systems. The simulation is achieved by using a Fortran computer program based on the developed equations.

The main objective in controlling an urban system is to minimize the delay that vehicles experience as they travel through the system. Usually this is achieved through the use of progressions. A progression is established by the settings of the traffic signals which comprise the primary control devices. In Chapter IV the problem of establishing progressions is investigated. Special attention is spent on one-way street progressions and on the unique problems presented by grids.

In Chapter V the problem of selecting the optimal progression design for a two-way street having variable demands along its length is studied. The criterion for this design is a minimization of the total vehicle-hours of travel time. Since the travel time is a non-linear function of the arterial geometry and signal settings, an iterative computer solution is required to obtain the optimal settings of the signals.

A simulation of the traffic system of Lansing, Michigan's central business area is given as a demonstration of the versatility of the model developed in Chapter II. A second example to demonstrate the optimal design procedure on a two-way street is also included. Flow charts for the simulation and design programs are given in the Appendix.

CHAPTER II

MODELING

The desire to know how a system behaves under a wide spectrum of conditions when direct experimentation and observation are not possible (for reasons of economy, time, safety, system inaccessibility, inadequate instumentation, etc.) is sufficient motivation for generating a simulation model. If the model accurately describes the phenomena exhibited by the system, it becomes a powerful tool for determining the response of the system to a variety of controls and for investigating the effects of various parameters.

A traffic system belongs to a huge category of systems which are difficult to study by direct means. Physically it is large: in a metropolitan area a traffic system of interest may cover several square miles. To study such a system requires an extensive instrumentation network. The vehicle detectors needed for measuring traffic variables are expensive and usually require costly installation. Unfortunately, they do not always provide the data in the form needed. Furthermore, many traffic variables cannot be adjusted at will. For example, average velocity and density result from the interactions of many vehicles and are not easily controlled. Finally, even though traffic signals are variable control devices, they should not be indiscriminately reset under the guise of scientific research.

The analysis situation is made worse by the fact that an accurate model is difficult to obtain. Certainly if better data were available, present mathematical descriptions of the system behavior would be more accurate. Secondly, within the system each driver, while constrained by the proximity of other vehicles and by legal and physical limits on speed and maneuvering, operates his vehicle according to his own driving habits. As a consequence of this freedom the system is to a greater or lesser extent stochastic in nature.

Furthermore, a traffic system exhibits more than one mode of operation so that under a given set of conditions certain behavioral aspects are dominant and others are minimal.

As a result of the inherent complexity of a traffic system one might rightfully conclude that its mathematical model needs to be extremely complex. If, however, a traffic system can be reduced to its essential characteristics, a tractable model is possible.

In defining a traffic system some of the distinct classifications become evident. A difference exists, for example, between traffic studies on urban streets and on limited-access freeways since the traffic signals used to control the flow of vehicles through intersections force the vehicles into behavioral patterns not observed on freeways. Moreover, traffic flow on surface streets exhibits several modes depending on the vehicular density. In a heavy density mode the queues which occur at each intersection are sufficiently long that they do not clear during a single green phase of the signal. Under these conditions vehicles travel only a short distance before coming to a stop and each must react instantly to the speed reductions of its predecessor to avoid collision.

In a medium density mode each vehicle is still constrained by the action of others, but a group of vehicles, called a platoon, often can travel through more than one intersection before stopping. In a light density mode the speed of an individual vehicle is almost independent of the speeds of other vehicles; queues and platoons, existing as random, transient phenomena, do not comprise a major feature of the flow.

2.1 Previous Modeling Approaches

Over the years many approaches have been taken to describe traffic flow. If it were a simple task, the work of these previous investigators would have included a complete simulation model for traffic systems. However, their efforts have concentrated on very specialized urban traffic problems and on the peculiar problems of open highways.

For example, Gazis, et al (GA1) considered individual vehicles of a line of moving vehicles and postulated the reaction of the average driver to the braking and accelerating behavior of the car ahead. Although this model was used satisfactorily for investigating local and asymptotic stability of the system of vehicles and "correctly" simulated car-following data observed in the Holland Tunnel, it has several weaknesses. Formulated as a linear model, it describes poorly the transitions between widely different steady state speeds. As a nonlinear model it overcomes this failing, but still does not account for certain pnysical constraints, such as the limited accelerating capability of a car.

In still another approach, Lighthill and Whitham (LW1) modeled traffic flow as a continuous process. They theorized on the existence of shock waves created, for example, at bottlenecks and signalized intersections but failed to get good correspondence with real data since the theory was based on an assumed flow-density function and neglected the detailed maneuvers of the cars in changing speed.

Others have viewed the traffic problem as a stochastic one and have attacked it with the tools of the statistician. Of particular interest is the work of Beckmann, Tanner, Herman et al., Haight and others, in which the problem of queuing at signalized (BE1, TA1) and non-signalized (HR1, HA1) intersections is investigated. These intersection models, however, are very limited in scope since most often only a single intersection can be effectively modeled.

None of these modeling approaches are addressed to the specific problem of modeling a complete signalized traffic system. They are inadequate for describing the peculiar platooning effects of the traffic signals (although there is some attempt to simulate the behavior of the platoon after it is formed). They do not simulate the traffic routing of an actual arterial system (i.e., turning movements). Finally most of them satisfactorily describe steady state behavior but fail to accurately describe the acceleration and deceleration transients occurring at intersections.

Goodnuff (GO1) has investigated traffic systems and established a model which simulates the peculiar components (e.g., multi-laned arteries and intersections) and behavior (e.g., turning movements) usually encountered in traffic systems. For heavy density operating conditions, he successfully formulated an optimization algorithm which clears a grid of queued vehicles in minimum time. The most important variables in Goodnuff's system are those describing the queues formed at each intersection. Since vehicles must stop for each signal, it is unnecessary to track them as they proceed through an intersection toward the tail of the next queue--their position as they traverse the space between queues provides no useful information. Only when the system is successfully reduced to a lower density mode do these movements become important, but at lower densities the model assumptions are no longer valid.

In the course of solving the control problem for medium to light density conditions, Chang (CH1) has developed a traffic model which suitably describes some of the phenomena of arterial traffic. Chang's model takes account of the queues and the vehicular flow between queues. This movement he considers as a continuous flow. Because his ultimate concern is the optimal setting of traffic signals, the approximations he makes for velocity (it is always constant), acceleration (he neglects it), etc., are justifiable. However, without this ulterior motive the model in its present form inadequately simulates vehicular flow and, even further, has no provisions for describing phenomena such as turning movements.

2.2 Present Model

The lack of a versatile model for simulating traffic flow within a system of signalized arteries led to this investigation. The goal, from a qualitative standpoint, is to develop a model which describes traffic flow for medium heavy to medium light density conditions such as might exist in morning and evening rush hours. It should be noted that in this mode the behavior of vehicles in transit is of equal importance with the behavior of those queued at the intersections.

Before pursuing details of the model, it is necessary to establish its nature. The model can be neither too elaborate nor simple. An elaborate model could achieve the stated goal by tracing the path of each vehicle through the system while maintaining a continuous surveillance of surrounding vehicles. Predictably, however, it becomes too complex and the computations inefficient as the system approaches any meaningful size. On the other hand, a continuous model is too simple since it does not depict all the phenomena that are important to the control problem.

Between these extremes there exists a suitable approach to modeling an urban system. Introducing platoons and queues as the smallest vehicular units allows studies to be made of large systems without becoming cumbersome. Equally important, it preserves enough of the identity of the vehicles that acceleration phenomena, turning movements and vehicle counts can be incorporated.

It has been observed that vehicles in close proximity to each other behave similarly in many respects, and many mathematical theories rely on this fact to describe average vehicular behavior and relative motion between vehicles. For the assumed densities, then, it is reasonable to model vehicles as platoons and queues and to describe the platoons and queues by average vehicular values.

The model can be either deterministic or stochastic in nature. It is assumed that the environment in which the vehicles move (the medium vehicular density and the relatively short distances between signals) and the platooning effects of the traffic signals constrain the individual's movements so that they are realistically described in a deterministic way. Some blending with statistical ideas is achieved in the model via the description of the generation of input vehicles and of the vehicle behavior at the intersections where turning is allowed.

2.3 Basic Component Introduced

A large system is often considered as an interconnection of primitive elements or components. The properties of such a component can be defined without reference to any other components. It is not necessarily the simplest such part since it may be possible to resolve it into a set of even simpler pieces.

Between every successive pair of traffic signals there lies a section of pavement which carries traffic in one direction. This length of pavement is an arterial section. It serves well as a base for a traffic system component since all the phenomena of a complete system can be observed on it. The complete component consists of the arterial section, the upstream signal, the upstream queue, and the platoons in transit on the section. Figure 2.1 illustrates such a component. All positions are measured positively with respect to the signalized end of the component.

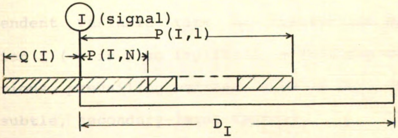


Figure 2.1. Traffic model component.

Associated with this component is a set of equations describing the vehicular units. The platoon description is complete if its length, its position, and its number of vehicles are known for every instant of time. A queue is described by its length and its number of vehicles. The set of equations given here are complicated since the platoons and queues are functions of many primary- and secondary-level factors.

$$\frac{dz}{dt} = f\,(z, \Delta_j, D, t) \qquad (2.1)$$

where

$$z = (P^k_{ld}, P^k_{tr}, n^k_p, Q, n_q)^T, \; k = 1,\dots, n \qquad (2.2)$$

The elements of the z vector are the position of the leading edge of each platoon, P^k_{ld}; the position of the trailing edge of each platoon, P^k_{tr}; the number of vehicles in each platoon, n^k_p; the queue length, Q; and the number of vehicles belonging to the queue, n_q. The jth phase of the traffic signal is denoted by Δ_j, the length of the arterial section is given by D, and the independent variable time is represented by t. The function f (·) is also implicitly a function of street conditions, prevailing weather, time of day, and other, more subtle, secondary-level factors.

Since more than one platoon may exist on an arterial section at a time, the index k is used to distinguish them. As each platoon is formed it is given a new index; thus the latest platoon has the highest index n. The dimension of the vector z is 3n + 2.

The relation expressed in equation (2.1) can be made more tractable if

(1) The number of elements in z can be reduced.

(2) Velocity and vehicular density on an arterial section are strongly correlated.

(3) The theory of queuing is applicable.

(4) Arterial streets have no inclines, banking or curves.

When these assumptions are incorporated into the model, it is possible to spell out the equations explicitly and yet not sacrifice accuracy.

The first statement suggests that either some of the elements of z are redundant or that a better set can be found.

The second statement, supported by theoretical and experimental studies, suggests that vehicular density is a first-order effect in the determination of platoon and queue positions. Conversely, the number of vehicles (or density) in the platoons and queues are determined almost wholly by the average velocity of vehicles. The second-order effects (arterial geometry, weather conditions, etc.) are in comparison negligible but accounted for implicitly in the velocity-density relation.

The third merely states that the description of the phenomena observed at the signalized intersections can be couched in the terminology of queuing theory. The last statement disallows peculiar arterial geometries and suggests that the effects of geometry in the determination of densities and velocities be relegated to a secondary role.

By confronting each of these assumptions in depth, the simplifications can be achieved. Before proceeding, however, it is helpful to note that vehicular density and lane occupancy are alternative measures for the number of vehicles. Vehicular density is the number of vehicles per unit pavement length. Lane occupancy is a normalized density defined in the following manner:

$$\text{LANE OCCUPANCY} = \frac{\text{total vehicular length}}{\text{total pavement length}}$$

2.4 Selection of State Variables

In a queue the vehicular density is at a maximum and the velocity is essentially zero. This value for density, the ratio of the number of vehicles to the queue length, is generally assumed to be a constant. Because the length is proportional to the number of vehicles, either is sufficient to serve as the state variable describing the queue. For computational ease, the length is selected.

The state description of a platoon must include any set of independent variables from which its length, position, and number of vehicles can be determined. The following list suggests most candidates.

1. Position of leading edge of platoon, (ft).

2. Average vehicular density of platoon, $(\frac{veh}{ft})$, or average lane occupancy $(\frac{\text{vehicle ft}}{\text{pavement ft}}) = (\frac{ft}{ft})$.

3. Number of vehicles in the platoon, (veh).

4. Pavement length of the platoon, (ft).

5. Total length of vehicles in the platoon, (ft).

6. Mean headway between vehicles, (ft or sec).

Variables (2, 3, 5), for example, are not independent.

An arterial section may hold several platoons at a time. A simplification results if one always assumes that all the space behind a platoon is occupied by other platoons, considering a free space as a platoon having a vehicular density of zero. Consequently, the platoon state at any time can be provided by any two variables listed above except for the pairs (1,4), (3,5), (1,6), and (2,6). The same two must be used to describe each platoon. For computational reasons the position of the leading edge and the vehicular length are selected to describe each platoon.

Since the states of the platoons and queues are derived from the density and velocity, the relations between these variables are established by experimental and theoretical investigations.

2.5 Velocity-Density Relations for an Arterial Section

Drivers in a traffic stream, aware or not, react to increasing density by lowering their speed. This natural control mechanism was studied closely by Greenshields in 1934 and led him to conclude a linear relation between the speed of vehicles in the traffic stream and the stream density. Subsequent experiments have substantiated that for many purposes Greenshields' linear model is realistic. Thus the speed is given by

$$v = v_f \left(1 - \frac{x}{x_j}\right) \qquad (2.3)$$

where v_f is the free speed, a mathematical value for speed as density approaches zero. The jam density, the density at which the speed goes to zero, is denoted by x_j. A typical value for x_j is 40 per cent of the bumper-to-bumper density. This relation, established under very restrictive conditions, applies to steady state conditions for vehicles moving on a highway (i.e., an uncontrolled artery) and it applies to average values of the variables.

Although the equation does not reflect it, Greenshields introduced a kink at the top of the graph to describe, in a realistic way, the region where the speed is unaffected by density below a certain limiting value.

This truncation is observed for any speed-density relation.

On a controlled artery when the densities are light and the speeds are normally higher, the conditions are not too unlike the steady state stream. With x_j about forty per cent equation (2.3) predicts velocity quite well.

As the density increases, however, the velocity does not approach zero as quickly as equation (2.3) predicts. The average velocity is zero on an artery when it is filled from intersection to intersection with a queue. Thus under these conditions x_j should equal x_q.

Stated mathematically, the equations describing velocity and density for arterial traffic are given as

$$v = v_{f_1} \left(1 - \frac{x}{x_{j1}}\right), \quad 0 \leq x \leq x_1 \tag{2.4}$$

$$v = v_{f_2} \left(1 - \frac{x}{x_{j_2}}\right), \quad x_1 \leq x \leq x_{j_2} = x_q \tag{2.5}$$

where x_{j1} is approximately 0.4. The constants v_{f_1}, v_{f_2}, and x_1 are selected to match observed traffic behavior on particular arteries. Figure 2.2 depicts a typical velocity - density characteristic.

The relations given in equations (2.4) and (2.5) are used to determine the average speed for all platoons on a particular arterial section. Speeds on other sections are determined similarly. Individual vehicle speeds, it must be emphasized, may be somewhat different from this average. Within the platoons, particularly when the density is light, accelerating and decelerating vehicles and passing phenomena may be present.

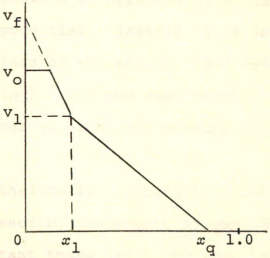


Figure 2.2. Velocity-density characteristic.

2.6 Queues Examined

In the terminology of queue theory an intersection is regarded as a rate-limited server which is subject to breakdowns. However, despite the impressive amount of literature available on queue theory in general and on traffic congestion in particular, most work has centered on the problem of gap acceptance (i.e., vehicle crossing or merging) (TA 1, MR1) or on the relatively simple problem of a single traffic signal on a two-lane artery (HA1, NE1). The problems encountered when queue theory is used as a primary analytic tool in the study of a complete traffic system are far too difficult (WE1).

The main difficulties result from the "artificial" behavior that the signals impose on the traffic. No longer can the distribution of arrivals at a signal be considered Poisson or exponential. Instead it is intimately related to the parameters of the signals (red and green times, relative phasing). For the same reasons the distribution of service times at each intersection involves intractable mathematics.

Nevertheless it is possible to utilize some queuing concepts to describe the events at the intersections. At any given instant there are n vehicles in a queue. The first vehicle in line enters the intersection. The time elapsing between this first entry and the entry of the vehicle next in line is the service time t_s for the first vehicle. During this time interval α vehicles arrive at the queue's end. The number of vehicles in the queue at the end of the service time is given by

$$n' = n - 1 + \alpha \tag{2.6}$$

It may happen that the original queue has zero length. If so, it is necessary to await a vehicle's arrival so that n=1, and consider the service time for it. Then,

$$n' = \alpha \tag{2.7}$$

The two above equations can be combined into the single equation

$$n' = \max(n-1,\ 0) + \alpha \tag{2.8}$$

or even more simply as

$$n' = n - 1 + d + \alpha \tag{2.9}$$

if d is defined as follows

$$\begin{aligned} d &= 1 \qquad \text{if } n = 0 \\ &= 0 \qquad\quad\ \, n > 0 \end{aligned} \tag{2.10}$$

Alternatively, during a time interval Δt a total vehicle length ρ leaves the queue and a length $\alpha\bar{\ell}$ arrives at its end so that equation (2.9) can be written

$$\frac{(n'-n)}{\Delta t}\,\bar{\ell} = \frac{-\rho}{\Delta t} + \frac{\rho d}{\Delta t} + \frac{\alpha\bar{\ell}}{\Delta t} \tag{2.11}$$

Average vehicle length is given by $\bar{\ell}$ and is used here to convert the number of vehicles to an equivalent vehicle length.

In the limit as Δt approaches zero

$$\dot{Q} = -V_q + V_q d + \gamma \tag{2.12}$$

where $\dot{Q}$ is the net rate of queue length change, v_q is the rate of outward flow, and γ is the rate of increase in queue length, all measured in feet/sec or some other equivalent units.

The effect of the signal can be introduced as a second, but imaginary, queue served by the intersection which has a "head of the line" priority. It has an arrival rate which is a constant, one per signal cycle. Its service rate equals a red phase. Viewed from the real queue the effect of the imaginary one is to cause the intersection to switch continuously between operation and breakdown.

Equation (2.12) describes the observed behavior at intersections. However, the rate v_q at which the vehicles are served must still be determined. As noted previously this rate is governed to a great extent by the signal timing since the timing determines the arterial velocity.

At the beginning of the green phase of a signal the queue, assuming that its length is not zero, injects the first vehicle into the next arterial section. After a moment the queue sends another vehicle into the section and continues to do so until the queue is dissipated or the signal changes phase, at which time the next platoon begins to form. The spacing between vehicles determines the vehicle density within the platoon. However, this density differs significantly from the average density observed on an arterial component.

Platoon density is closely tied to the velocity prevailing during its formation. A rule of thumb suggested by safety advertisements, etc., advises that a driver allow a vehicle length between vehicles for each ten miles per hour of speed. This relation stated mathematically gives the lane occupancy of a platoon as

$$x_p = \frac{1}{1 + \frac{v}{10}}\text{, v in mph} \tag{2.13}$$

or

$$x_p = \frac{1}{1 + \frac{v}{14.7}}\text{, v in fps} \tag{2.14}$$

The "ten" figure is not rigid, it could be some more accurately determined value. (One suspects, however, that this advice is not the result of idle daydreaming but corresponds closely to the natural tendencies of the average safe driver.) In order to have equation (2.13) consistent with the requirement that vehicles at rest have a density x_q the following modified equation is used instead.

$$x_p = \frac{1}{\frac{1}{x_q} + \frac{v}{10}} \tag{2.15}$$

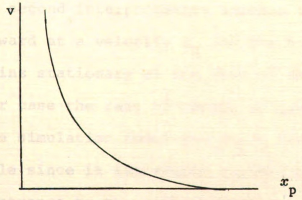


Figure 2.3. Velocity-platoon density characteristic

It is now possible to demonstrate the close correlation of the rate at which vehicles are discharged from a queue and the speed of the platoon on the arterial. This results from the fundamental requirement that the flow into an intersection must equal the outward flow. Equating the inward flows at an intersection during T seconds,

$$x_q \; v_q \; T = x_p \; v \; T \tag{2.16}$$

where x is average density and v is average velocity. Obviously, then

$$v_q = \frac{x_p}{x_q} \; v \tag{2.17}$$

This equation can be interpreted in two ways. As noted earlier the queue can be regarded as standing still but becoming shorter as a "shock wave" moves backward through it at a velocity v_q. The shock wave is the discontinuity resulting from the difference between the queue and platoon densities. The second interpretation assumes that the queue moves forward at a velocity v_q and the so-called shock wave remains stationary at the foot of the intersection. In either case the rate of change of queue length is $-v_q$. For the simulation model the first interpretation is less desirable since it introduces a platoon between the queue and the entrance to the intersection.

2.7 Component Equations

Some vehicles which leave the queue turn rather than continue straight. The fraction that continues straight is given by the constant "a."

Using the state variables selected earlier and the constraint conditions imposed by queue theory, the component equations are as follows for the ith arterial section.

$$\Delta_{ij} = \begin{cases} 0, & \text{if jth phase of ith signal is red} \\ 1, & \text{otherwise} \end{cases} \tag{2.18}$$

$$\dot{P}(i,k) = v_i,\ k=1,\ldots,n;\quad 0 \leq P(i,k) < D_i - Q(i+1) \tag{2.19}$$

$$\dot{P}_x(i,k) = \begin{cases} -P_x(i,k)\,\delta\,(t-T_{i,k}), & k=1,\ldots,n-1 \\ a_i v_i \dfrac{(x_p)_i}{x_q}\Delta_{ij} - P_x(i,n)\,\delta\,(t-T_{i,n}), & k=n \end{cases}$$

$$Q \geq 0 \tag{2.20}$$

$$\dot{Q}(i) = \frac{-v_i (x_p)_i}{x_q}\Delta_{ij} + P_x(i-1,k)\,\delta(t-T_{i-1,k}),$$

$$Q(i) \geq 0 \tag{2.21}$$

$$x_i = \frac{x_q\, Q(i+1) + \sum_{k=1}^{n} P_x(i,k)}{D_i} \tag{2.22}$$

$P(i,k)$ is the position of the front edge of the kth platoon; $P_x(i,k)$ is the vehicular length of the kth platoon. $Q(i)$ is the queue length for the ith section. The characteristic velocity, $v_i = f(x_i)$, is determined according to Figure 2.2 once the average density x_i is known. D_i is the pavement length of the ith section. The platoon density x_p is determined from equation (2.15).

The impulse function is denoted by $\delta(t)$; $T_{i,k}$ is the time of occurrence for $P(i,k) = D_i - Q(i+1)$.

The use of the indices i-1 and i+1 implies that the ith queue extends into the (i-1)st arterial section and the (i+1)st queue extends into the ith arterial section.

Equations (2.19 - 2.22) are less formidable if it is kept in mind that vehicles released from the ith queue during a green phase (2.21) appear as part of the nth platoon (2.20). The platoon travels the length of the arterial section (2.19) until it joins the upstream queue where it ceases to exist as a platoon (2.20). In the meantime new platoons are added to the ith queue (2.21). It is this cyclic conversion of the vehicles from a queue to a platoon to a queue which characterizes the model.

It is precisely this cycling of vehicles which allows a traffic system to be represented as an interconnection of similar components. Also it permits the use of iterative techniques to obtain solutions. Because of the size of the problem, the non-linearities involved, and the randomness of certain of the variables, these solutions are best obtained using a digital computer.

CHAPTER III

SIMULATION

Between the phases of modeling a component and programming a computer there lies the important process of simulation. This simulation is a demonstration of how a complete traffic system can be interpreted as an interconnection of basic components. In this chapter the simulation models for several traffic structures are developed. With the addition of acceleration phenomena, random input generation, and a varying turning pattern the simulation can be made adaptable to most situations encountered in real systems.

3.1 Acceleration Phenomena

With the inclusion of acceleration effects the model has a greater potential for simulating traffic behavior. The model presently employs a constant acceleration function which could be generalized to any function if future experiments dictate a change.

The acceleration process is simple to describe. When the signal phase becomes green the first vehicle in the queue crosses the intersection to begin a platoon.

The velocity of this platoon is initially zero, increasing to the prevailing average velocity v_i corresponding to the average density on the section.

In both real systems and models the process of decelerating is more complicated. When the signal turns green, the vehicles accelerate or remain still if their path is blocked. On the other hand, the driver of a vehicle approaching an intersection at a constant velocity must decide whether to begin braking. He must consider the phase of the approaching signal, the behavior of the traffic ahead, the distance to the next intersection, and his own speed. In a simulation model, these same factors must be weighed.

In the present model, a driver makes the decision to brake or continue at the same speed through an intersection when he is at a critical distance from the intersection. This critical distance D_{cr} is a function of the vehicle's velocity v and the number of vehicles before it. It is braked for any one of the following reasons.

1. A queue lies ahead.

2. The signal ahead is red and will remain red for a time T greater than D_{cr}/v.

3. The signal ahead is green but will change phase in time T less than D_{cr}/v.

4. The vehicles ahead are decelerating.

In the model, after this decision is made, the vehicles involved are transferred from their own platoon to one of two transition platoons which carries them from the critical point to the intersection at either (1) a constant velocity or (2) a decreasing velocity.

3.2 Turning Movements

The turning movements at an intersection observed over a long period of time can be used to establish a value p which is the ratio of the number of vehicles that turn to the total number that enter the intersection per unit time interval (a green phase). This average value can be interpreted as a probability estimate that a vehicle turns. Alternatively, p is the expected value of the fraction of a queue that turns. Although this fraction can assume any value in the closed interval [0.0, 1.0], for computation it is easier to discretize the set to the finite set $\{\frac{n}{100} : n = 0, 1, \ldots, 100\}$. The problem of arriving at the fraction "a" of turning vehicles for a particular green phase can be handled as follows.

The queue in question is normalized to one hundred vehicles. A binomial probability function is used to establish the probability that k vehicles turn

$$P(X = k) = \binom{100}{k} p^k (1-p)^{100-k} \tag{3.1}$$

where the random variable X has possible values 0,1,...,100.

A complete table of these random variables and their corresponding probabilities is generated. In the simulation a random variable r which is uniformly distributed on the interval [0.0, 1.0] is generated. The turn ratio $a = \frac{k}{100}$ is determined by searching the table for k such that

$$F_X(k-1) < r \leq F_X(k) \qquad (3.2)$$

Here $F_X(\cdot)$ is the distribution function defined as $F_X(k) = P(X \leq k)$.

3.3 Input and Output Elements

The portion of the traffic network being simulated may have several sections connecting it to other streets or parking facilities. The sections by which vehicles enter the simulated traffic network are the inputs, and those by which the vehicles leave are the outputs.

The queues of the input sections are formed in a way uniquely different from the queues of the other sections. Vehicles are assumed to arrive at these queues in a random manner. The traffic signals and traffic behavior outside the system, location of parking lots, time of day, etc., play significant parts in establishing the distribution of arrivals. However, as a simplified representation a Poisson probability function is used in this model to describe the arrivals.

An expected value λt describes the average demand observed. The Poisson probability function given by

$$P(X = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \tag{3.3}$$

is used to determine the number of arrivals in the time interval t. A random number r having a uniform distribution on the interval [0.0, 1.0] is generated every t seconds. The number of arrivals k is then established every t seconds by

$$F_X(k-1) < r \leq F_X(k) \tag{3.4}$$

In this latter relation, the distribution function is denoted by $F_X(\cdot)$. For an input section it is easy to continuously adjust the probability parameters to simulate a changing arrival pattern.

The output sections are hardly different from the internal sections. Instead of allowing the vehicles to form into platoons after crossing the intersection, they are accumulated in a counter or sink. This implies that the sinks are of great enough capacity (infinite) that there is no cumulative congestion which reflects back into the system. If this is not true, it is necessary to extend the portion of the network being simulated until such a section is reached or to reduce the flow rate into the output section as congestion builds up.

3.4 System Simulation

In the simulation of traffic systems, components like the one modeled in the previous chapter are connected together according to the geometric pattern of the system. This model has associated with it a connection matrix which details the manner in which these connections are made and the direction of traffic flow. In all respects it is analogous to the incidence matrix used in graph theory (KT1). The matrix has dimensions 2 x m, a column for each of the m components. The first entry of the ith column is the index of the component following the ith component. The second entry of the ith column is the index of the component preceding the ith. An input or output element has a zero for its second or first entry, respectively.

In the following examples the components are represented by the simplified symbol shown in Figure 3.1.

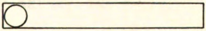


Figure 3.1. Traffic component symbol.

One-way artery

A one-way artery with m signals is the simplest of all traffic structures to simulate. The complete model consists of m traffic components joined end to end as in Figure 3.2. The first one is an input component and the

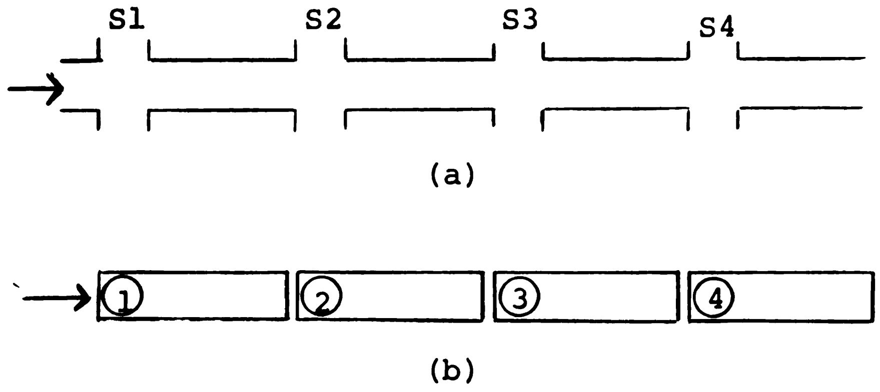


Figure 3.2. (a) Artery with m signalized intersections.
(b) Model of artery (a).

last an output. No indexing scheme is implied by the formulation, so for simplicity the components are numbered successively beginning with the input element. For the system of Figure 3.2 the connection matrix is

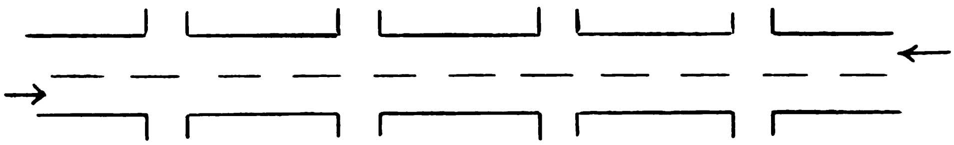
$$KM = \begin{bmatrix} 2 & 3 & 4 & 0 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

Two-way artery

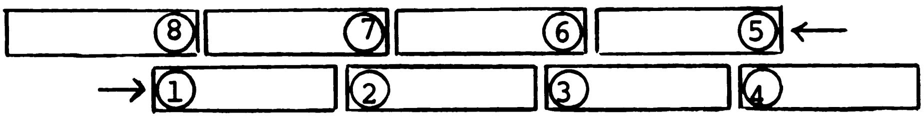
A two-way artery can be considered to be two one-way arteries side by side as shown in Figure 3.3. It is important to remember, however, that at each intersection the signal phase, red or green, is the same for both the inbound and outbound directions.

Intersection of two one-way arteries

The model for a pair of intersecting one-way streets can be considered as two arterial components at



(a)



(b)

Figure 3.3. (a) Two-way artery.
(b) Model of artery (a).

right angles to each other as shown in Figure 3.4. Each has a turning coefficient defined for it. Since there is only one signal for the two components, they must share it.

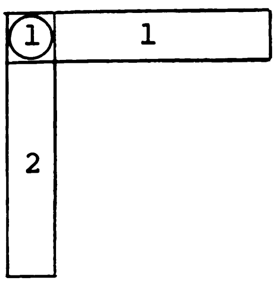


Figure 3.4. Basic component for grid structures.

The component equations for this basic grid element are as follows. In these equations i = 1, 2.

$$\Delta_{1k} = \begin{cases} 0, \text{ if kth phase is red} \\ 1, \text{ otherwise} \end{cases} \tag{3.5}$$

$$\bar{\Delta}_{1k} = \text{ logical complement of } \Delta_{1k} \tag{3.6}$$

$$\dot{P}(i,j) = v_i,\ j = 1, n;\quad 0 \leq P(i,j) < D_i - Q(i+1) \qquad (3.7)$$

$$\dot{P}_x(i,j) = \begin{cases} -P_x(i,j)\,\delta\,(t-T_{i,j}), & j = 1, n-1 \\ \dfrac{a_i v_i (x_p)_i \Delta}{x_q} + (1-a_{\bar{i}})\,\dfrac{v_i (x_p)_i}{x_q}\,(1-\Delta) \\ -P_x(i,n)\,\delta\,(t-T_{i,n}), & j = n \end{cases} \qquad (3.8)$$

$$Q(i) \geq 0$$

$$\dot{Q}(i)\ \frac{-v_i (x_p)_i}{x_q}\,\Delta + P_x(i-1,\ j)\,\delta\,(t-T_{i-1,j});\ Q(i) \geq 0 \qquad (3.9)$$

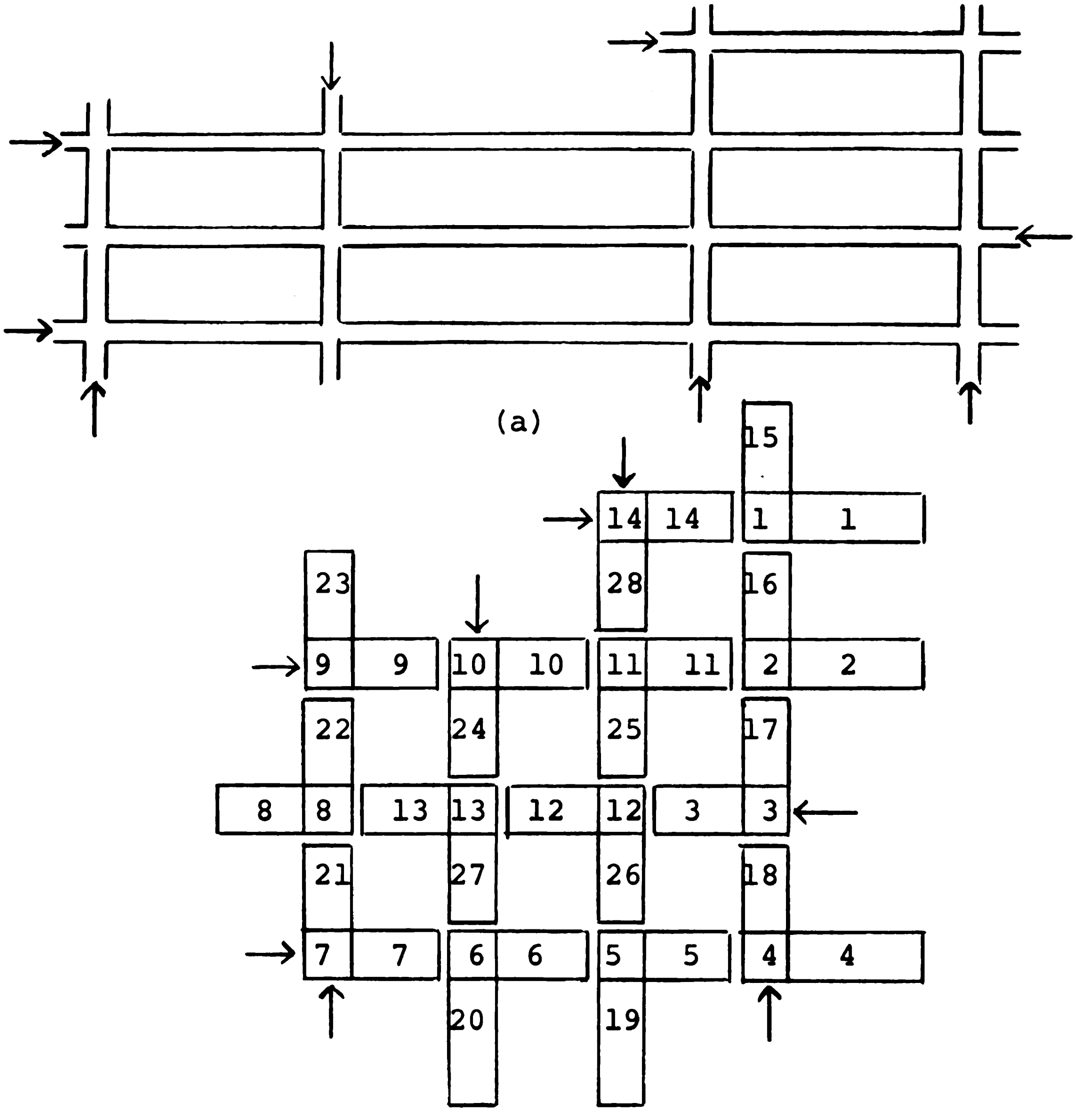
$$\Delta = \begin{cases} \Delta_{1k} & \text{if } i = 1 \\ \bar{\Delta}_{1k} & \text{if } i = 2 \end{cases} \qquad (3.10)$$

$$\text{where } \bar{i} = \begin{cases} 1 \text{ if } i = 2 \\ 2 \text{ if } i = 1 \end{cases} \qquad (3.11)$$

The turning split factor is a_i.

<u>Grid of one-way arteries</u>

Within a grid of one-way arteries the signal at each intersection is shared by the two competing directions of traffic. In the model, therefore, the grid components described above is the basic building element. For convenience, especially in a computer simulation, the following numbering scheme is suggested. Label the signals in any order from 1 to m. The horizontal component associated with signal i is labelled i and the vertical component is labeled m + i, as shown in Figure 3.5. The actual values



(a)

(b)

$$KM = \left[\begin{matrix} 0 & 0 & 12 & 0 & 4 & 5 & 6 & 0 & 10 & 11 & 2 & 13 & 8 & 1 \\ 14 & 11 & 0 & 5 & 6 & 7 & 0 & 13 & 0 & 9 & 10 & 3 & 12 & 0 \end{matrix}\right.$$

$$\left.\begin{matrix} 0 & 15 & 16 & 17 & 0 & 0 & 22 & 23 & 0 & 27 & 26 & 19 & 20 & 25 \\ 16 & 17 & 18 & 0 & 26 & 27 & 0 & 21 & 22 & 0 & 28 & 25 & 24 & 0 \end{matrix}\right]$$

(c)

Figure 3.5. (a) Traffic grid of one-way arteries.
(b) Model for grid (a).
(c) Connection matrix.

for a_i ($i = 1,\ldots,2m$) must be determined through observation. Note that there are several input components in a grid. The connection matrix for the system is also given in Figure 3.5.

Intersection of a two-way artery and a one-way artery

A vehicle on a one-way artery which crosses a two-way artery can turn either right or left while a vehicle on the two-way artery can turn only right or left depending on which direction it is traveling. To simulate these turning options in a model, a dummy is introduced at the intersection on the one-way artery as shown in Figure 3.6. This dummy (labeled 1) has a length $D_1 = 0$, and the right turning coefficient for the one-way artery is defined on it. Signals 1 and 2 may work in unison or the red phase on the one-way artery of signal 1 may be delayed slightly in order to simulate the amber phase during which vehicles from the two-way artery making left turns have the

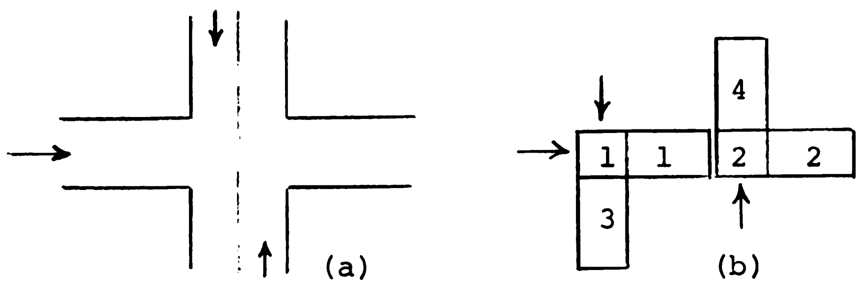


Figure 3.6. (a) Intersection of a one-way artery and a two-way artery.
(b) Model of intersection (a).

opportunity to complete their turn. These vehicles turning left, therefore, do not interfere with the oncoming platoons.

Intersection of two two-way arteries

The model for a pair of intersecting two-way arteries uses four dummy components to simulate the traffic flow. As shown in Figure 3.7 the dummies are numbered 1, 4, 6, and 7; the right turning coefficients are defined on them. All four signals must operate in unison. Unfortunately there is no simple way of simulating an amber phase for all four directions simultaneously.

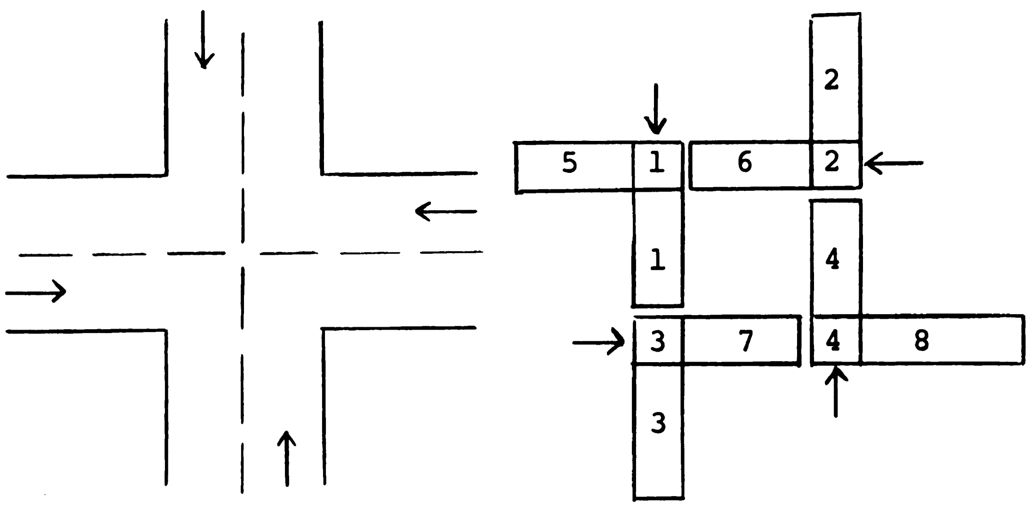


Figure 3.7. (a) Intersection of two two-way arteries.
(b) Model of intersection (a).

Multiple lane arteries

If an artery has more than one lane then some modifications are necessary in the component equations. If m_i is the number of lanes in the ith arterial section, then equations (2.20-2.22) are modified in the following manner, respectively.

$$\dot{P}_x(i,j) = \begin{cases} P_x(i,j)\ \delta\ (t-T_{i,j}),\ j = 1,\ldots,n-1 \\ m_i v_i \frac{(x_p)_i}{x_q}\ \Delta_{1k} - P_x(i,n)\ \delta\ (t-T_{i,n}),\ j = n \end{cases}$$

$$Q(i) \geq 0 \qquad (3.12)$$

$$\dot{Q}(i) = -m_i v_i \frac{(x_p)_i}{x_q}\ \Delta_{ik} + P_x(i-1,j)\ \delta\ (t-T_{i-1,j});$$

$$Q \geq 0 \qquad (3.13)$$

$$x_i = \frac{x_q Q(i+1) + \sum_{j=1}^{n} P_x(i,j)}{m_i D_i} \qquad (3.14)$$

The simulation ideas presented in this chapter have been incorporated into a digital computer program which is capable of handling a significant traffic area. In the Appendix a flow chart of this program is given along with a list of the symbols used to represent the variables in these chapters.

3.5 Example

Under the best of conditions the worth of a simulation model can be demonstrated by comparing data from a real system and from the model. Evaluating the model developed in these chapters would be easy if adequate traffic data were available. Unfortunately, much of the data from large traffic systems consist merely of hourly vehicle counts. Usually these are for widely separated points and are acquired over a period of several months. Consequently, one cannot know accurately how events at one point affect the vehicular movements at other points in the system. Little of the data is concerned with velocity and density variations over short time intervals. Thus, evaluating the wealth of data generated by the model is difficult.

The system used in the following simulation is the CBD (Central Business District) of Lansing, Michigan. The arterial component network is developed from a street map supplied by the Traffic Division of the City of Lansing (GE 1). These maps are shown in Figures 3.9 and 3.8, respectively. In Figure 3.8 there are 72 signalized intersections. Due to the additional dummy elements, the number of traffic signals in Figure 3.9 is increased to 100 and the number of arterial components is 200.

The number of lanes and arterial lengths are accurately depicted in the model. The other necessary data are estimated as accurately as possible. Input rates, based on typical hourly counts supplied by the traffic department, range from 0.1 to 0.2 vehicles/sec/lane. The expected values for the turning split factors are based on these counts and on the system geometry.

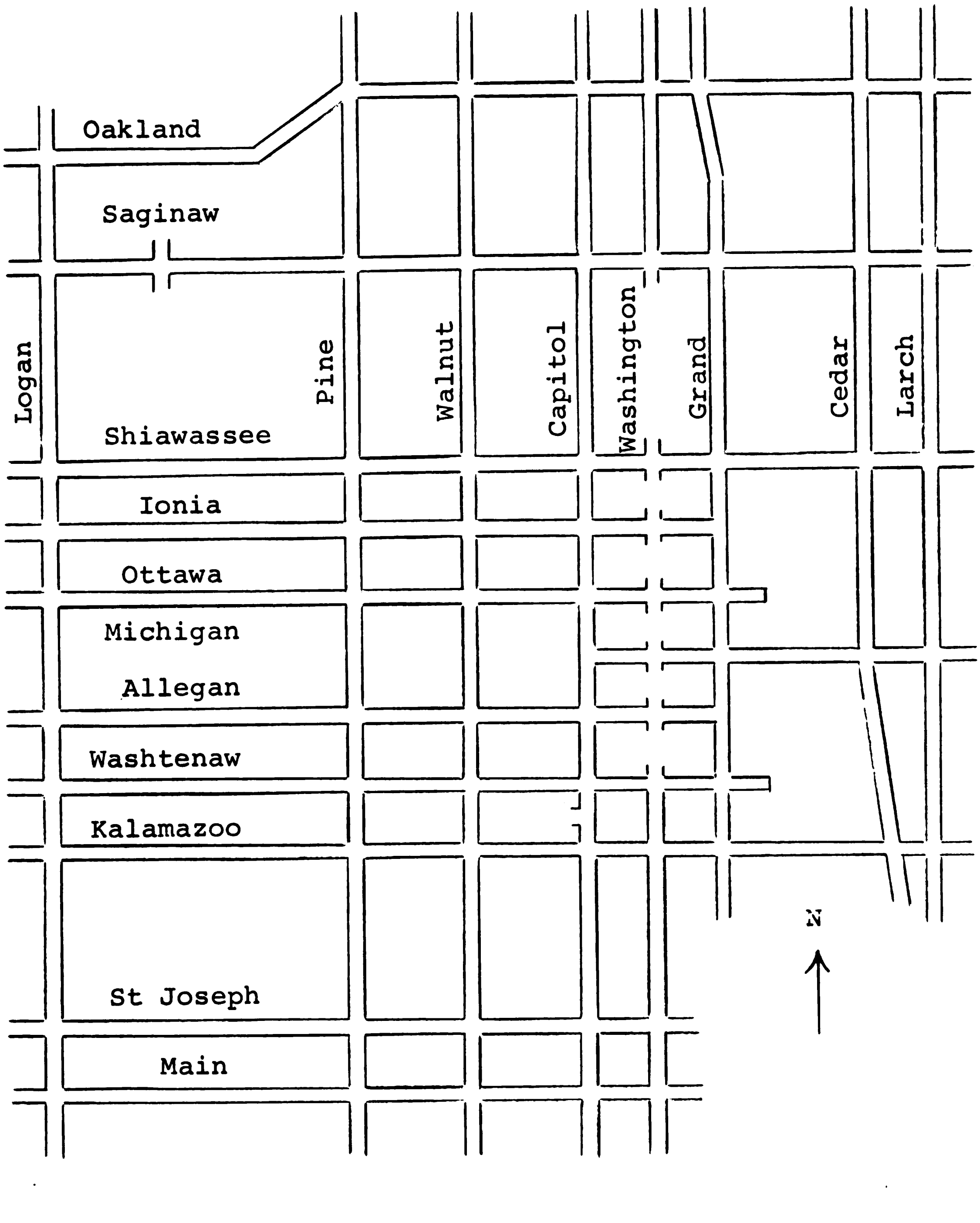


Figure 3.8. Traffic grid of Lansing, Michigan's central business district.

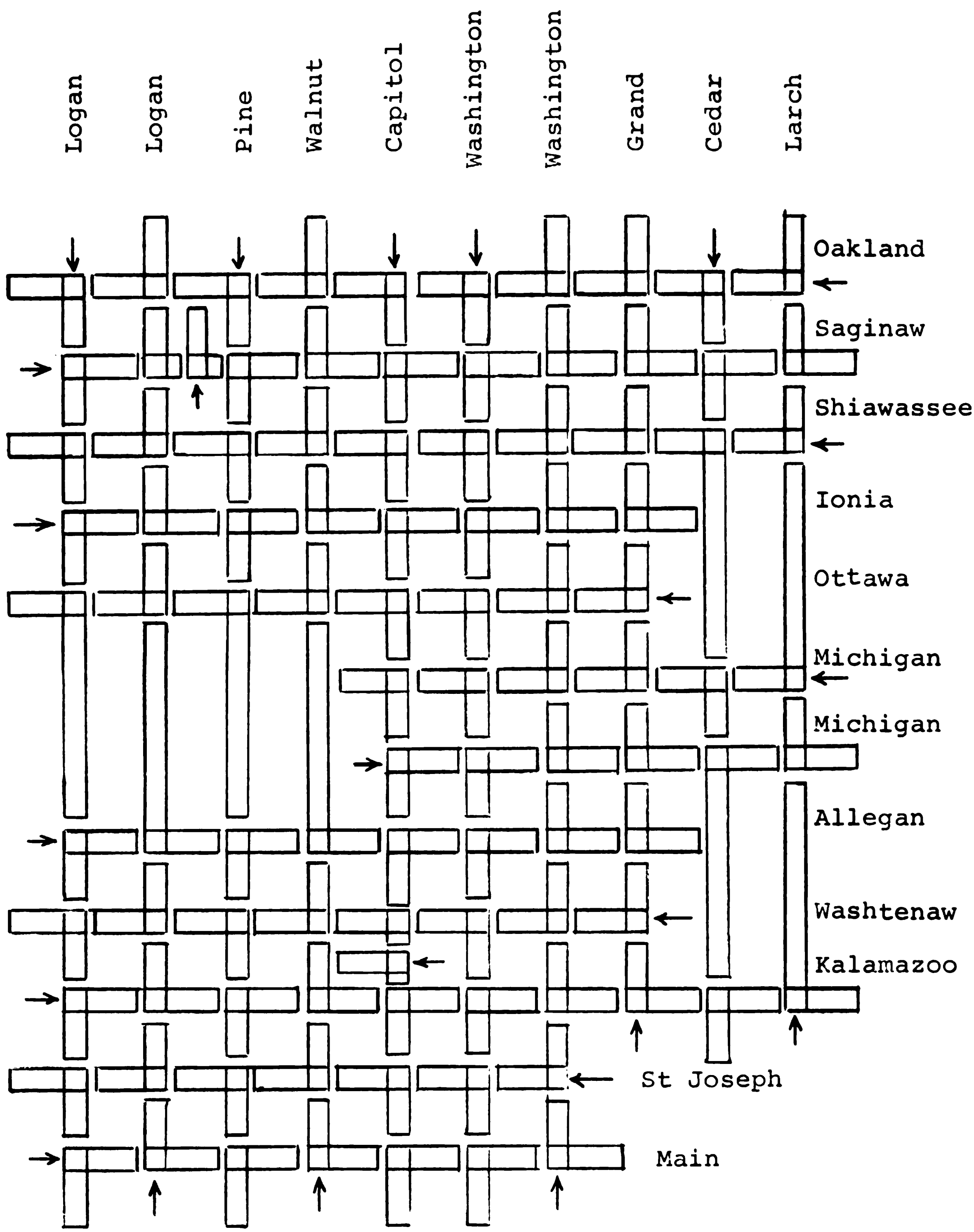


Figure 3.9. Model of grid in Figure 3.8.

For example, all vehicles traveling west on Michigan Avenue must turn left at Capitol Avenue, and thus the turning split factor is zero. It is assumed that all the traffic signals operate with a 60 second cycle and an equal red-green split. The relative timing of these signals is more or less random.

A constant acceleration of 5 ft/sec^2 and a free speed of 60 ft/sec is used throughout the model. Referring to Figure 2.2, v_o, v_1, x_1 and x_q are 57 ft/sec, 30 ft/sec, 0.5 and 0.85, respectively. The vehicles in the system are assumed to have an average length of 20 feet.

This hypothetical study of Lansing's traffic has two main objectives. The first is to demonstrate the effectiveness of the model in simulating a large, realistic system and to establish a measure for the ratio of computer time to real time. The second objective is to illustrate the variety of investigations which are possible with the model. These investigations may be either macroscopic--dealing with such variables as vehicle counts and velocities for an extensive area of the system--or microscopic--dealing with the detailed behavior on a small portion of the system.

Simulation

The simulation was performed on a CDC 6500 computer. Using a time increment of 1.0 second, 2000 seconds were simulated in 252 seconds. Based on these figures, approximately 0.63 millisecond is required to simulate the traffic behavior on a single component for one increment of time. As indicated in the following studies the data available as output varies widely.

Microscopic Study

Between Saginaw and Shiawassee streets on Logan, the southbound platoon and queue behavior were studied as a function of time. Data were printed every second. The leading and trailing edges of the platoons and queues are plotted on space-time coordinates in Figure 3.10. It is easy to follow the cyclic behavior of the vehicles: their accumulation at Saginaw, their acceleration, their transit to Shiawassee and their deceleration. In the figure one also notes that during the red time vehicles are appearing on the street due to the turning movements from Saginaw.

A cross-section of Figure 3.10 taken at a particular time produces a picture of the platoon and queue states like Figure 2.1. As more vehicles are added to the street the average vehicular velocity decreases, and this is reflected in the decreasing slopes of the leading and trailing edges of the platoons: at $t = 0$ the velocity is 57 ft/sec and at $t = 140$ it is 42 ft/sec.

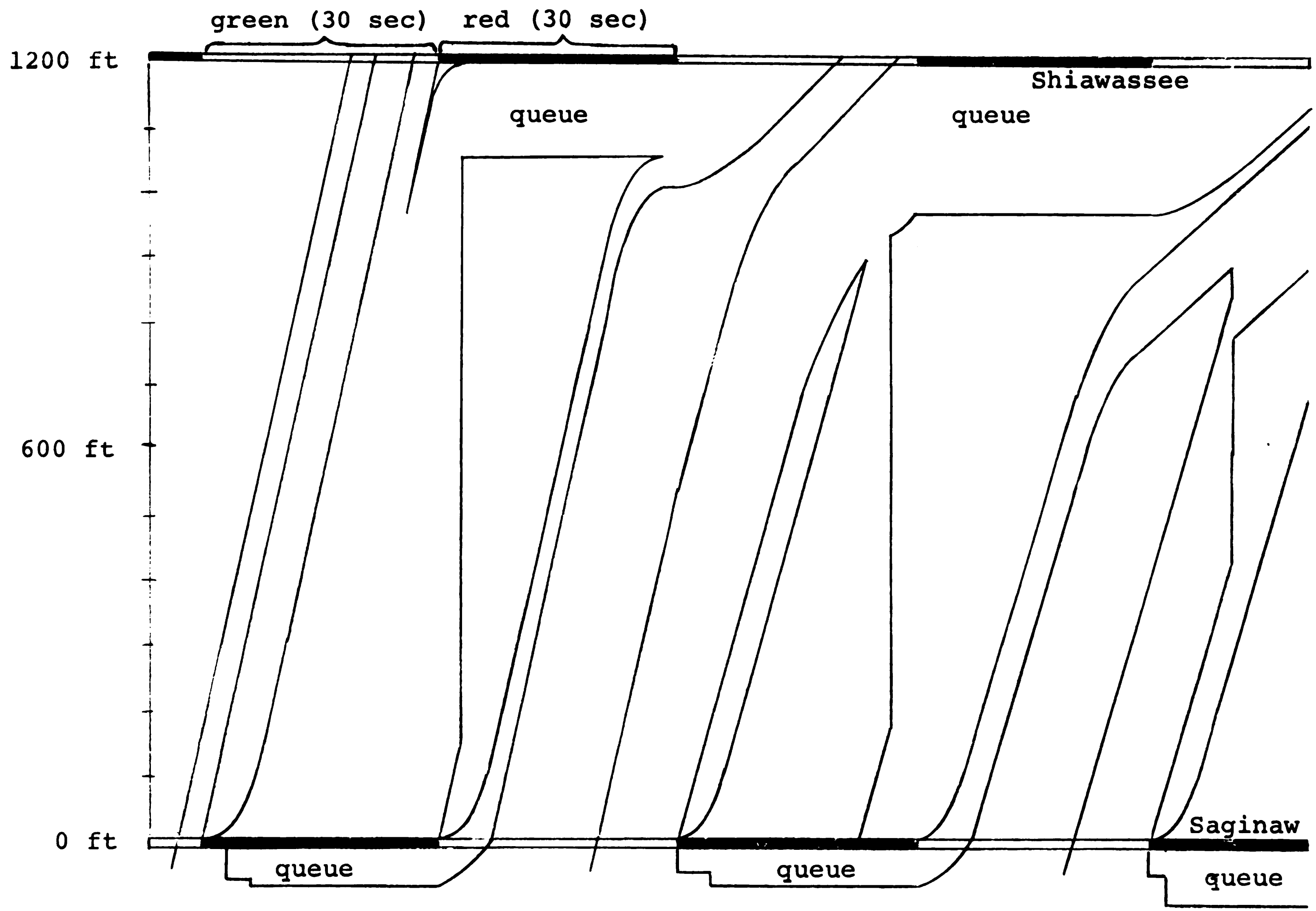


Figure 3.10 Platoon and queue behavior

Note that the ends of the queues, having zero velocity during the red times, are represented by lines with zero slope.

Since it is assumed that the whole platoon becomes part of a queue once its leading edge has reached the end of the queue, an impulse appears in the position of the platoon's trailing edge. Although this description is unrealistic, it is felt that no serious consequences result. First, the important variables are the average transit time per vehicle, which is determined for the leading edge, and the time headway between the two edges. If it were necessary to approximate the true behavior of the trailing edge, using the above data this would be easy. Secondly, since the queue serves the vehicles on a first come-first served basis, the vehicles from the end of the platoon (even though it is assumed that they arrive early) will not be served until their turn.

Macroscopic Study

The input rate to Capitol Avenue was assumed to have a normal value of 0.15 veh/sec/lane. After an initial period of 300 seconds, in which the system reached a more or less steady state, this input rate was increased by a factor of 3 for an interval of 100 seconds and then returned to normal. The average velocity for several sections of Capitol Avenue are plotted. The velocity on the input section (between Oakland and Saginaw) drops considerably due to the increased load. On the following section (between Saginaw and Shiawassee) the effect is less evident. This is probably due to the greater distance between intersections (1200 versus 800 feet). Between Shiawassee and Ionia the velocity demonstrates the same drastic response to the increased load as the input section.

Between Ottawa and Allegan the disturbance is still strongly felt, but the velocity in this region is also influenced considerably by the turning movements onto Capitol from Michigan. Note, for example, that the velocities in this region, even for normal operation, tend to be lower than observed elsewhere. Finally between Washtenaw and Kalamazoo the effect of the disturbance has been greatly diminished--only a slight depression is noted. The effect of the signal timing is one factor which influences the results--e.g., the rate of the disturbance propagation.

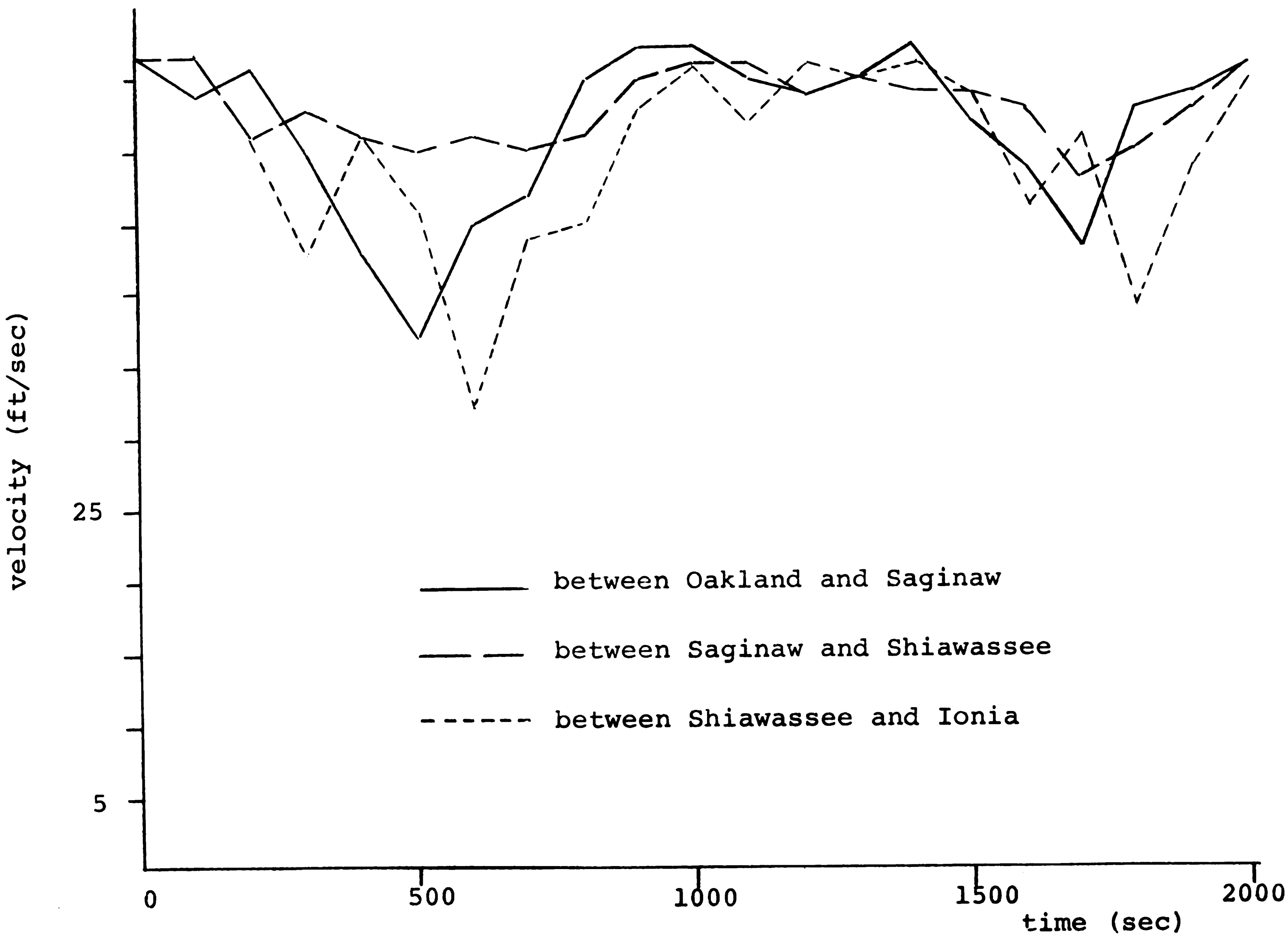


Figure 3.11. Velocity along Capitol Avenue

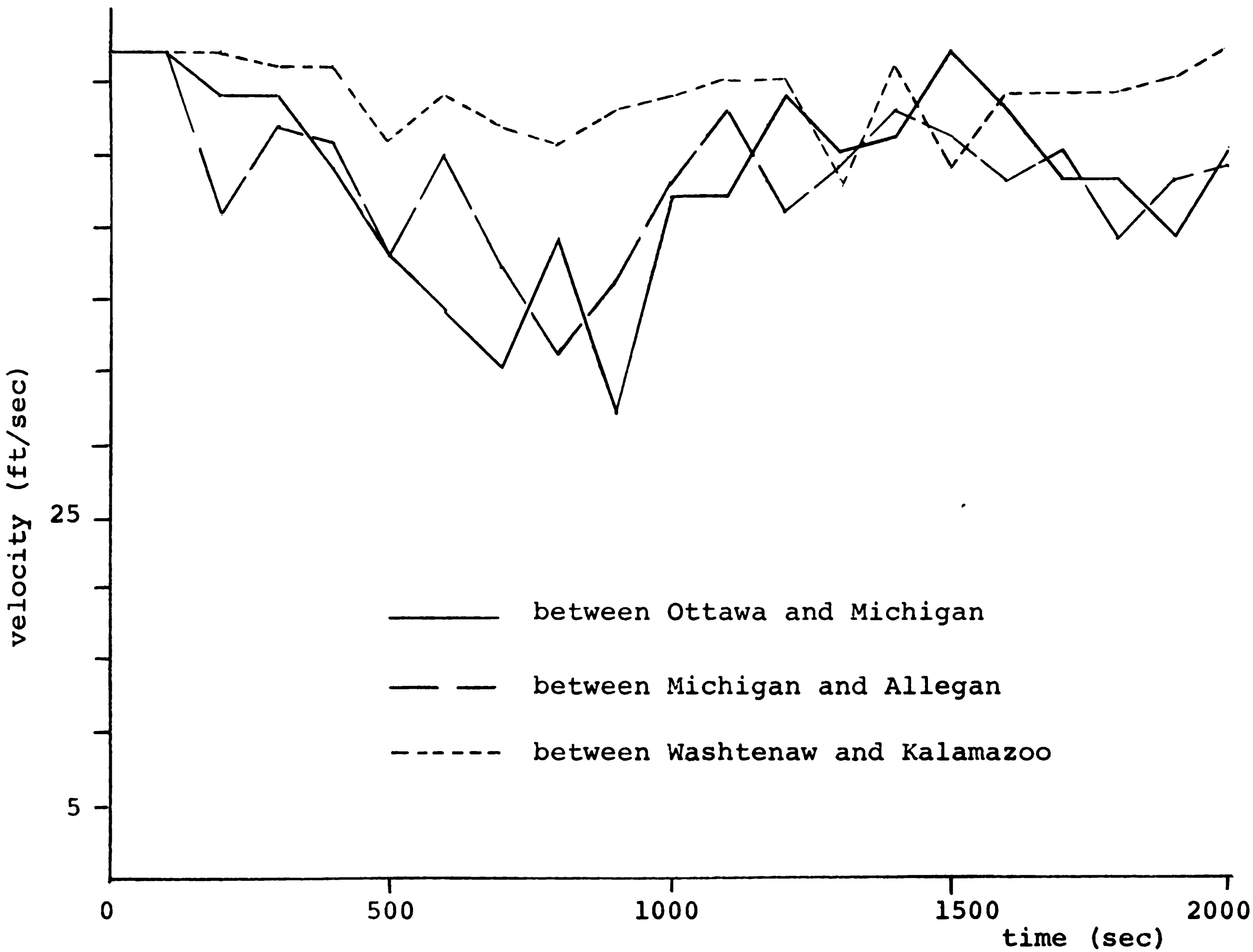


Figure 3.12. Velocity along Capitol Avenue continued

CHAPTER IV

CONTROL OF URBAN TRAFFIC SYSTEMS

Within a system of signalized intersections traffic flow is governed almost entirely by the traffic signals. By a judicious choice of signal variables one can minimize total vehicle travel time and time spent waiting in queues while maximizing vehicle counts. For medium density traffic conditions an effective control strategy based on minimal queue build-up is the establishment of progressions on the arteries. A progression can be defined as a steady state mode of operation which allows vehicles to travel at a specified velocity (the design speed) from one end of an artery to the other without stopping. The portion of the signal cycle for which this is possible is called the bandwidth.

A method developed by Morgan and Little (ML1) is a useful basis for determining a particular progression design on a two-way artery. With the introduction of two important theorems, it is possible to inspect a wide range of designs with a minimum of calculations.

While a prerequisite for smooth, efficient flow on an artery is the establishment of a progression, the possibility that queues develop always exists. Since the signal parameters which satisfy the design specifications of bandwidth and velocity are not unique, adjustments within the framework of a particular design can discourage the build-up of queues. In situations where queues have developed, it may be necessary to perturb the progression settings to eliminate them.

A discussion of traffic control would be incomplete if it did not touch on the special problems encountered on traffic grids. The extension to a grid of the control methods used on arteries is possible if some preliminary ground rules are established.

This chapter presents some of the important ideas pertaining to progressions. It also presents methods for establishing progressions on one-way and two-way streets as well as grids and, finally, some auxillary techniques for maintaining progressions in the face of disturbances.

4.1 Space-Time Diagrams and Traffic Signals

In studying the motion of a body in a one-dimensional space, a plot of its displacement from some reference point as a function of time is often helpful. On such a graph the velocity at any instant is given by the slope of the plot.

Engineers, studying the behavior of vehicle platoons on an artery, have long used space-time graphs as a visual aid. These graphs display the locations of the leading and trailing edges of the platoon as functions of time. The spatial length of the platoon is measured as l_p; its length in time (headway) is measured as t_p. The ratio of these variables is the platoon velocity v.

When the behavior of vehicular platoons in the two directions of a street are displayed on a space-time diagram at the same time, a complete picture of traffic flow on the street is obtained (GA2). However, the diagram's usefulness is limited to illustrating the flow for a given set of traffic signal parameters and design velocities. Under limited circumstances it may be possible to use the diagram for noting how a change in a parameter affects the flow. For example, in Figure 4.1 it can be seen that

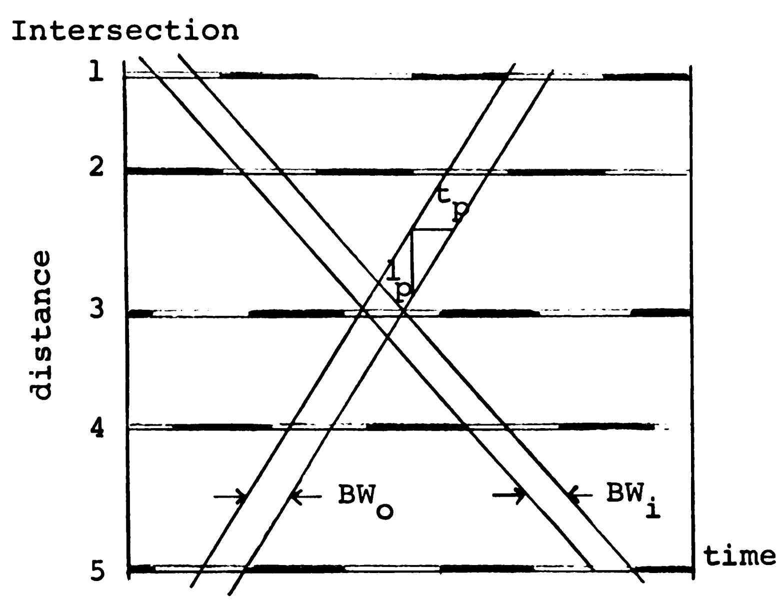


Figure 4.1. Typical space-time diagram.

if the timing of signal 5 is advanced slightly the size of the platoons which can be accommodated increases in both directions. Such observations produce only limited qualitative information for improving the system's flow.

It is appropriate to introduce a list of some basic traffic terms and the symbols which represent them. These terms occur frequently enough in what follows to warrant their inclusion.

CYC - common cycle of signals (sec); a cycle consists of successive red and green phases, the amber phase being relegated to the red or green.

G_i - green time of the ith signal (sec).

D_{ij} - distance between intersections i and j (ft).

v_{ij} - design velocity for vehicles traveling from intersection i to intersection j $(\frac{ft}{sec})$

T_{ij} - transit time for vehicles traveling from intersection i to intersection j (sec); $T_{ij} = \frac{D_{ij}}{v_{ij}}$.

BW - bandwidth, the measure of the band for which vehicles can travel the length of the artery without stopping (sec); the bandwidth-cycle ratio is $B = \frac{BW}{CYC}$.

β_{ij} - offset of signal j measured with respect to signal i (sec); β_{ij} is measured from the center of a green of signal i to the center of the first green of signal j such that $(0 \leq \beta_{ij} < CYC)$.

These parameters are illustrated on the space-time diagram of Figure 4.2. Often it is convenient to normalize the time parameters by dividing by "CYC." Thus, $g_i = \frac{G_i}{CYC}$, $0 \leq g_i < 1$; etc.

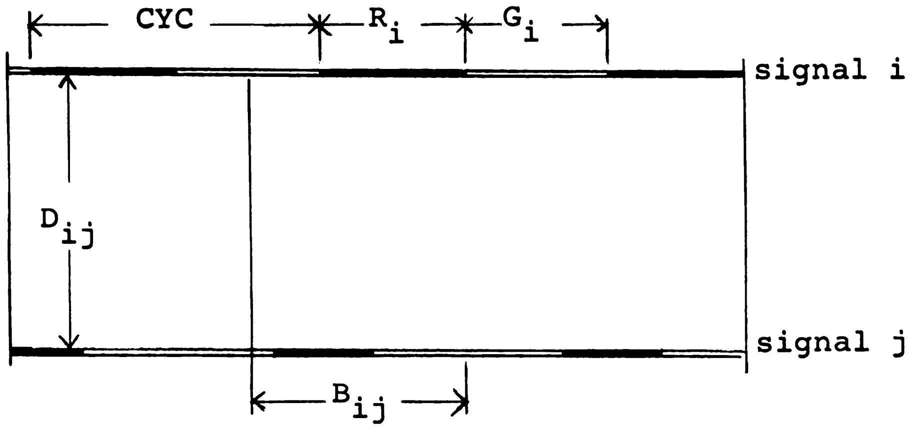


Figure 4.2. Traffic parameters defined.

4.2 Steady State Queuing

The problem of attaining an efficient traffic system and of maximizing flow is closely related to the problem of queuing. Therefore, a useful (though incomplete) measure of a control system's effectiveness is the total time that the vehicles spend waiting in queues.

As previously indicated determining vehicle behavior at the intersections of a large system using mainly statistical methods is nearly impossible. A study of steady state queue behavior, however, is a reasonable objective. For the purposes of the following discussion steady state implies that the vehicles are flowing continually through the intersections during the green phases and that the velocity is always assumed to be the same. A constraint implied by these assumptions is that the green phases of all signals are equal and the cycle length is the same for all signals.

It will be shown that the formation of a queue at an intersection is a function of the offset β of that intersection's signal measured with respect to the previous signal, the velocity v of the vehicles and the green time to cycle length ratio, g.

Used as an aid in the discussion of the queuing at a single intersection, Figure 4.3 is a space-time diagram for two intersections illustrating the vehicle flow between them. From the figure, it is evident that four cases exist depending on the values of g and β. The first two cases (a and b) correspond to situations where the trailing portion or the leading portion, respectively, of platoons leaving the first intersection encounters the red phase at the second intersection. In the third case (c) the entire platoon encounters the red phase, and in the last case (d) the center portion of a platoon arrives at the second intersection during the red phase.

For each case three new variables are defined. Vehicles arriving at the second intersection when the signal is red form a queue. The queue continues to grow until

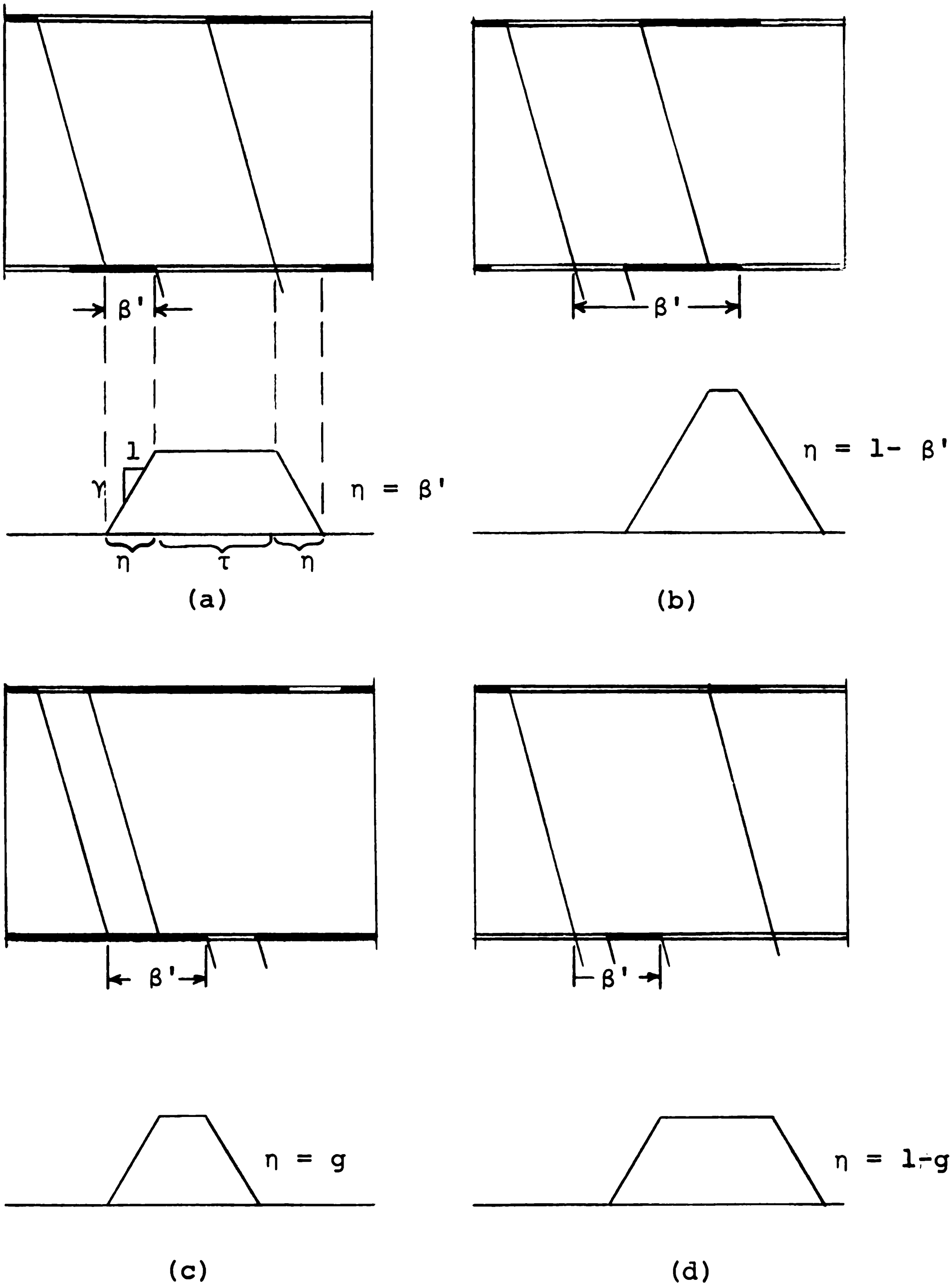


Figure 4.3. Steady state queuing at an intersection.

the signal changes to green (Figures 4.3 a, d) or until vehicles cease to arrive at its end (Figures 4.3 b, c). In either event, the variable η represents the time interval of queue formation.

If the signal turns green and if vehicles are still arriving at its end (Figures 4.3 a, d), the queue length remains constant until vehicles cease to arrive at its end. Alternatively, if vehicles cease to arrive at its end but the signal remains red (Figures 4.3 b, c), the queue length remains constant until the signal becomes green. In either event, the time interval that the queue exists with a constant (nonzero) length is τ.

Finally, the queue begins to shorten at the first instant when both the signal is green and no vehicles are arriving at its end. The time required to dissipate the queue is assumed to be the same as that required to form it, η.

The third variable defined is β'. Measured at the second intersection, β' is the time between the arrival of the first vehicle in the platoon and the start of the next green phase.

These variables are defined mathematically as

$$\eta = \min\,[\beta',\ 1 - \beta',\ g,\ 1 - g] \qquad (4.1)$$

$$\tau = |\beta' - g| \qquad (4.2)$$

$$\beta' = \beta - T + n\mathrm{CYC} \qquad (4.3)$$

where T is the transit time between the signals and n is an integer selected so that $0 \leq \beta' < CYC$. These relations are more apparent after examining Figure 4.3

In Figure 4.3 there is presented also a graph depicting the queue length behavior during a cycle. The average vehicle time spent waiting in a queue per cycle is proportional to the integral of queue length with respect to time over a cycle.

$$T_q = [(\eta\gamma)\frac{\eta}{2} + (\eta\gamma)\tau + (\eta\gamma)\frac{\eta}{2}] \cdot CYC^2$$

$$= [\eta^2 + \eta\tau]\ \gamma \cdot CYC^2 \qquad (4.4)$$

Substituting (4.1) and (4.2) into (4.4) results in

$$T_q = \min\ [\beta'^2 + \beta' \cdot |\beta' - g|,\ (1-\beta')^2 + (1-\beta') \cdot |\beta' - g|,$$
$$g^2 + g \cdot |\beta' - g|,\ (1-g)^2 + (1-g) \cdot |\beta' - g|]\gamma \cdot CYC^2 \qquad (4.5)$$

For a fixed value of g the minimum total wait per cycle of all vehicles is given by

$$T_{q(min)} = \min_{\beta'} \{ \min[\beta'^2 + \beta' \cdot |\beta' - g|,\ (1-\beta')^2$$
$$+ (1-\beta') \cdot |\beta' - g|\ ,\ g^2 + g \cdot |\beta' - g|,\ (1-g)^2$$
$$+ (1-g) \cdot |\beta' - g|] \}\ \gamma \cdot CYC^2 \qquad (4.6)$$

The value for β' which achieves this minimum is 0 so that $T_{q(min)} = 0$. This result is readily seen from equation

(4.6) or from Figure 4.4 which illustrates equation (4.5).

The conclusion that can be gleaned from the foregoing discussion is that the optimal selection of β occurs

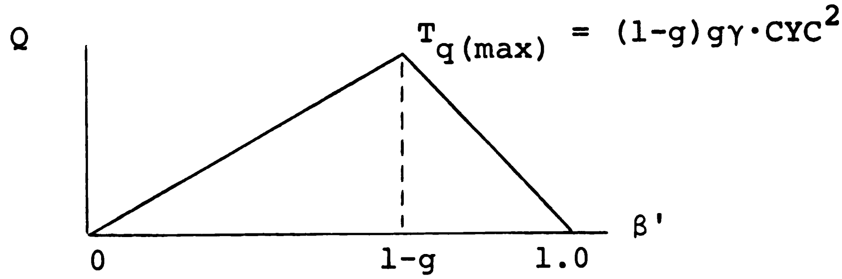


Figure 4.4. Queue integral as a function of offset β'.

when $\beta' = 0$, (i.e., β = T-nCYC). At this value no queues are formed and the flow γ is maximum. For all other choices queues are present and they have the detrimental effect of reducing γ. For the assumptions in this example it is clear that the best way to control an artery is to set the signal variables so that vehicles can travel from one end to the other without stopping somewhere in between to form a queue. For a real artery the same conclusion applies: it is desirable to set the signal variables in conjunction with the prevailing vehicle speed so the vehicles do not need to stop at intermediate intersections.

4.3 Progressions

A method attributed to Morgan and Little can be used to determine the signal settings for progressions on a two-way artery when the velocities are specified everywhere.

Briefly summarized, they have shown that with each offset value of either zero or one half of a cycle (that is, half cycle synchronization), there results inbound and outbound bandwidths which are equal. Among all the possible half cycle synchronizations there exists a combination which maximizes this equal bandwidth value.

The procedure suggested by them to achieve maximum equal bandwidths is basically simple: A bandwidth is established between the first and second signals. This bandwidth can be maximized by selecting the proper half cycle offset value (either 0 or $\frac{1}{2}$). The third signal can then be selected with either 0 or $\frac{1}{2}$ cycle offset so that the bandwidth is reduced as little as possible. The procedure is continued until all signals are considered. The procedure is then repeated for every pair of initial signals and from the resulting bandwidths the combination of offsets producing the maximum bandwidth is selected.

The method has not been fully exploited for design purposes. For example, they have shown in a corollary to the main presentation how a design having equal bandwidths for the two directions can be modified by reapportioning the total available bandwidth between the two directions. However, no sound criterion is given for this redistribution.

Another shortcoming is the lack of information regarding bandwidth as a function of velocity (or of any variable, for that matter). If the progression velocities are fixed precisely in advance, this information is not needed to determine a progression design. However, this is generally not true in a genuine design situation. The better approach is to consider all designs for a range of acceptable velocities and to select the one which provides the most bandwidth. Since the bandwidth is measurably affected by even small changes in velocity, it is worthwhile to have this information.

The simplest, yet most common, problem encountered on a two-way street is to establish a progression in each direction when only two velocities are specified, one for each direction of flow. A progression design exists for each point of the subset defined by

$$V^2 = \{(v_1, v_2): 0 < v_i \leq v_{max}\}$$

For this important case, the following theorem demonstrates how the bandwidth can be depicted as a function of a single variable, thereby simplifying the design problem.

Theorem 1. Between a pair of intersections i and j, if the bandwidths BW_1 and BW_2 are realizable for a design velocity pair (v_{ij}, v_{ji}), then these same bandwidths are realizable with the design velocities (v'_{ij}, v'_{ji}) where

$$\frac{1}{v_{ij}} + \frac{1}{v_{ji}} = \frac{1}{v'_{ij}} + \frac{1}{v'_{ji}} \tag{4.7}$$

Proof: Assume that the progression bands 1 and 2 are associated, respectively, with the design velocities v_{ij} and v_{ji} between the intersections i and j. The distance between the intersections is designated D_{ij} (or D_{ji}). β_{ij} is the offset value for which the original design is realized.

Transit times for vehicles in the two bands are

$$T_{ij} = \frac{D_{ij}}{v_{ij}} \tag{4.8}$$

$$T_{ji} = \frac{D_{ji}}{v_{ji}} \tag{4.9}$$

If the offset β_{ij} is altered so that

$$\beta'_{ij} = \beta_{ij} + \Delta T \tag{4.10}$$

new transit times can be defined

$$T'_{ij} = T_{ij} + \Delta T = \frac{D_{ij}}{v'_{ij}} \tag{4.11}$$

$$T'_{ji} = T_{ji} - \Delta T = \frac{D_{ji}}{v'_{ji}} \tag{4.12}$$

Thus

$$T_{ij} + T_{ji} = T'_{ij} + T'_{ji} \tag{4.13}$$

and

$$\frac{1}{v_{ij}} + \frac{1}{v_{ji}} = \frac{1}{v_{ij}^{T}} + \frac{1}{v_{ji}^{T}} \tag{4.14}$$

The change in transit time does not affect the relative time spacing of the leading and trailing edges of the progression bands, thus the bandwidths BW_i remain fixed.

With this result it is possible to define a new velocity v_e such that

$$\frac{1}{v_{ij}} + \frac{1}{v_{ji}} = \frac{2}{v_e} \tag{4.15}$$

where v_e is the value when the two velocities are equal. For the two-way street example cited above, the design problem is reduced to examining the equal bandwidth possibilities corresponding to the points of V where

$$V = \{v_e\colon\ 0 < v_e \leq v_{max}\}$$

This information can be set forth in a graph. The ordinate, bandwidth-cycle ratio, is the total available bandwidth which can be freely apportioned between the two directions, subject only to the constraint, that the bandwidth in either direction must not exceed the minimum green time, g_{min}. Similarly, each abscissa point (velocity-cycle product) represents a set of inbound and outbound velocities constrained only by equation (4.14). Once a point on the graph is selected and the individual velocities are fixed, one needs only to determine the cycle length to complete the design of the progression.

As an example of such a plot consider the artery used by Morgan and Little in their presentation: a two-way street having ten intersections with specified green phases. An immediate observation is that the graph of

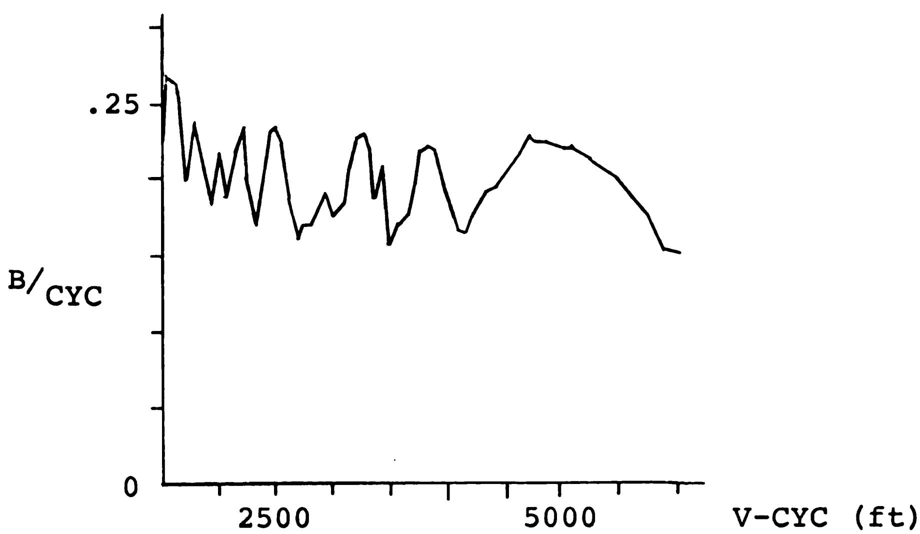


Figure 4.5. Bandwidth/cycle versus velocity-cycle.

Figure 4.5 is very erratic having no truly periodic components. Despite this lack of mathematical periodicity, however, most values of the function are repeated many times suggesting that the total bandwidth range may be realized over a relatively narrow velocity domain.

This fact is especially fortunate since it is likely that the choices of velocities and cycle length come from relatively narrow ranges. That is, the velocities may range from 20 to 80 ft/sec and the cycle length from 30 to 100 sec; thus the product may range from 600 to 8000 ft. However, low velocities are usually associated with long cycles and vice-versa. Under these circumstances, the abscissa interval of interest is more likely to be 2000 to 4000 ft. Over this interval all values from the bandwidth range are realized and only this part of the graph needs to be determined.

In the general problem the desired velocity along the street may not be constant. In this case, it is useful to divide the street into n segments, each segment having constant inbound and outbound velocities. For each point in the set V^{2n} there exists an equal bandwidth progression design where

$$V^{2n} = \{(v_1,\ldots,v_{2n}):\ 0<v_i\leq v_{max}\}$$

where

$$i = 1,\ldots,\ 2n$$

The previous theorem permits reducing the dimension of this set by a factor of two, thereby cutting the search time. A greater simplification for this n-segment street can be achieved with the result of the following theorem.

Theorem 2. If a street for which a progression is being designed is partitioned into n segments such that a different design velocity v_e^k prevails over each segment, the band is equivalent to one having a universal velocity v_e obtained by defining an equivalent length for each section such that

$$D_{ij}^e = D_{ij}\left(\frac{v_e}{v_e^k}\right) \tag{4.16}$$

Proof: The transit time for vehicles in any segment is given by

$$T_{ij} = \frac{D_{ij}}{v_e^k} \tag{4.17}$$

If the transit time is kept constant, then

$$T_{ij} = \frac{\alpha D_{ij}}{\alpha v_e^k} = \frac{D_{ij}^e}{v_e} \tag{4.18}$$

Thus

$$\alpha\, v_e^k = v_e \tag{4.19}$$

and

$$D_{ij}^e = \frac{v_e}{v_e^k} D_{ij} \tag{4.20}$$

The bandwidth remains unchanged with these transformations of velocity and length.

The preceding result is useful in determining the progression design for a street which may be partitioned into segments, each having distinct design velocities. Although the result does not reduce the dimension of the space to be searched, it allows a more systematic search to be made. It is possible to make a plot of available bandwidth as a function of the single velocity v_e. This, however, does not constitute the complete picture.

4.4 One-way Streets

Although the problems encountered when establishing a progression on a one-way street can be readily solved using the methods for two-way arteries, the wide usage of one-way streets, particularly in central business areas, justifies a separate discussion.

The specifications for a one-way street include the velocity over each arterial segment and the bandwidth. The bandwidth can be set equal to the minimum green time on the artery no matter what the specified velocities are. Morgan and Little's method could be used to determine the offset values which accommodate this specified design. However, these offsets can be determined more simply by the following relation.

$$\beta_{ij} = \frac{D_{ij}}{v_{ij}} - n\ CYC \qquad (4.21)$$

where n is selected so that $0 \leq \beta_{ij} < CYC$. Such a set of offsets locates the progression band centrally within each signal's green time.

Consider the one-way street progression of the space-time diagram of Figure 4.6 For simplicity a constant velocity is assumed along the street so that $v_{12} = v_{23} = \ldots = v$. Intersection 1 is the primary input for vehicles. However, at each intersection vehicles are injected onto the artery from cross-streets by turning movements (secondary inputs). A vehicle leaving intersection 1 at the

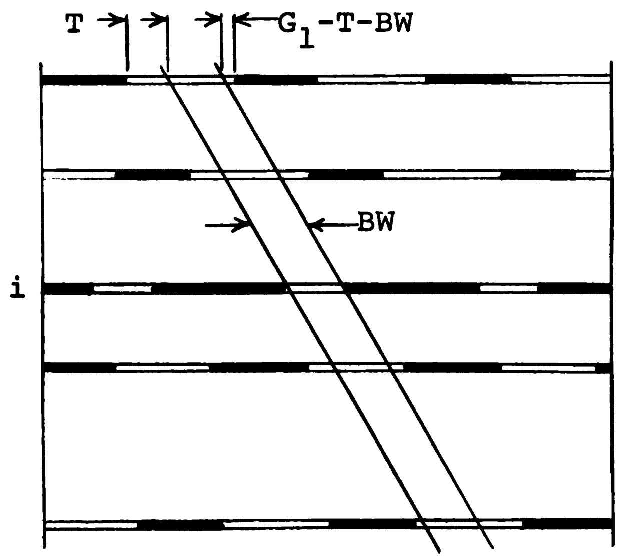


Figure 4.6. Progression bandwidth.

progression velocity v during the interval ($T \leq t \leq T + BW$) can travel the artery without stopping. The group of vehicles leaving during this opportune time is a progression platoon.

Vehicles within the system which are not part of a progression platoon result from one of the following phenomena.

1. Vehicles which enter the artery at the primary input at time t where ($0 \leq t < T$ or $T + BW < t \leq T_{G_1}$).

2. Vehicles which enter the artery at one of the secondary inputs. (These vehicles necessarily enter the artery during a red phase.)

3. Vehicles, which fail to maintain progression speed, falling away from the platoons.

These vehicles are (probably) stopped at one of the succeeding intersections (especially intersection i) forming queues. Unless these queues are dissipated before the platoons arrive, they interfere with the movement of the platoons. Under severe conditions, the queues cause the breakdown of the progression, and, for this reason, their effects must be minimized. Several steps can be taken to this end.

Light Density Conditions

The first intersection of an artery is the primary input for vehicles. Forcing vehicles from this intersection into progression platoons tends to minimize the number of vehicles which form queues. Therefore, by reducing the green time of this first signal so that it has the smallest green time of all signals, any vehicle entering the artery at this point is in a progression platoon and can travel to the other end without stopping.

This corrective action may be less desirable if there are a significant number of vehicles which could use the non-progression band and leave the artery before being queued. Figure 4.7 shows some of these non-platoon vehicles which leave the artery before reaching the critical signal i.

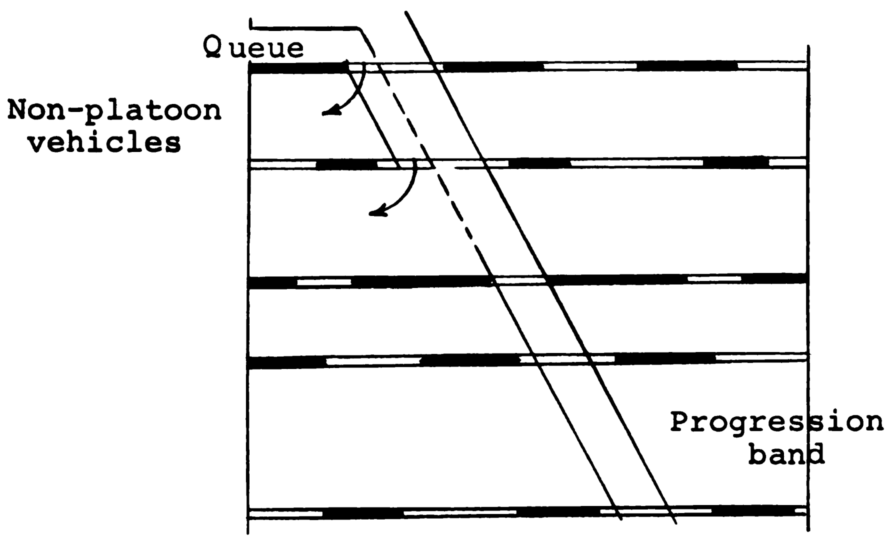


Figure 4.7. Non-platoon vehicles and queue on one-way artery.

The band occupies the entire green phase of the minimum green signal. For the other signals it is wise to distribute the excess green time so it occurs to the left of the band; that is, so it occurs earlier in time. (See Figure 4.8.) This shift in offset provides time for any

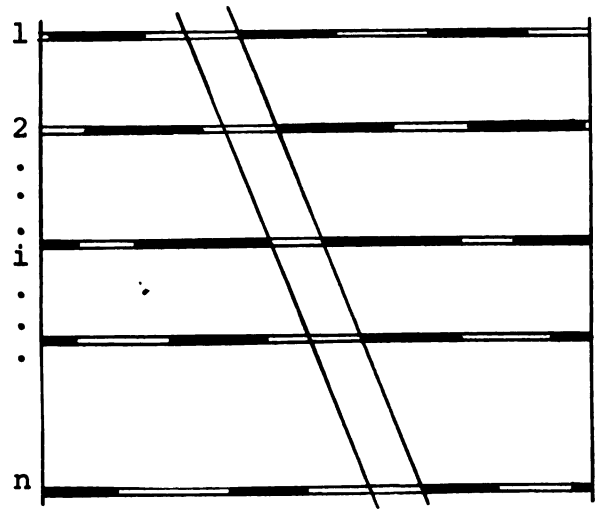


Figure 4.8. Progression with excess green distributed to left of band.

queue formed at the intersection to start moving before the arrival of the scheduled platoon, thereby minimizing the interference between queue and platoon. The new offset value is given by

$$\beta_{ij} = \frac{D_{ij}}{v_{ij}} + \frac{G_i - G_j}{2} - n\ \mathrm{CYC} \qquad (4.22)$$

where n is selected so that $0 \leq \beta_{ij} < \mathrm{CYC}$. Figure 4.9 illustrates how this relation is obtained.

At troublesome intersections (e.g., intersection i in Figure 4.7) the above measures are not sufficient to completely dissipate the queues. In such cases a transient control can achieve the desired result. If all the offsets are changed by the same amount the steady state settings of the signals remain unchanged. During the transient time, however, the queues are provided an opportunity to clear.

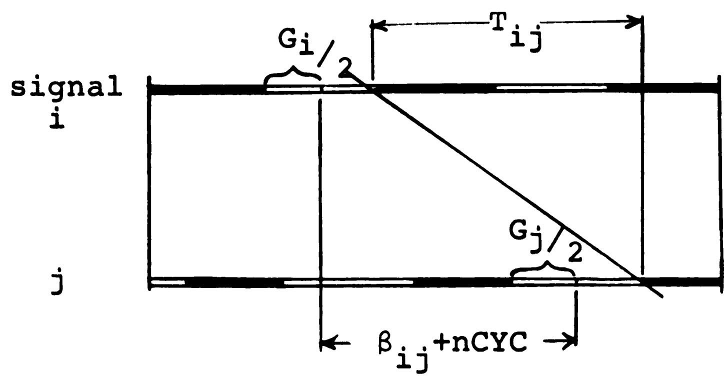


Figure 4.9. Illustration for β_{ij} determination.

The phasing of the signals should be performed in the following manner.

1. Increase, successively, the red times of signals k (k = 1, 2,..., i - 1) by α seconds.

2. Then increase successively the green times of signals k (k = i, i + 1,..., n) by α seconds.

This procedure increases temporarily the time for which the queue at intersection i can move while maintaining the usual number of vehicles entering at the primary input. Repeated intermittenly, this procedure helps to clear queues. Figure 4.10 illustrates this transient control.

Heavy Density Conditions

In a well-designed progression the timing of the signals is such that as a platoon approaches an intersection

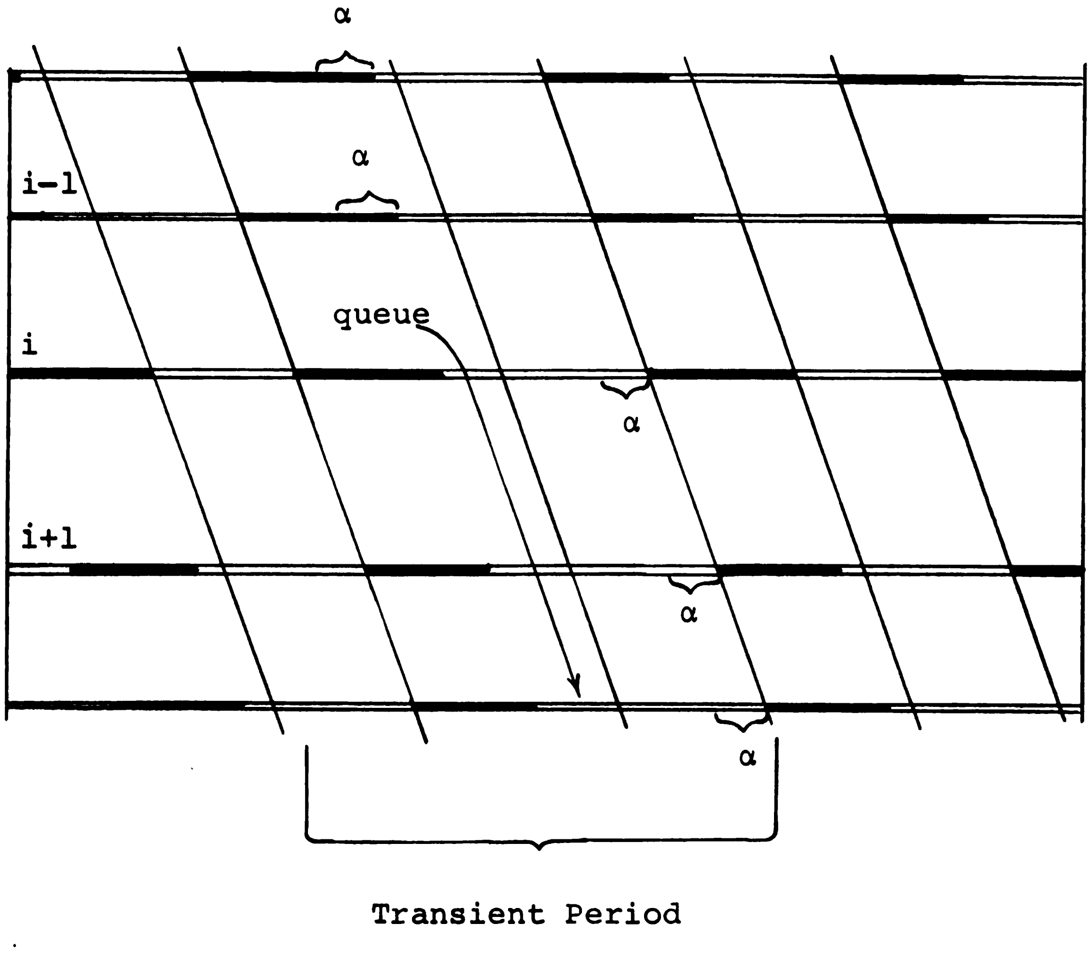


Figure 4.10. Clearing of queues by transient effects.

the signal turns green. Under heavy density conditions in which vehicles are queuing at the intersections a good design is difficult to maintain. When a platoon reaches the end of a queue it has for all purposes reached the intersection, and the signal should be turned green at this time. As a general policy, therefore, as the queue at an intersection increases, the offset of that intersection should decrease with respect to the preceding intersection. Equation (4.21) becomes

$$\beta_{ij} = \frac{D_{ij} - \alpha Q_{ij}}{v_{ij}} - n\ \mathrm{CYC},\ \alpha \geq 1.0 \qquad (4.23)$$

If α is selected greater than one, the queued vehicles have a chance to accelerate before the platoon arrives. Under extreme conditions where a queue is formed at every intersection and extends over the entire artery, the offsets should be reduced to zero so that all signals turn green simultaneously. Goodnuff discusses this problem in detail (GO1).

By increasing the green phase of all signals but the first, the vehicles on the artery have more time to pass through each intersection. At the same time, all the inputs, primary and secondary, are regulated (i.e., the number of vehicles permitted onto the artery is decreased).

Lengthening the cycle of all signals results in a lower progression speed, hopefully coinciding with the lower natural speed dictated by the heavier density conditions. (See Figure 2.2.) This policy is very effective and easy to implement on an artery which has progression settings. Figure 4.11 illustrates a progression for which the cycle is increased at t = T.

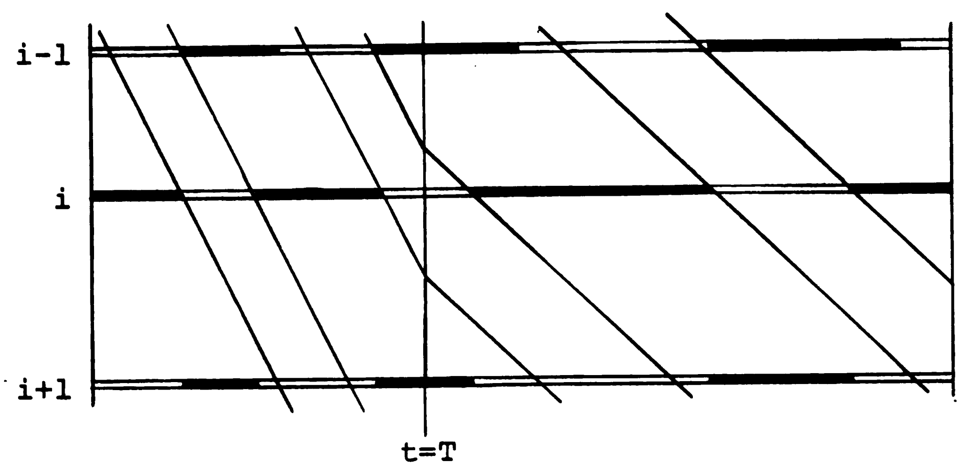


Figure 4.11. Effect of increasing cycle length.

4.5 Grids

The study of arterial traffic leads naturally to the study of arterial networks having at least one complete circuit. A grid, as such a system of intersecting arteries is called, operates most effectively when a progression is established on each of its arteries. The same techniques for establishing progressions on isolated arteries can be applied to the arteries of a grid, but there are constraints for the signal parameters which cannot be violated.

A grid can be visualized as a mesh of arteries having one or more internal circuits or loops. A simple grid loop may consist of four arterial sections formed into a closed path as in Figure 4.12. The following definitions are useful in establishing the important relation for the signal offsets around such a loop.

Let the underlined index refer to the green phase for north-south flows and the other index to the green phase for east-west flows. Thus $\beta_{i\underline{j}}$ is the relative offset of signal j with respect to signal i, measured from the center of the east-west green for the ith signal to the center of the north-south green of the jth signal.

The following relations are obvious from any space-time diagram.

$$\beta_{i\underline{i}} = \frac{CYC}{2} \tag{4.24}$$

$$\beta_{ij} = \beta_{ik} + \beta_{kj} + \alpha\ CYC,\ \alpha = -1,0 \tag{4.25}$$

$$\beta_{ij} = \alpha CYC - \beta_{ji},\ \alpha = 0,1 \tag{4.26}$$

Since offsets are positive fractions of a cycle, the α's are necessary to maintain this status.

The offsets around a closed loop sum to an integer number of cycles. To show this, the following relations can be used. (See Figure 4.12.)

$$\beta_{13} = \beta_{12} + \beta_{2\underline{2}} + \beta_{\underline{23}} + \beta_{\underline{3}3} + \alpha_1 CYC,\ \alpha_1 = -1,-2 \tag{4.27}$$

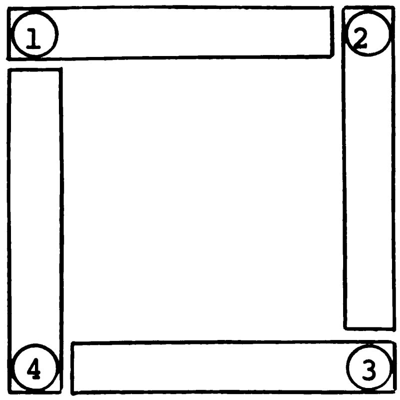


Figure 4.12. A grid loop model.

and

$$\beta_{13} = \beta_{1\underline{1}} + \beta_{1\underline{4}} + \beta_{\underline{4}4} + \beta_{43} + \alpha_2 \mathrm{CYC}, \quad \alpha_2 = -1,-2 \qquad (4.28)$$

Using the relations (4.24-4.26) above and equating (4.27) and (4.28) the following can be stated.

$$\beta_{12} + \beta_{\underline{23}} = \beta_{\underline{14}} + \beta_{43} + (\alpha_2 - \alpha_1)\mathrm{CYC} \qquad (4.29)$$

$$\beta_{12} + \beta_{\underline{2}2} + \beta_{23} + \beta_{3\underline{3}} - \mathrm{CYC} = \beta_{\underline{1}1} + \beta_{14} + \beta_{4\underline{4}} - \mathrm{CYC}$$
$$+ \beta_{43} + (\alpha_2 - \alpha_1)\mathrm{CYC} \qquad (4.30)$$

$$\beta_{12} + \beta_{23} + \beta_{34} + \beta_{41} = (\alpha_3 + \alpha_2 - \alpha_1)\cdot\mathrm{CYC},$$
$$\alpha_3 = 0,1,2 \qquad (4.31)$$

Since the β's are positive, $(\alpha_3 + \alpha_2 - \alpha_1)$ must equal (0, 1, 2, or 3).

By using Morgan and Little's method iteratively it may be possible to establish progressions on a grid, which satisfy the geometrical constraints of the arteries and equation (4.31) for every closed grid loop.

It was noted earlier that on an arterial more than one design could be realized. The designs that are possible for a grid are even more varied. One may seek a design which establishes a progression on every part of the grid. Theoretically this is possible but it may require an enormous amount of computation time and ultimately result in very narrow bandwidths. It is reasonable, therefore, to establish less restrictive objectives for a grid and to devise methods to achieve them.

One effective technique is to divide the procedure into two stages. In the first, the grid is broken into simple subsystems of single grid loops and arteries. Complete progressions can be established on these pieces using material presented previously. In the second stage, when the system is re-joined, one may not be able to maintain the progressions established; however, one may be able to minimize the delay on each artery by a compromise shift of offsets at the tie points.

Consider the hypothetical grid consisting of three north-south arteries crossing three east-west ones. Assume that in order of demand priority, the highest is labeled A and so on to the lowest which is labeled F. This grid is broken into two subsystems as indicated in

Figure 4.13. For the first stage of the design the procedures presented previously produce progressions on each of the arteries. In the second stage the system is reassembled an artery at a time beginning with A. The offsets on A, B, and C remain unchanged after assembly. An

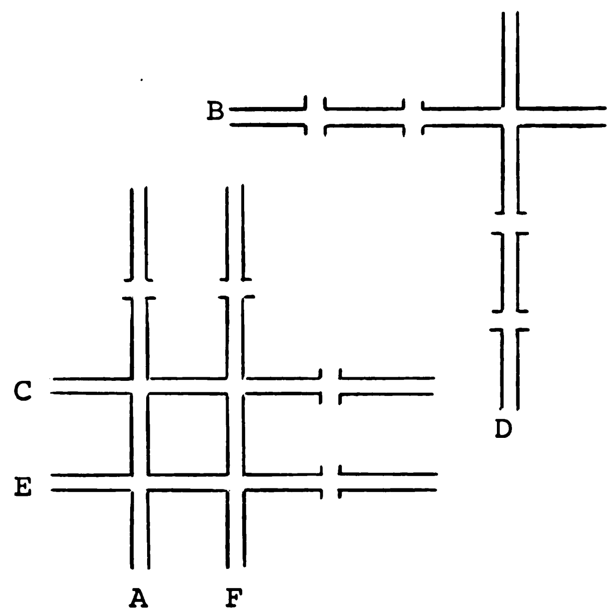


Figure 4.13. Subdivided grid.

interruption in the progression on D may result where D crosses C. This is due to the fact that the progression on C has already fixed the offset of the signal at the intersection. Most likely, the progression established on D fixed this offset at some other value. Since only one offset is possible the progression on either artery (or both) must be disrupted at this intersection. Similarly one disruption results in the progression on E where E crosses D and on F where it crosses B. If any of these arteries are one-way arteries or two-way arteries with traffic flow predominantly in one direction, it may be desirable to have these disruptions occur either upstream or downstream depending on the particular arterial traffic conditions. This would dictate to some extent how the grid is sub-divided in the first stage.

Fortuitous geometries of certain grids having a large number of one-way streets make the problem of establishing progressions less formidable. For very regular geometries it is always possible to satisfy equation (4.31) using zero, half-cycle or quarter cycle signal offsets. However, even with a good choice of cycle length, the service on the established progressions may be low in quality.

4.6 Example

As a further illustration of the diverse applications of the simulation model developed in the preceding chapters a simple control problem is considered. Within one section of the computer program it is possible to adjust the timing of the traffic signals on a one-way street to obtain a progression. The basis for these settings is equation 4.22

Along Walnut Street the signal offsets were adjusted initially for a progression velocity of 40 ft/sec, and these were not changed during the run. On Pine Street, which is parallel to Walnut, the offsets were set arbitrarily initially but during the run were intermittently (every 60 seconds) adjusted by the program so that a progression was obtained. The input rate for each street is 0.25 vehicles per second. Figure 4.14 depicts the input component velocities of these two streets. Succeeding figures (4.15-4.19) illustrate successive downstream component velocities. The average steady state velocity is 9 per cent higher when the signal settings are adjusted (36 vs 33 ft/sec).

One notes, however, that the velocities observed along Pine are more susceptible to oscillations. This illustrates a serious problem which must be dealt with in the future: In an attempt to attain a certain steady state control strategy the signals have to be changed. During the transient period resulting from this change the state of the system may change enough that a different steady state strategy is called for. This stability problem has not been adequately considered in the program.

In Figure 4.20 the output as a function of time is plotted. The steady state outputs (determined from the slope of the linear region) are 650 and 595 veh/hr/lane for Pine and Walnut, respectively. The output is improved by 9.2 per cent when the adjustments are made. In Figure 4.21 the accumulated time spent by all vehicles in a queue is shown. More time (7 per cent) is spent waiting in queues when no adjustments are made. The steady state rates are 41.4 and 44.3 veh-sec/sec for Pine and Walnut, respectively. (It seems reasonable to expect that a more sophisticated adjustment of the settings could result in a reduction of the "41.4" figure.)

For an individual vehicle the improvement is more pronounced. In the steady state the average vehicle must spend a total of 115 and 135 sec in queues while traveling the lengths of Pine and Walnut, respectively--a 17.4 per cent improvement with the adjustment. The average total trip time on Walnut is 12 per cent greater.

The conclusion that can be drawn from this example is obvious. The better the progression matches the conditions existing on the artery the smoother will be the traffic flow. The velocity will be increased, the wait time decreased and the output increased. Although one cannot determine a synchronization which will be good for all conditions, any synchronization is better than a random setting of signals. On the other hand, stability becomes a serious problem when corrections are attempted.

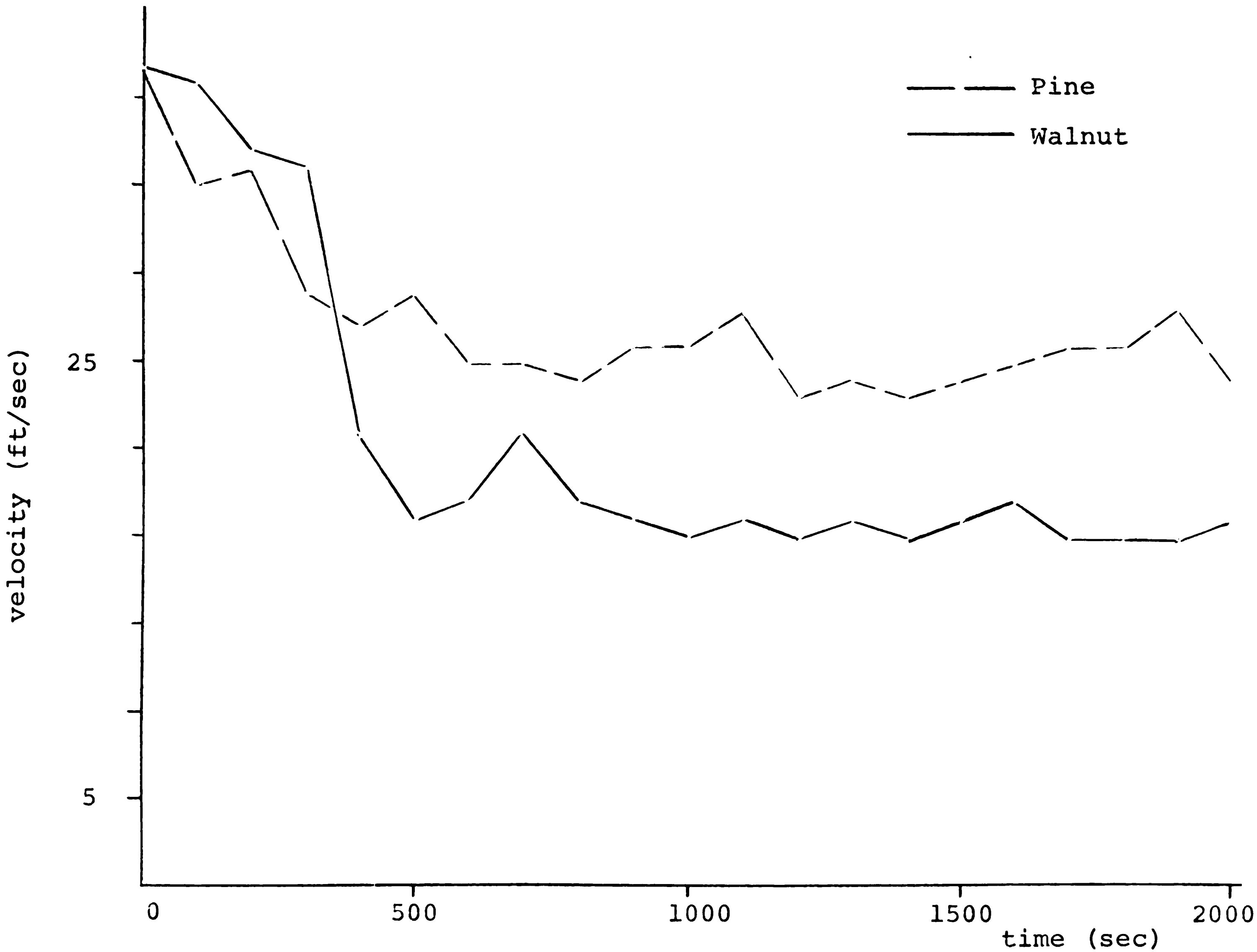


Figure 4.14 Input component velocity.

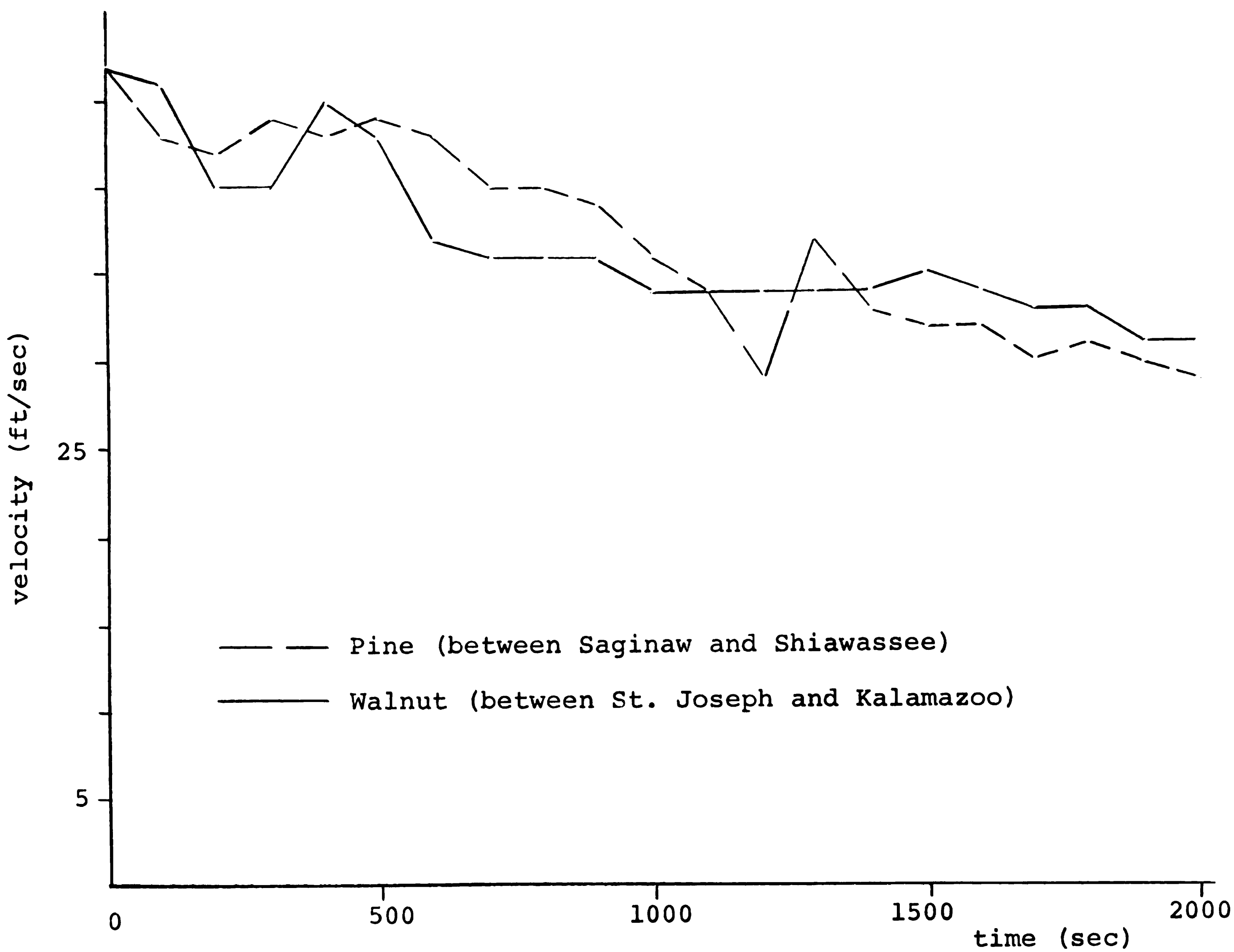


Figure 4.15 Downstream velocity.

Figure 4.16 Downstream velocity continued.

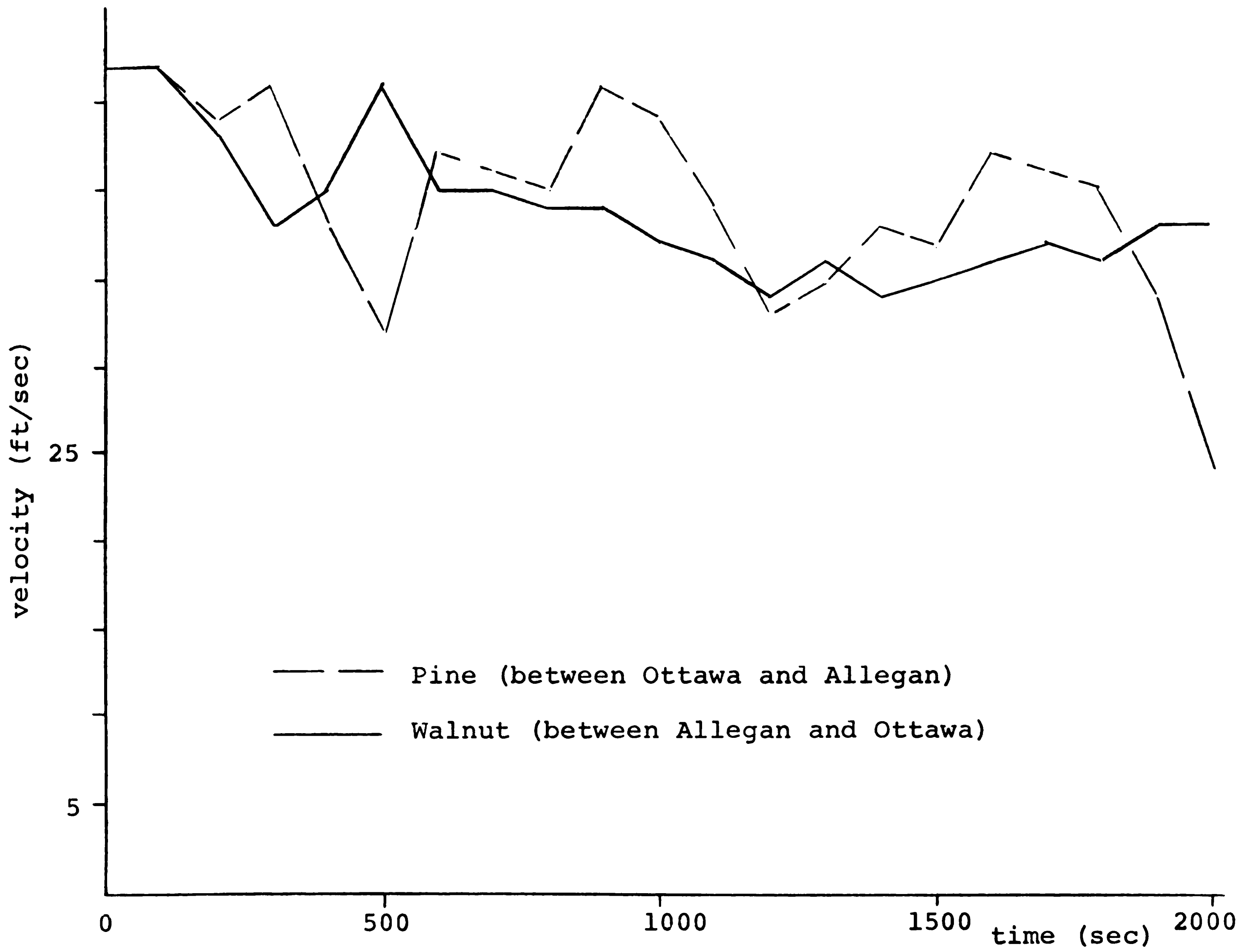


Figure 4.17 Downstream velocity continued.

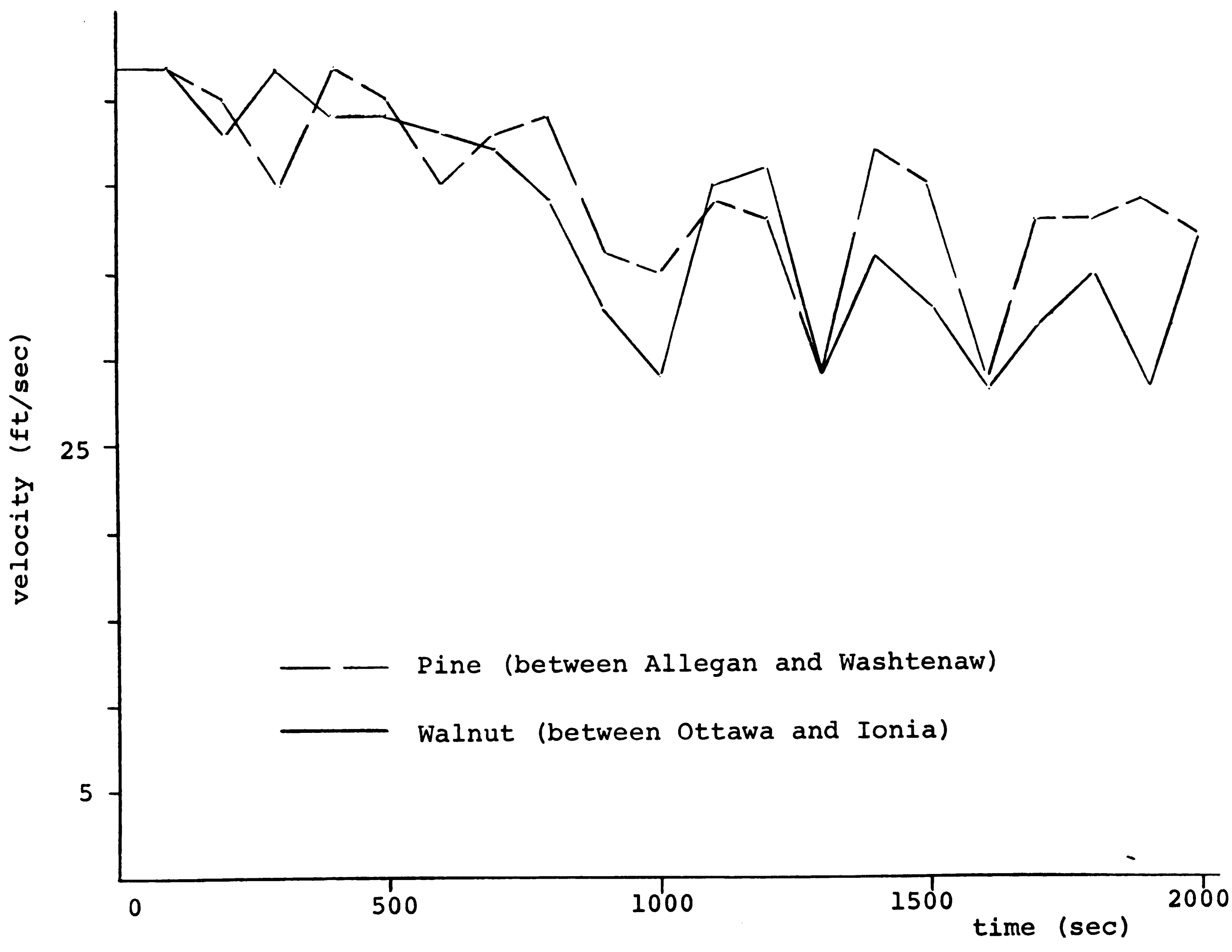


Figure 4.18 Downstream velocity continued.

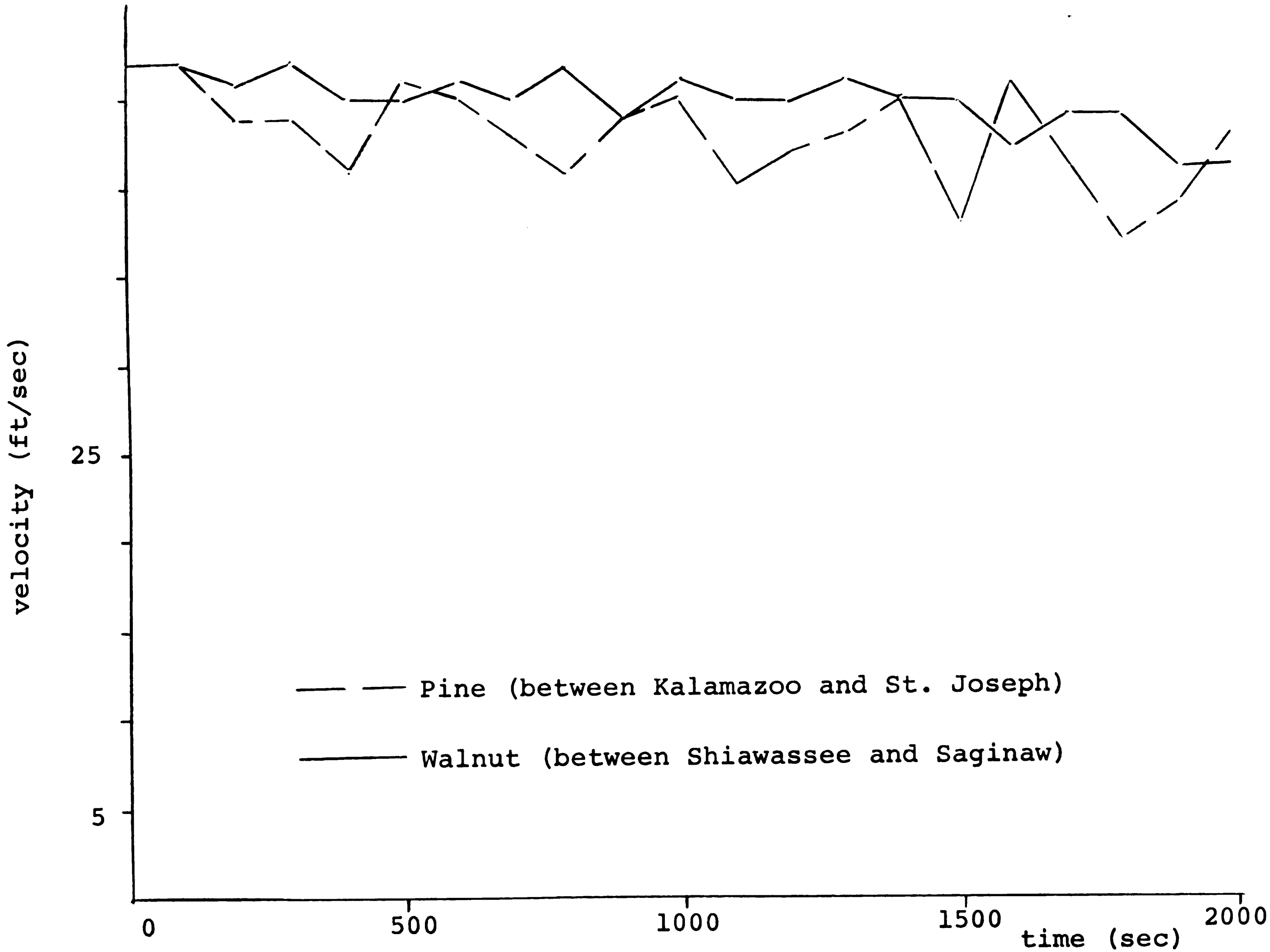


Figure 4.19 Downstream velocity continued.

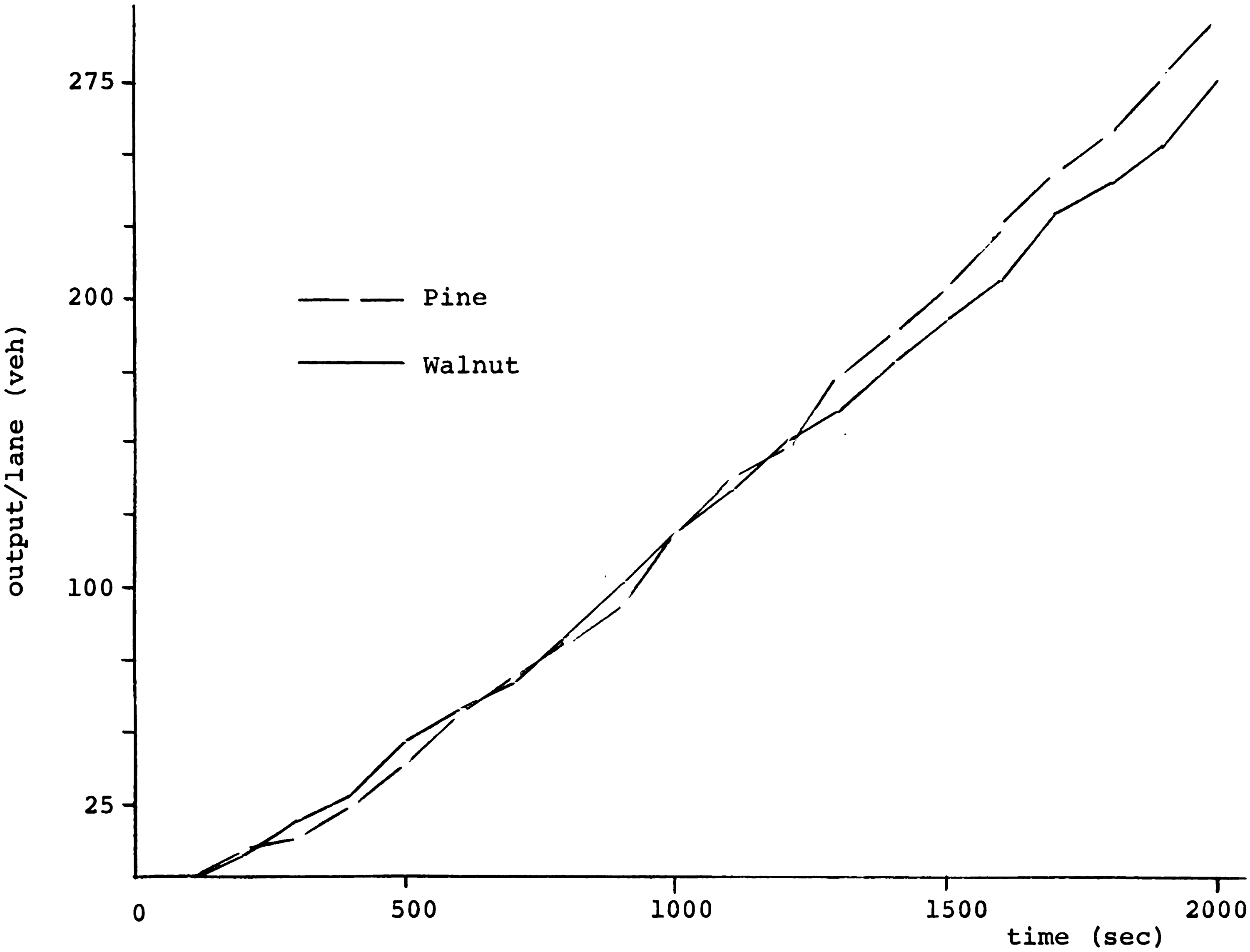
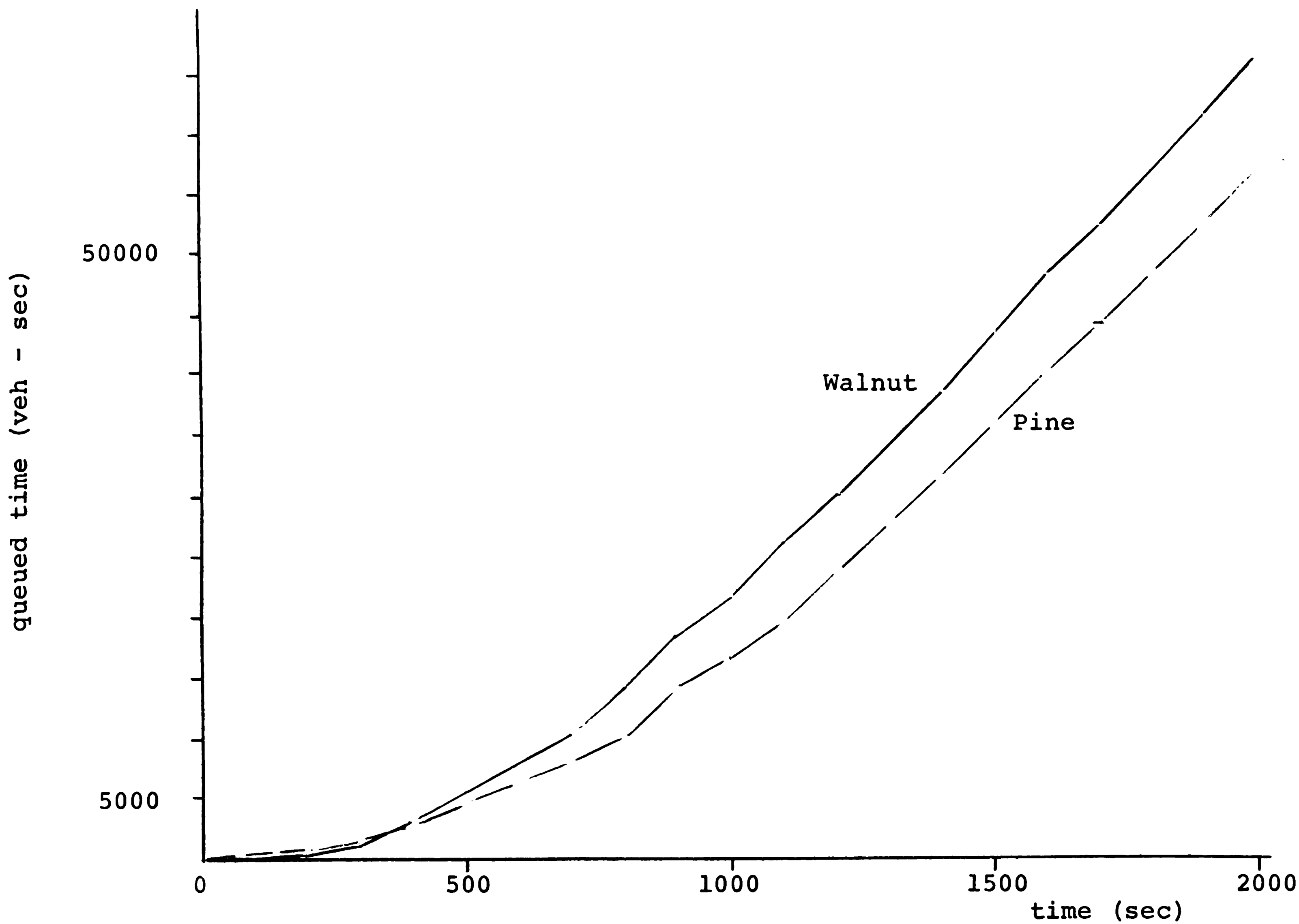


Figure 4.20 Output.



Figures 4.21 Time spent in queues.

CHAPTER V

OPTIMAL PROGRESSION DESIGN

Although the theory of optimal control is a fairly well-established one, its application to real systems is often severely limited. Many times the equations for the system are not known with a great degree of accuracy, and without an accurate description nothing like an optimal control is possible. On the other hand, if all the facts were accounted for, by sheer magnitude the problem of optimality for certain (large, nonlinear) systems would overwhelm even the most ambitious engineer.

It would be unfair to generalize by concluding that an optimal control for a traffic system cannot be devised for (1) a lack of knowledge and (2) an overabundance and complexity of variables. Sufficient evidence indicates that a traffic system can be efficiently controlled using a progression. A reasonable next step is to select the progression design in a manner which best satisfies the demands of the artery. This chapter sets forth a practical method for arriving at this optimal design.

5.1 Preliminary Remarks

The specifications for a progression design include the bandwidths and velocities over every segment of the artery. If this design is achieved without regard to the prevailing traffic conditions, it is likely to be inadequate. For example, the velocities cannot be arbitrarily specified since they are closely correlated to the number of vehicles demanding use of the artery. It is necessary and reasonable, therefore, to include this demand somewhere in the procedure.

In the discussion to follow a demand, speaking quantitatively, is a measure of the number of vehicles per unit of time desiring the use of an artery. Similarly, a flow rate is a measure of the number of vehicles per unit time actually using an artery.

For a one-way street it may be possible to find a progression design which handles a maximum demand at a good average velocity. For a two-way artery on which neither flow demand is negligible, selecting a good design is more difficult. There are many combinations of design parameters (bandwidths and velocities) to try and many that satisfy the given traffic conditions. The problem of selecting a best design is further complicated when the demand does not remain constant along the artery.

A space-time diagram is useless in this instance.

While it illustrates the structure of a particular design, it fails to show if the system is effectively meeting the flow demands. Likewise, while the methods of Morgan and Little are useful for determining the signal settings for a two-way street progression, they too fail to give an indication to the user as to which sets of parameters satisfy the flow demands on the artery.

In fact present methods for establishing a progression on two-way streets avoid this question almost completely. They (1) a priori specify the velocities based on gross observations of the traffic, and then (2) determine the bandwidths that these velocities produce. For the two-way street either the bandwidths in the two directions are made equal and maximum or one is maximized subject to the restriction that the other does not drop below a pre-assigned minimum. (This latter alternative hinges on the well-established fact that an increase in the bandwidth of one direction is usually accompanied by a decrease in the bandwidth of the other direction.) In no way, however, do these techniques provide an answer to the optimal progression design problem. Yet very little, if any, work on progression design uses more.

A progression that is optimally adjusted to the prevailing traffic demands must be implemented with traffic signals which can be automatically changed by a central digital controller. Initial steps toward a central traffic control have already been taken in Toronto, San Jose, and Witchita Falls (HU1, SJ1, CA1) with significant gains reported.

5.2 The Mathematical Model

For a progression to operate effectively no congestion can occur and the demand everywhere must be matched exactly by the flow rate. To achieve this when the demand does not remain constant complicates the problem.

Often the ends of an artery do not comprise the only vehicular entry and exit points on the artery. A large number of vehicles may enter at one or more internal intersections, and possibly many vehicles leave the artery before reaching its end. As a result of these turning movements both onto and off the artery or due to a change in the number of traffic lanes, the demand along each direction of flow may not remain constant. An excellent approximation is that this demand varies in a stepwise fashion along the artery's length. It is natural, therefore, to assume each point (intersection) where the demand changes significantly represents a boundary between segments having unequal demands for one or both directions. Thus an artery is, for the present purposes, divided into n segments if at n-1 internal intersections there is a change in either the inbound or outbound demand.

The first point of concern is to determine the flow rate that a given progression can sustain. The number of vehicles per unit of time that can be served by the street depends not only on the bandwidth of the progression but also on the density of the vehicles in the band. The relationship between average vehicular velocity and density has already been presented in Chapter II. Greenshields and others have established by theoretical and experimental studies that for a single lane of traffic there is an almost parabolic relationship between the flow rate and the lane occupancy as shown in Figure 5.1 (GR1). The corresponding speed of the traffic

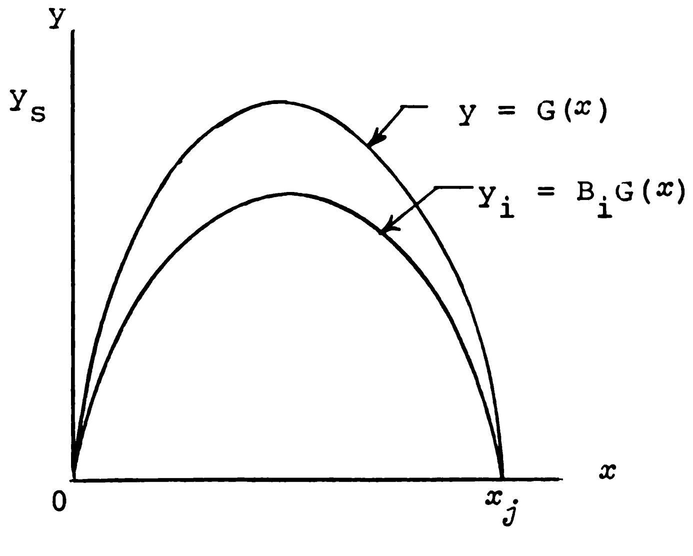


Figure 5.1. Flow-density characteristic based on Greenshields' linear model.

stream varies with occupancy as shown in Figure 5.2.

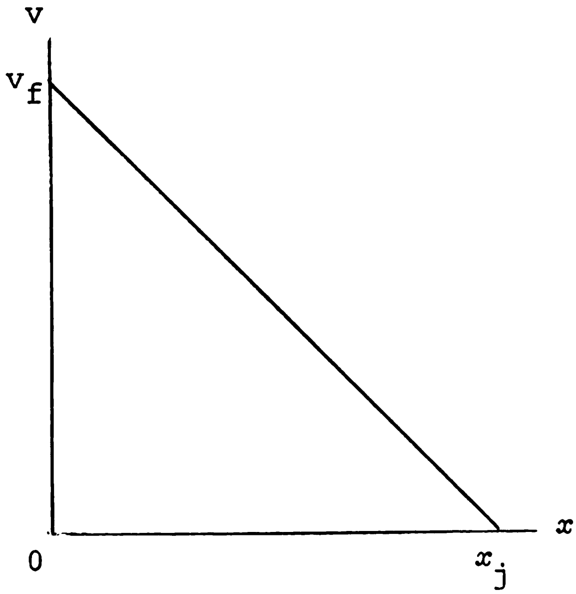


Figure 5.2. Velocity-density characteristic.

Inasmuch as the same basic phenomena governing speed and density are effective in a platoon as in the continuous stream of vehicles, the flow-occupancy characteristic for the progression system is given by the product of the continuous stream's characteristic and the bandwidth to cycle length ratio. Figure 5.1 illustrates both characteristics. The continuous stream flow rate is given by $y_s = G(x)$, the lane occupancy by x and the ratio of bandwidth and cycle length by B. There is a unique graph for each distinct bandwidth: $y_i = B_i\ G(x)$.

In a continuous steady state traffic stream the vehicles adjust their speed so that the flow rate matches the demand exactly, provided the capacity of the traffic lane is not exceeded. If the demand exceeds the capacity of the street, congestion quickly occurs upstream. Below the saturation level, an increase in demand is accommodated by a lower individual vehicle speed and higher lane occupancy, as demonstrated in Figures 5.1 and 5.2.

Analogously, within a progression system the speed should be selected so that the service flow rate always matches the demand and so that the platoons within the band are filled with vehicles at the density dictated by the speed-occupancy characteristic. Setting the progression speed higher than the value for which demand and service rate are balanced invites queuing and congestion, while setting it lower implies a needlessly lower quality of service.

If the demands on each part of the partitioned two--way street do not exceed the capacity then many combinations of speeds and bandwidths for a progression design are possible. The problem is to determine the unique combination which offers the highest quality of service.

In the analysis which follows, Greenshields' traffic model

$$y = 4y_{sm} \left(\frac{x}{x_j}\right) \left(1 - \frac{x}{x_j}\right) \tag{5.1}$$

$$v = v_f \left(1 - \frac{x}{x_j}\right) \tag{5.2}$$

is used where y_{sm} is the peak flow rate, x_j is the jam lane occupancy and v_f is the free speed. Comparable results may be obtained using other models which have been proposed.

If x is eliminated from (5.1) and (5.2) the result is

$$y = 4y_{sm} \left(1 - \frac{v}{v_f}\right) \left(\frac{v}{v_f}\right) \tag{5.3}$$

Applying the principle that bandwidths and progression speeds should be selected such that service rate equals demand on each arterial segment, requires

$$y_{2k-1} = 4B_{2k-1}\, y_{sm} \left(1 - \frac{v_{2k-1}}{v_f}\right) \left(\frac{v_{2k-1}}{v_f}\right), \quad k = 1,\ldots,n \tag{5.4}$$

and

$$y_{2k} = 4B_{2k}\, y_{sm} \left(1 - \frac{v_{2k}}{v_f}\right) \left(\frac{v_{2k}}{v_f}\right), \quad k = 1,\ldots,n \tag{5.5}$$

where y_{2k-1} and y_{2k} are the inbound and outbound demands and v_{2k-1} and v_{2k} are the progression speeds on the kth segment.

5.3 The Optimization

A reasonable expectation for the progression is that it satisfy the demands of the artery while minimizing the total vehicle-hours of travel time in the system.

A measure of total vehicle-hours of travel time under steady state operation is

$$F = \sum_{k=1}^{n} \left(\frac{y_{2k-1}}{v_{2k-1}} D_k + \frac{y_{2k}}{v_{2k}} D_k \right) \tag{5.6}$$

Here y is a measure of the number of vehicles served in an interval of time. The individual trip time over each segment of the artery is given by D/v, the ratio of the segment length and steady state velocity. The function F must be minimized while taking into account the constraints of realizability of the bandwidths and progression speeds selected.

Before proceding with the optimization process, the relation established in Chapter IV, equation (4.16) is recalled.

$$\frac{1}{v_{2k-1}} + \frac{1}{v_{2k}} = \frac{2}{v_{ek}} \qquad k = 1,\ldots,n \tag{5.7}$$

Within each segment the progression speeds can vary within this constraint without changing the bandwidths of that segment. Note that v_{ek} is the progression speed which would prevail if v_{2k-1} and v_{2k} are equal. Inspection of equation (5.7) indicates that allowing v_{ek} to increase without disturbing bandwidths results in higher permissible values of v_{2k-1} and v_{2k} .

If equations (5.4) and (5.5) are substituted into (5.6), the result is

$$F = \sum_{k=1}^{n} \left\{ \frac{4B_{2k-1}\, y_{sm}\, D_k}{v_f} \left(1 - \frac{v_{2k-1}}{v_f}\right) + \frac{4B_{2k}\, y_{sm}\, D_k}{v_f} \left(1 - \frac{v_{2k}}{v_f}\right) \right\} \tag{5.8}$$

If F is minimized by the method of Lagrange multipliers subject to the constraint of (5.7), the optimum occurs when

$$\frac{\partial}{\partial v_{2k-1}} \left(F + \sum_{j=1}^{n} \mu_j \left[\frac{1}{v_{2j-1}} + \frac{1}{v_{2j}} - \frac{2}{v_{ej}} \right] \right) = 0 \quad k = 1,\ldots,n \tag{5.9}$$

$$\frac{\partial}{\partial v_{2k}} \left(F + \sum_{j=1}^{n} \mu_j \left[\frac{1}{v_{2j-1}} + \frac{1}{v_{2j}} - \frac{2}{v_{ej}} \right] \right) = 0 \quad k = 1,\ldots,n \tag{5.10}$$

where μ is a Lagrange multiplier.

The simultaneous solution of (5.9) and (5.10) after substitution of (5.8) results in

$$\left(\frac{v_{2k-1}}{v_{2k}} \right)^2 = \frac{B_{2k}}{B_{2k-1}} \qquad k = 1,\ldots,n \tag{5.11}$$

To be sure that the extremum obtained is a minimum and not a maximum, note that a maximum for F requires all the velocities to be zero--clearly this is not the case. The relationship given in equation (5.11) is for a minimization of F.

This latter relation must be satisfied within each segment on the artery. Taking the ratio of (5.5) and (5.4) results in

$$\frac{y_{2k}}{y_{2k-1}} = \frac{B_{2k}}{B_{2k-1}} \frac{v_{2k}}{v_{2k-1}} \frac{(v_f - v_{2k})}{(v_f - v_{2k-1})} \qquad k = 1,\ldots,n \tag{5.12}$$

Substituting (5.11) into (5.12) yields

$$\frac{y_{2k}}{y_{2k-1}} = \frac{v_{2k-1}}{v_{2k}} \frac{(v_f - v_{2k})}{(v_f - v_{2k-1})} = \frac{\left(\frac{v_f}{v_{2k}} - 1\right)}{\left(\frac{v_f}{v_{2k-1}} - 1\right)} \qquad k = 1,\ldots,n \tag{5.13}$$

Simultaneous solution of (5.7) and (5.13) results in

$$\frac{v_{2k-1}}{v_f} = \frac{\left(1 + \frac{y_{2k}}{y_{2k-1}}\right)}{\frac{y_{2k}}{y_{2k-1}} + \frac{2v_f}{v_{ek}} - 1} \qquad k = 1,\ldots,n \tag{5.14}$$

and

$$\frac{v_{2k}}{v_f} = \frac{1 + \dfrac{y_{2k-1}}{y_{2k}}}{\dfrac{y_{2k-1}}{y_{2k}} + \dfrac{2v_f}{v_{ek}} - 1} \qquad k = 1,\ldots,n \qquad (5.15)$$

A realizable design, obtained using the Morgan and Little procedure and the theorems introduced in Chapter IV, accounts for the physical dimensions of the artery, the signal green times and the maximum allowable speeds. In applying these methods it is implied that the total bandwidth along the artery is constant,

$$B_{2k-1} + B_{2k} = C \qquad k = 1,\ldots,n \qquad (5.16)$$

The optimal design results from an iterative search employing the exact equations (5.14) and (5.15) and the design method of Morgan and Little, subject to the minimization of the cost function of equation (5.8).

5.4 The Search

The search is initiated by selecting a point V_o having n coordinates, each coordinate representing the equal progression velocity on a segment of the artery. Each coordinate is restricted to the interval

$$0.5\ v_f \leq v_{ko} \leq v_f \qquad k = 1,\ldots,n \qquad (5.17)$$

A design with one or more velocity coordinates in the neglected interval, $0 \leq v_{ko} < 0.5\ v_f$, always provides lower quality service than one with all coordinates in the interval (5.17). This means that the flows are always achieved on the left half of the flow-density curve of Figure 5.1.

The second theorem of Chapter IV permits transforming the point V_o associated with the given artery A to the equivalent point V_o' associated with the equivalent artery A'. The point V_o' is given by

$$V_o' = (1,\ 1, \ldots, 1)\ v_o$$

Theorem 1 permits the determination of the maximum bandwidth which is realizable on the artery A' for the velocities V_o', or equivalently, on A for V_o. Equations (5.14) and (5.15) determine the inbound and outbound progression velocities on each segment. Finally, these values are substituted into the cost function expression and stored as F.

The coordinates of V_o are successively incremented and in each instance the above procedure is repeated to determine F_i. This F_i is compared with F_o to determine if better service is achieved. The procedure is repeated until no improvement is obtained. The incremental step size is then reduced for finer resolution of the function space. A steepest descent approach is used to speed the rate of convergence.

The combination of the arterial geometry and the signal variables produces a bandwidth-velocity function which is not convex as noted in Chapter IV, see Figure 4.5. As a result the function $F(\cdot)$ to be minimized is not convex. Therefore, one cannot achieve a global minimum for $F(\cdot)$ by a direct application of the above search procedure, although a local one is assured. The global minimum is obtained by determining the minimum for the function resulting from a number of different starting points V_o.

5.5 Example

Present attempts to establish progressions on two-way streets are mostly concerned with the heavier demand directions. To adjust to a change in the demand once the system is in operation, the signal cycle is modified. Such action decreases or increases the progression velocities in both directions by the same factor. This unduly penalizes the traffic in the low demand direction.

Euclid Avenue in Cleveland, Ohio, is a typical street on which progressions might be established. It has ten intersections, unequally spaced over its 6050 feet, and traffic signals having a variety of red-green splits. Initially it has been assumed that the two demands are constant along the street length. The results in Figure 5.3 illustrate that it is possible to give better service to traffic in both directions in terms of a higher progression speed by taking account of the light demand as well as the heavy.

When the street is partitioned into two or three sections corresponding to lengths of constant demands, similar results are obtained and examples are given in Figure 5.3

	$\frac{y_i}{y_{sm}}$	$\frac{y_o}{y_{sm}}$	B_i	B_o	$\frac{v_i}{v_f}$	$\frac{v_o}{v_f}$
	0.23	0.23	0.237	0.237	0.589	0.589
	0.20	0.23	0.226	0.249	0.669	0.638
	0.15	0.23	0.210	0.265	0.767	0.682
	0.10	0.23	0.196	0.280	0.850	0.711
	0.07	0.23	0.188	0.288	0.896	0.724
			2-sections			
section 1	0.23	0.20	0.257	0.235	0.663	0.693
2	0.15	0.10	0.259	0.229	0.824	0.875
section 1	0.15	0.10	.266	0.237	0.831	0.880
2	0.15	0.18	0.238	0.257	0.804	.773
			3-sections			
section 1	0.23	0.20	0.259	0.237	0.668	0.698
2	0.15	0.20	0.233	0.265	0.798	0.747
3	0.18	0.10	0.270	0.221	0.788	0.870

Figure 5.3. Optimal progression characteristics.

CHAPTER VI

CONCLUDING REMARKS

The primary objectives at the outset of this research were to develop a simulation model for arterial traffic systems and to establish a steady state control strategy for the system for medium to light density traffic conditions. During the research many other facets of the problem were recognized. The intent of this concluding chapter is to summarize the material presented in this dissertation and to indicate potential areas for future research.

6.1 Traffic Model

Crucial to the development of the traffic model were the identification of certain entities and the selection of variables to describe them. The features which distinguish the arterial system are the platoons and queues formed as a consequence of the traffic signal controls. Describing these phenomena in the most direct manner dictated to a great extent the choice of position and density as the basic variables for the system. Thus the description of each platoon and queue in the system requires a pair of these variables. The differential equations for these variables were based on published experimental and theoretical investigations and other approximations due to the author. These equations given for the platoons and queue of an arterial section comprise a component state model. When a number of these component models are combined according to the procedures of system theory, they form a state model (KT1). With such combinations the simulation of all traffic structures (e.g., two-way streets, grids) is possible.

Despite the relative simplicity of the simulation model several factors make analysis difficult. (1) The simulation model for ordinary traffic systems quickly becomes large. Its order at any time is given by $2p + q$, where p is the number of platoons and q the number of queues. The number of platoons is not constant--it is affected by the density conditions and the signal controls. (2) The velocity-density relationship is a non-linear one. (3) Certain random phenomena used to give more realistic results make the system only quasi-deterministic.

The present model has several advantages over other models making it a desirable and effective simulation tool.

1. Variety of Applications. A large number of traffic situations which might require investigation by

simulation have densities ranging from light to medium. The model is developed especially for these conditions. Without requiring a prohibitively large amount of data, it functions equally well in the study of single arteries or complete grids.

2. Design and Analysis Capabilities. Since the effects of some changes in a traffic system cannot be predicted, the capability to evaluate these changes before they are instituted proves to be an economic, time-saving, safety, and even political advantage. The simulation model can be used to determine the best locations and settings of traffic signals, to specify speed limits, to evaluate the desirability of one-way streets and other proposed changes in the traffic system, and to predict effects, for example, of a parking lot on traffic patterns. Further, it may prove useful in the investigation of disturbance propagation along an artery and throughout a grid.

3. Ease of Computer Programming. To simulate a traffic system requires a minimum of data preparation, and the variables (e.g., vehicular velocity, position, and counts) which are time and space dependent, are readily available as outputs. Although velocity, acceleration, and vehicle input behavior have been simulated in the model, it is possible to rapidly change these

descriptions if special studies are required or if further research dictates better ones.

6.2 Control

In Chapter IV it was shown how the procedure developed by Morgan and Little can be used as a basis for designing progressions. The question of selecting design parameters is easier to answer with the inclusion of two results which provide a means for displaying the velocity and bandwidth dependence and the effects of a nonconstant velocity along an arterial, respectively. These results find the greatest application in the optimal design procedure presented in Chapter V.

The optimal design is the steady state control of an artery which minimizes the total travel time on the artery. The iterative search procedure used to accomplish this considers not only the optimality conditions relating velocities and bandwidths but also the constraints imposed by the geometry of the artery and the traffic signal variables.

6.3 Future Investigations

Although many aspects of the arterial problem have been considered, several studies are suggested by this dissertation.

Most of the control strategies presented apply to steady state operation. It is possible to use the model to investigate stability when the input demands change or the signal parameters are varied. Such a study would reveal how quickly the system controls should be adjusted to meet new demands while minimizing transients.

The biggest menace to progression systems is the formation of queues. Although some strategies were suggested in Chapter IV for avoiding queue development, investigations for the future could place greater emphasis on the queue phenomenon and how it contributes to congestion. These studies may suggest the best control strategy to return a congested system to a normal one having progression. Transient stability would be a significant factor in this analysis.

In the event that it is not possible to establish progressions throughout a traffic system, a simulation test can be made to determine alternative traffic settings which will minimize travel and wait time. Although this has been done before, vehicle acceleration and deceleration have always been neglected. It is likely that taking these into account will reveal significant differences between the results obtained when acceleration is considered and when it is ignored.

APPENDIX A

The primary goal achieved in Chapters II and III is the development of an efficient simulation model for urban traffic systems. The equations which comprise this model are best solved using a digital computer. A Fortran program for this purpose is available. The program consists of a main program called TRAFIK and three subroutines: BINOM, POISSON, and GENERAT.

Within the program the following scalars (constant) are used.

ACL	average car length
B	v_1 (see Figure 2.2)
BRAKE	deceleration constant
D1	x_1 (see Figure 2.2)
F	lane occupancy for queues
H	time increment
KN	length of table for Poisson distributed inputs
MN	artery (1) or grid (2) indicator
MPN	number of outputs
MSF	number of expected values of turn coefficients
NEX	ramp (0) or exponential (1) acceleration indicator

NOO	number of inputs
NP	number of arterial components
NS	number of traffic signals
TF	total simulation time
TQUEUE	accumulated queue waiting time
ZOOM	acceleration constant

The following vectors (constant) are used. Unless otherwise specified, the vectors have NP coordinates.

CM	free velocity
DT	length of arterial section
EXPVAL	expected inputs per unit time (NOO coordinates)
KM3	integer denoting index of appropriate expected value of turn coefficient ($1 \leq$ KM3 (I) $\leq$ MSF)
M	number of traffic lanes
TURN	expected value of turn coefficient (MSF coordinates)

The matrix KM is used to specify how the system is connected. Its dimensions are NPx2.

KM(I,1)	index of the component succeeding the Ith (equals 0 if Ith is an output)
KM(I,2)	index of the component preceding the Ith (equals 0 if Ith is an input)

The variable T is time.

The following vectors (variable) each have NP coordinates.

D	output counter (if KM(I,1) = 0), distance from lead vehicle of lead platoon to end of queue (otherwise)

DCR	critical distance at which vehicles either begin braking or continue at constant velocity
DENMAT	average lane occupancy for component
Q	queue length
S	current turn coefficient
X	position of leading edge of lead platoon

In the following matrices (variable) the number of rows is always NP. The number of columns is indicated in each instance.

ACCEL	accelerated velocity, (2 columns) (col 1) for platoon; (col 2) for queue
C	steady state velocity, (2 columns) (col 1) for platoons; (col 2) for queue
CRIT	description of transition platoon existing between platoons and queue and having a constant velocity, (4 columns) (col 1) length of platoon (col 2) vehicular length of platoon (col 3) distance to end of queue (col 4) velocity of platoon
DECEL	description of transition platoon existing between platoons and queue and having a decelerated velocity, (4 columns) (col 1) length of platoon (col 2) vehicular length of platoon (col 3) distance to end of queue (col 4) velocity of platoon
P	length of platoon, (KP columns, see below)
PD	vehicular length of platoon, (KP columns, see below)

The vector NDECL having dimension NP is an indicator for the presence of CRIT and DECEL platoons

(a) neither type platoon present, indicated by 0

(b) only a CRIT platoon present, indicated by 1

(c) both type platoons present, indicated by 2

(d) only a DECEL platoon present, indicated by 3.

Equations (2.17 - 2.20) for the states of the platoons and queues are deceptively simple, but require meticulous care in programming. As a result several variables, not explicit in the equations, have to be introduced.

KQ is a vector with NP coordinates which describes the nature of the queue.

(a) the absence of a queue, indicated by 0.

(b) a queue whose vehicles are accelerated across the intersection, indicated by +1.

(c) a queue whose vehicles move at a constant velocity across the intersection, indicated by -1.

Since the number of platoons on the arterial section is variable, the vector KP of dimension NP denotes this number.

The variables of the traffic signals are contained in the matrix MT having dimensions NS x 5.

(col 1) the current time such that $0 \leq MT(I,1) < MT(I,2)$

(col 2) the signal cycle length

(col 3) the green time

(col 4) indicator for current phase

(a) signal is green, indicated by 0.

(b) signal is red, indicated by 1.

(col 5) indicator for phase of previous time increment.

(Note: If MT(I,4) = MT(I,5), then signal has not changed phase during the time increment.)

Three subroutines are used in the program TRAFIK. Two of them, BINOM and POISSON, determine tables of distribution functions, and the third, GENERAT, using these tables selects random variables which are Poisson or binomially distributed.

BINOM: For each expected value, TURN, for the turn ratio, a row of the matrix ATPOI is determined such that

$$\text{ATPOI}\ (\cdot,\text{I}) = \sum_{\text{J}=0}^{\text{I}-1} \text{P}(\text{X} = \text{J})$$

where $\text{P}(\text{X} = \text{J}) = \binom{100}{\text{J}}\ \text{TURN}^{\text{J}}\ (1\text{-TURN})^{100-\text{J}}$

POISSON: For each expected value, EXPVAL, for the input arrival rate, a row of the matrix THRESH is determined such that

$$\text{THRESH}\ (\cdot,\text{I}) = \sum_{\text{J}=0}^{\text{I}-1} \text{P}(\text{X} = \text{J})$$

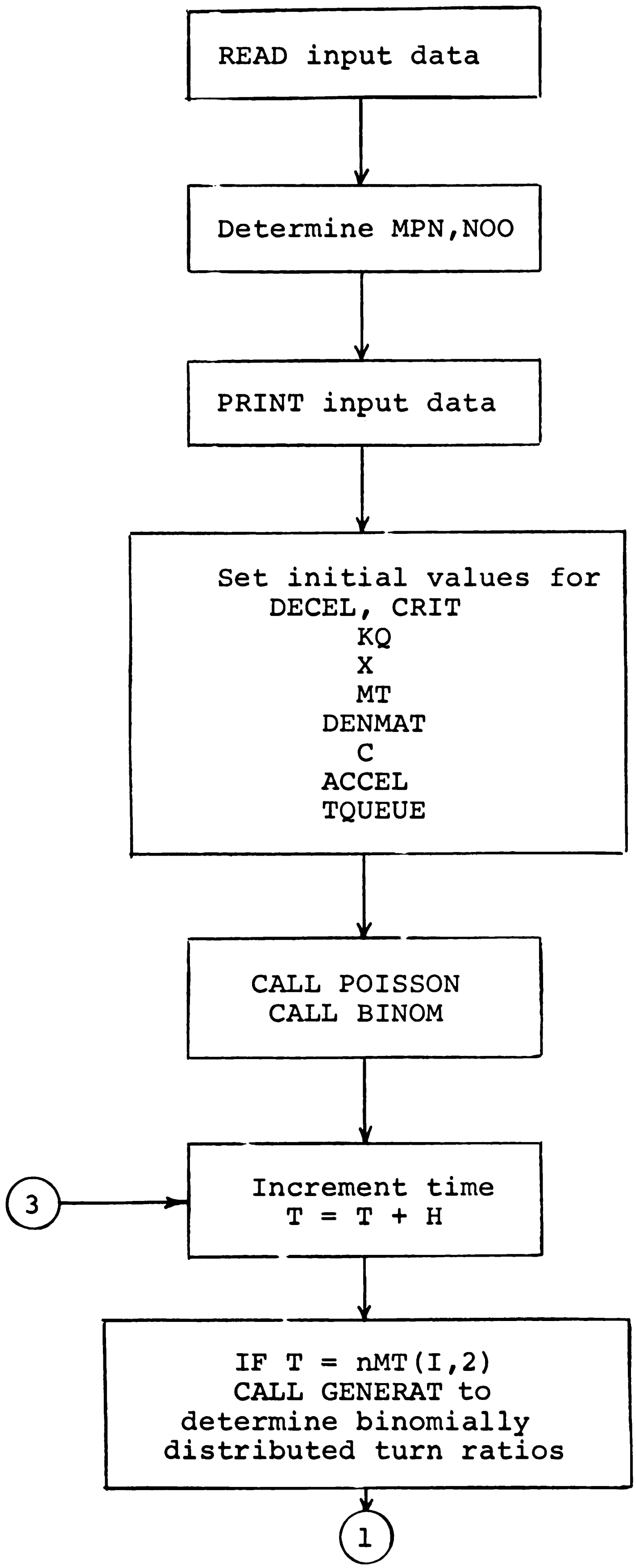
where

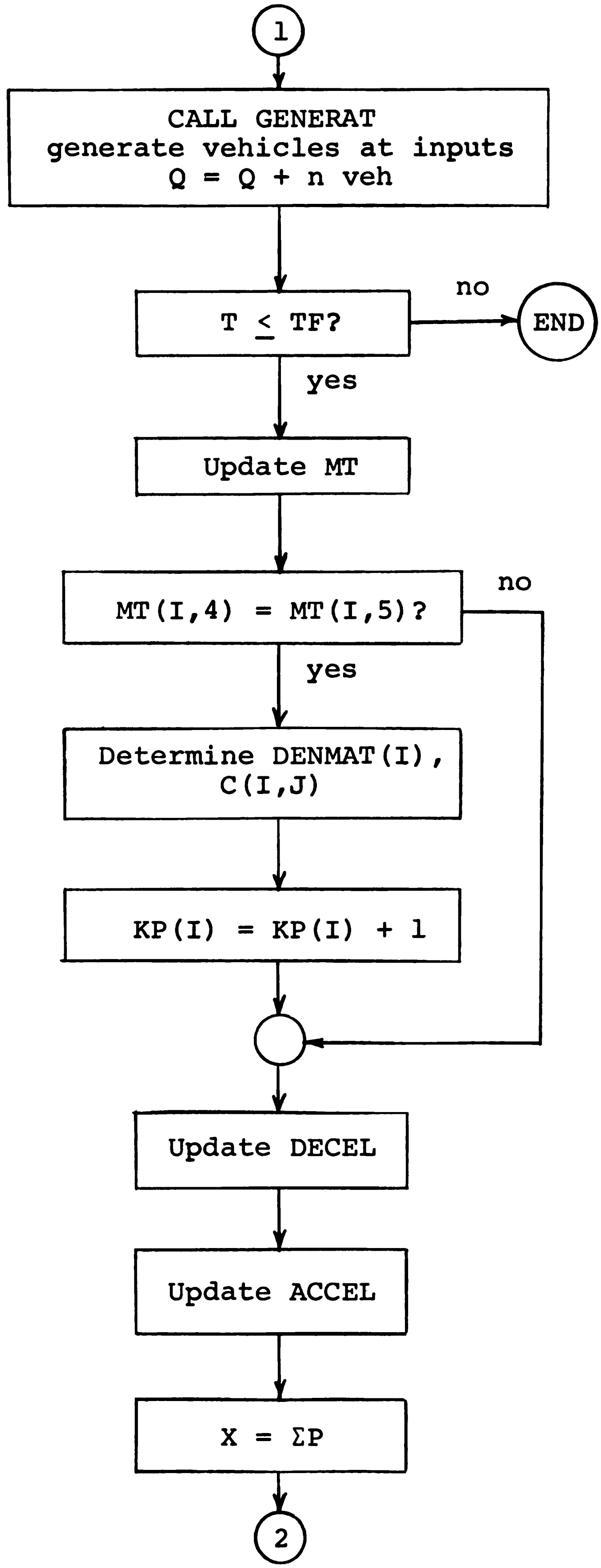
$$\text{P}(\text{X} = \text{J}) = \frac{\exp\ (\text{EXPVAL})(\text{EXPVAL})^{\text{J}}}{\text{J}!}$$

GENERAT: A uniformly distributed variable R is generated (R = RANF(-1)) and is compared with a table to determine another random variable NGEN which is either Poisson or binomially distributed, depending on the table used.

$$\text{THRESH}\ (\cdot,\ \text{NGEN}-1) < \text{R} \leq \text{THRESH}\ (\cdot,\ \text{NGEN})$$

An abbreviated flow diagram of the program is presented in Figure A.1. Since the program is long (over 500 statements) and uses over 150 "IF" statements, only the major logical branches are included. Most of the blocks in the diagram represent a "DO" loop with "J = 1, NP" or "J = 1, NS." It is clear in each instance which one is intended.


READ input data
Determine MPN,NOO
PRINT input data
Set initial values for
DECEL, CRIT
KQ
X
MT
DENMAT
C
ACCEL
TQUEUE
CALL POISSON
CALL BINOM
3
Increment time
T = T + H
IF T = nMT(I,2)
CALL GENERAT to
determine binomially
distributed turn ratios
1


1
CALL GENERAT
generate vehicles at inputs
Q = Q + n veh
T ≤ TF?
no
END
yes
Update MT
MT(I,4) = MT(I,5)?
no
yes
Determine DENMAT(I),
C(I,J)
KP(I) = KP(I) + 1
Update DECEL
Update ACCEL
X = ΣP
2

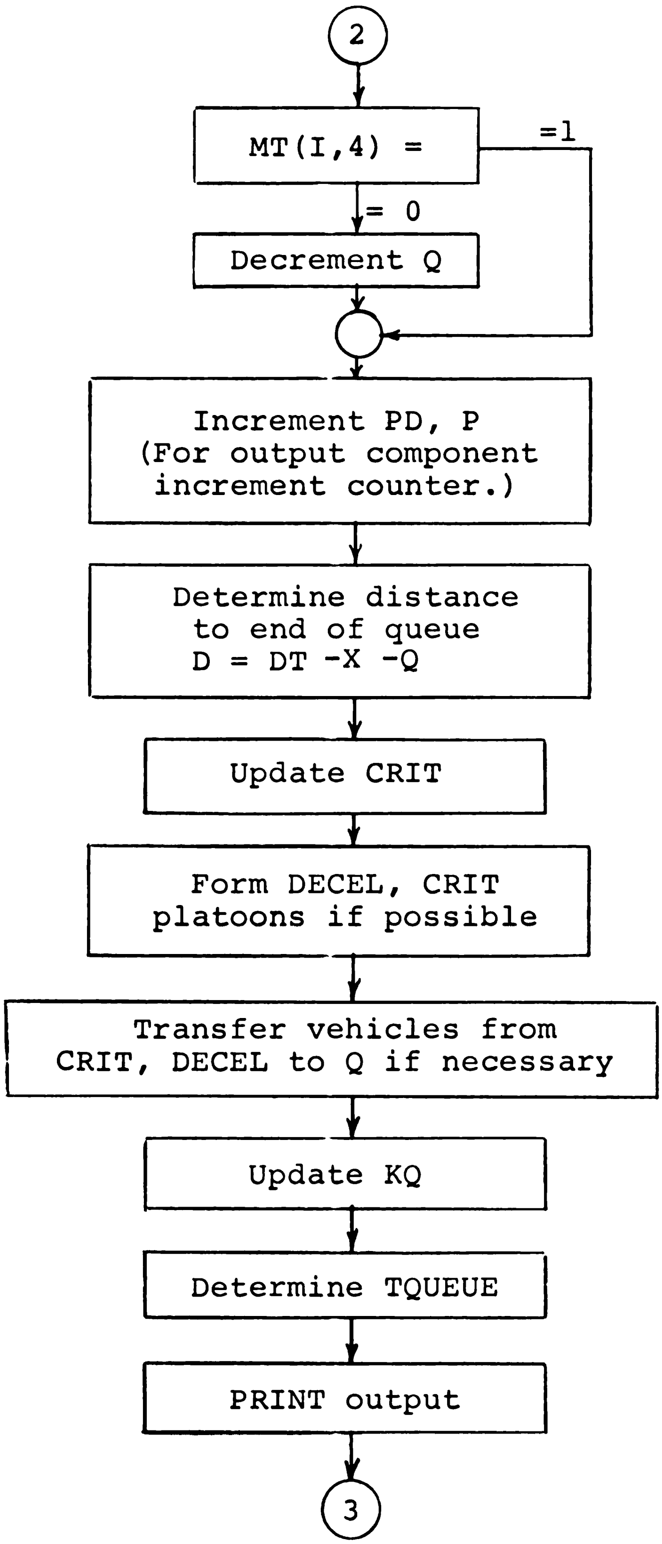


Figure A.1 Flow diagram for TRAFIK

APPENDIX B

To find the optimal set of velocities and bandwidths for progressions on a two-way street when the demand is not constant the Fortran program SEARCH was written. The program uses two subroutines: INITIAL and MORLIT. The variables used in SEARCH are listed below. The dimensions of the matrix variables are clearly indicated.

BW	bandwidth requirement for inbound (row 1) and outbound (row 2) directions, (2 x M)
BMAX	maximum available bandwidth
D	length of section, (M)
DELTA	step increment (3)
DIST	pavement length between successive traffic signals, (N-1)
M	number of sections
N	number of traffic signals
TWOB	total required bandwidth
VA	normalized velocity of inbound (row 1) and outbound (row 2) traffic flow, $0.5 < VA \leq 1.0$, (2 x M)
XE	reciprocal equal velocity, (M)

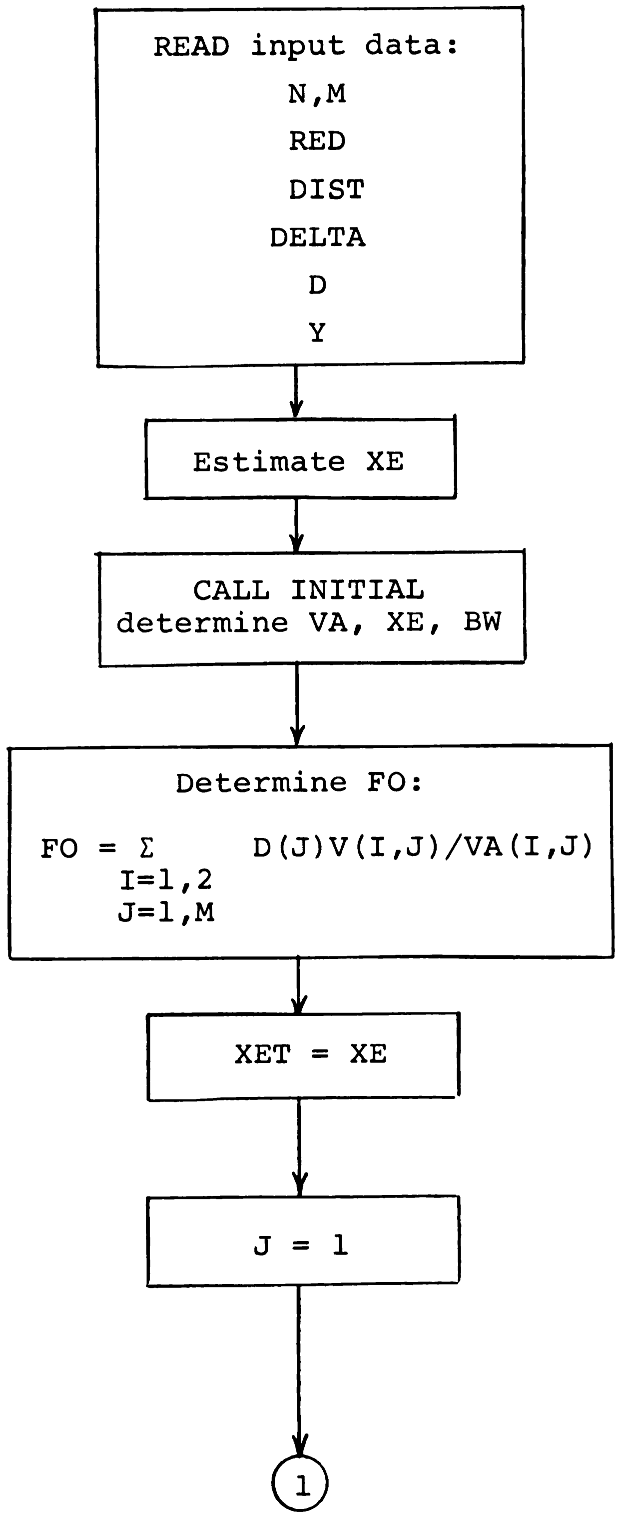
XET trial reciprocal equal velocity, (M)

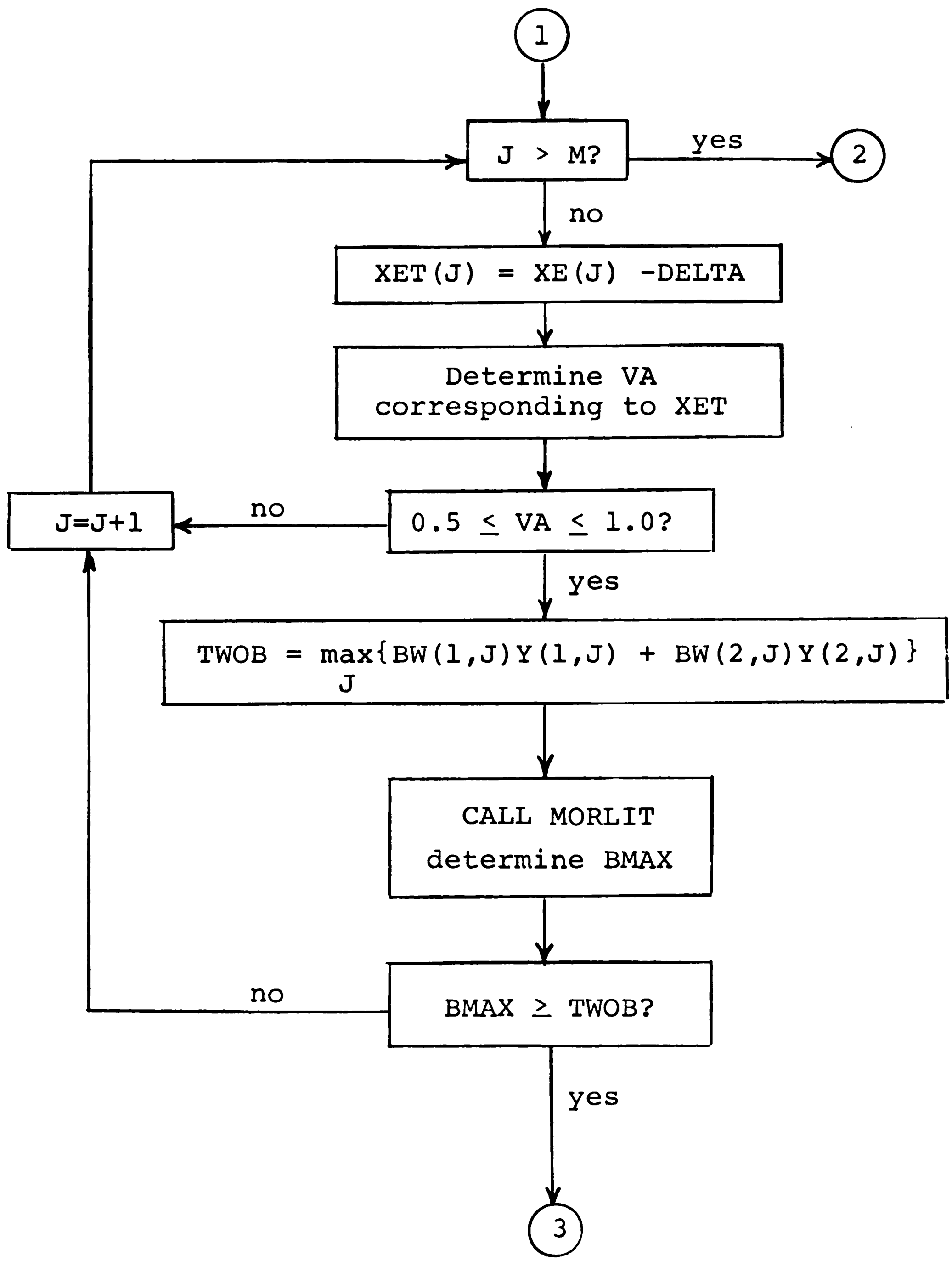
Y flow demand of inbound (row 1) and outbound (row 2) traffic, (2 x M)

The purpose of the subroutine INITIAL is to determine a set of velocities VA (and XE) and bandwidths BW (which meet the realizability requirements of a solution) to serve as a starting point in the search.

The subroutine MORLIT is based on the equations developed by Morgan and Little. It has been written so that for a range of equal velocities the maximum available bandwidth BMAX is determined. This subroutine considers the constraints imposed by the green-red splits of the traffic signals and the distances between them.

The following flow diagram, based on the ideas developed in Chapter V, outlines the main features of the program SEARCH.


READ input data:
N,M
RED
DIST
DELTA
D
Y
Estimate XE
CALL INITIAL
determine VA, XE, BW
Determine FO:
$FO = \sum_{\substack{I=1,2\\J=1,M}} D(J)V(I,J)/VA(I,J)$
XET = XE
J = 1
1


1
J > M?
yes
2
no
XET(J) = XE(J) -DELTA
Determine VA
corresponding to XET
J=J+1
no
0.5 ≤ VA ≤ 1.0?
yes
TWOB = max{BW(1,J)Y(1,J) + BW(2,J)Y(2,J)}
J
CALL MORLIT
determine BMAX
no
BMAX ≥ TWOB?
yes
3

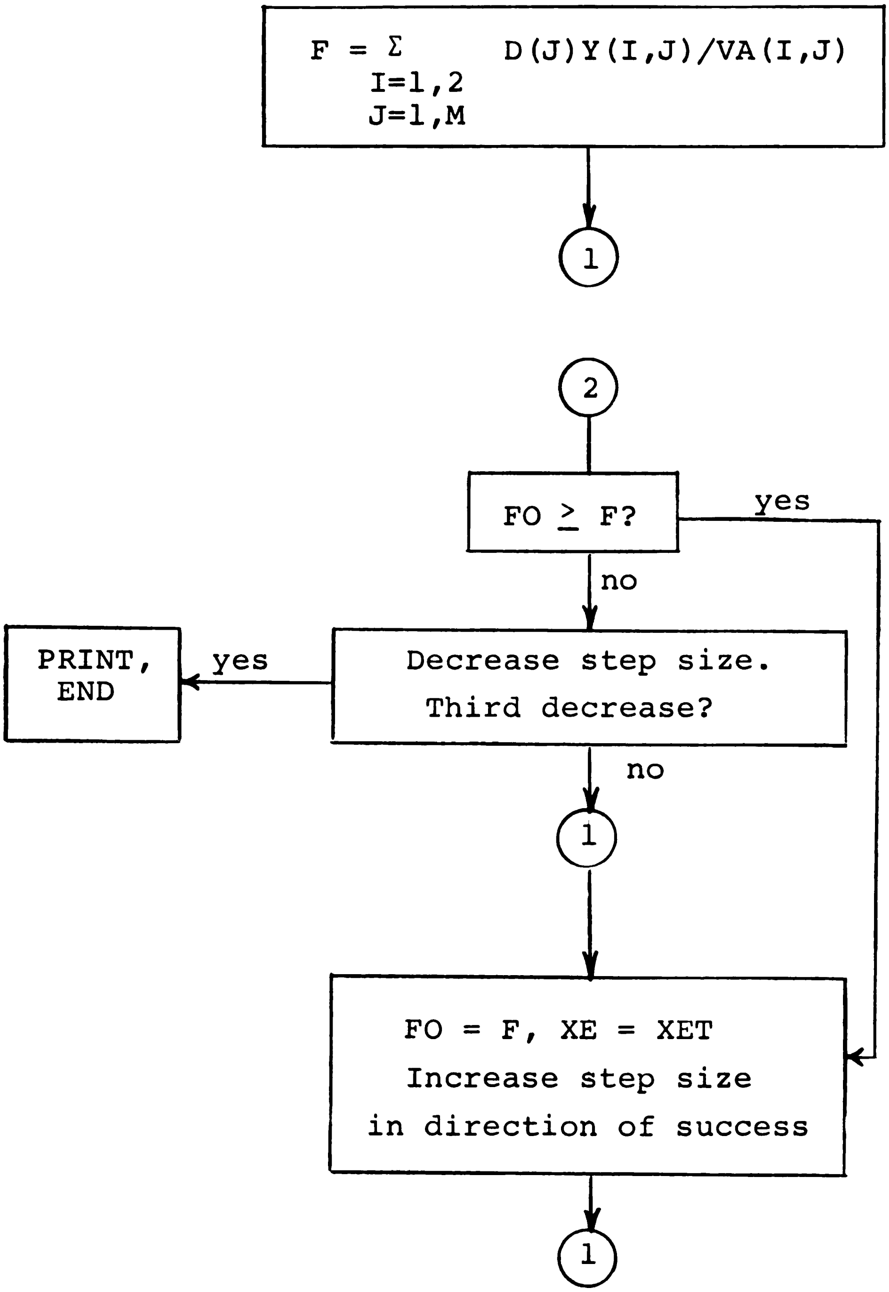


Figure B.1 Flow diagram for SEARCH

BIBLIOGRAPHY

(BE1) Beckmann, M. J., *et al*. *Studies in the Economics of Transportation*. New Haven: Yale University Press, 1956.

(CA1) Casciato, L. and Cass, S. "Pilot Study of the Automatic Control of Traffic Signals by a General Purpose Electronic Computer," *Highway Research Bulletin*, 338 (1962), 28-39.

(CH1) Chang, A., "Synchronization of Traffic Signals in Grid Networks," *IBM Journal*, 11 (July, 1967), 436-441.

(GA1) Gazis, D. C., Herman, R., and Rothery, R. "Mathematical Car-Following Theory of Traffic Flow," *Proceedings American Society of Civil Engineers* (Engineering Mechanics Division), 89 (December, 1963), 29-46.

(GA2) Gazis, D. C. "Traffic Control, Time-Space Diagrams and Networks" in Horton, T. R. (ed.), *Traffic Control-Theory and Instrumentation*. New York: Plenum Press, 1965.

(GE1) Gervais, E. F., and Associates; and Wiley, T. T., and Associates, Traffic Study--Central Business District Lansing, Michigan, Report, February, 1969.

(GO1) Goodnuff, J. L. *An Application of System Theory to the Optimal Control of Vehicular Traffic Networks*, Ph.D. Thesis, Michigan State University, 1967.

(GR1) Greenshields, B. O. "A Study of Traffic Capacity," *Proceedings of the Highway Research Board*, 14, (1934), 448-477.

(HA1) Haight, F. *Mathematical Theories of Traffic Flow*. New York: Academic Press, 1963.

(HRl) Herman, R., Montroll, E. W. "Traffic Dynamics: Analysis of Stability in Car-Following," Operations Research, 7 (1959), 96-106.

(HUl) Heumann, G. H. "Digital Computers Rev Up Traffic Control," Control Engineering, 14 (August, 1967), 95-100.

(KTl) Koenig, H. E., Tokad, Y. and Kesavan, H. K. Analysis of Discrete Physical Systems. New York: McGraw-Hill, 1967.

(LWl) Lighthill, M. J. and Whitham, G. B. "On Kinematic Waves: II. A Theory of Traffic Flow on Long Crowded Roads," Proceedings of the Royal Society, London, A 229 (1955), 317-345.

(MLl) Little, J. D. C., Martin, B. V., and Morgan, J. T. "Synchronizing Traffic Signals for Maximal Bandwidths," Highway Research Board Record, 118 (1966), 21-47.

(MRl) Miller, A. J. "Analysis of Bunching in Rural Two-Lane Traffic," Operations Research, 11 (March-April, 1963), 236-247.

(NEl) Newell, G. F. "Statistical Analysis of the Flow of Highway Traffic through a Signalized Intersection," Quarterly Applied Mathematics, 13 (1956), 353-369.

(SJl) "San Jose Traffic Control Project, Final Report," Data Processing Division, IBM, Kingston, New York.

(TAl) Tanner, J. C. "A Theoretical Analysis of Delays at an Uncontrolled Intersection," Biometrika, 49: 1 and 2 (June, 1952), 163-170.

(WEl) Weiss, G. H. "A Survey of Some Recent Research in Road Traffic," Proceedings of the Symposium on Congestion Theory, Smith, W. L. and Wilkinson, W. E. (ed.). Chapel Hill: University of North Carolina Press, 1965.

(WLl) Wilde, D. J. Optimum Seeking Methods. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.