FREEWAY MANAGEMENT
WORKING PAPER
CURRENT DEVELOPMENT OF TRAFFIC MODELS

by

Mark E. Hallenbeck
Research Engineer

Chien Lin
Research Assistant

Washington State Transportation Center (TRAC)
The Corbet Building, Suite 204
4507 University Way N.E.
University of Washington, JE-10
Seattle, Washington 98105

Washington State Department of Transportation
Technical Monitor
Les Jacobson
FAME Project Manager

Draft Final Report

Research Project Y-3399
Task 30

Prepared for

Washington State Transportation Commission
Department of Transportation
and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

May 17, 1988

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

This paper provides an overview of traffic models that are available for use by WSDOT. It is intended to provide a general review of modeling capabilities that exist today, and present a future vision of what those capabilities will be in the near future. The information collected for this paper was used to select models to be tested for improving the department's ability to manage traffic during construction projects as well as during normal operation.

Material presented in this document is based primarily on an extensive literature search, supplemented by the review of the documentation for a limited number of models, and the experience of the project staff. Many of the models reviewed are still undergoing modification and refinement and many of these changes are not addressed in the literature. Thus, the reader should be aware that some models discussed below have capabilities not mentioned in this paper.

BACKGROUND

Congestion has been increasing on almost every major urban freeway in the United States. Because freeways were originally conceived and designed as free-flowing, limited-access facilities, little consideration was given during their design to the possibility of their needing traffic control systems. Growing traffic demand and its resulting congestion have forced transportation engineers to search for ways to control freeway traffic that maximize the capacity of existing freeways while minimizing traffic impacts on surrounding arterials.

Methods currently used for improving freeway operations include incident management, aid to stranded motorists, driver information systems, ramp metering and high occupancy vehicle (HOV) facilities. These activities help improve operation and safety during both peak and off-peak periods [92]. A significant problem is deciding which strategies or combinations of strategies are most effective in improving freeway levels of service.

Increasing urban development has also significantly impacted traffic levels on arterial streets. Congestion levels on urban arterials have grown substantially in the last decade, and many traffic control plans do not maximize the use of the available facilities. Because of the availability of a variety of signal timing programs, alternative improvements to arterials are more easily studied than options for freeways. However, even for arterial signals, the available analysis tools have limitations.

Considerable efforts are being undertaken, both within this country and abroad, to develop tools for examining the operation of freeway and arterial systems. Much of the work being done involves the use of computer models to examine the impacts of different traffic control strategies and to assist in the development of better traffic control systems.

In addition, advanced traffic surveillance and control systems (for both freeways and urban street systems) have become widely accepted in the U.S. as a means of improving traffic flow. Such acceptance has been accelerated by revolutionary advances, and associated cost reductions, in computer and electronic technology. Recent work within the field of traffic surveillance and control has advanced beyond experimentation to the point where such systems are considered effective tools of operation. Both the SCOOT system in Britain and SCAT system in Australia both have been used successfully. These two systems have demonstrated the importance of computer models in the evolutionary process of real-time control on integrated freeway and urban street traffic systems.

This paper presents the findings of an extensive literature search and review into these computerized tools and techniques. It provides an introduction to the types of computer models that are available, briefly describes the more common models, and presents the project team's opinion on the future improvements that need to take place in the area of computerized traffic modeling.

STRUCTURE OF THE PAPER

The primary objective of this working paper is to examine the existing tools for analyzing the impacts of alternative freeway management strategies. The paper is broken into a number of chapters discussing the various types of traffic models that are currently available. This first chapter introduces modeling concepts and the basic categories of traffic models.

To provide a background for the review of specific models, traffic control strategies and techniques are described in Chapter 2. These traffic control strategies and techniques are associated with one of the following traffic situations:

- urban street traffic control,
- freeway traffic control,
- freeway corridor traffic control, and
- integrated freeway and area traffic control.

Most existing traffic control computer models fall into the first category, and most freeway simulation models are in the third category. The fourth situation is generally recognized as the most difficult task and is the frontier of traffic model development.

The most widely used off-line, arterial traffic control models are described in Chapter 3, and the three most important demand-responsive (on-line) urban traffic control systems (UTCS, SCOOT, SCAT) are described in Chapter 5. The description of off-line freeway simulation modes in Chapter 4 is based primarily on previously published work by Adolf May of the University of California at Berkeley's Institute for Transportation Studies. Information about on-line freeway traffic control models was not found in the literature search, and thus they are not reviewed in this report.

Chapter 6 presents an overview of the models most recently released or currently under development. The models in this chapter represent the latest modeling trend in the U.S., the integration of existing models. The AAP (Arterial Analysis Package), Integrated

Simulation Model (TRAF) and Integrated Traffic Data System (ITDS) are basically updates, enhancements and integrations of existing models rather than new models.

Chapter 7 provides a look into potential modeling improvements to expect in the near future. It discusses the impacts of microcomputers, computer graphics, database management systems, and artificial intelligence. It also presents the research team's outline of the subjects about which additional research (at both the state and national levels) needs to occur in the areas of traffic control and modeling.

INTRODUCTION TO MODELING CONCEPTS

Numerous methods exist for categorizing traffic models. Among the more important distinctions of models are the following:

- on- versus off-line operation,
- computation versus simulation algorithms,
- empirical versus analytical algorithms,
- deterministic versus stochastic modeling,
- microscopic versus macroscopic scales,
- event scan versus time scan model timing, and
- optimization versus evaluation objectives.

All traffic models consist of some combination of the above "methods." In the following paragraphs, each of the above terms will be briefly defined. These terms will then be used to describe the various modeling alternatives presented in Chapters 3 through 5.

On-line versus Off-line Models

On-line computer models are connected to the traffic control system, collect data in real time and interact with the control devices. An on-line model can be implemented using either fixed time-of-day or traffic-responsive control. Fixed time control utilizes different operating parameters at specified times of the day. Traffic responsive control uses a computer to

- collect data on traffic flow conditions,
- make calculations to determine a desired timing plan, and
- implement or adjust the timing plan over short time intervals such as each cycle, every 5 minutes, or every 15 minutes.

Off-line computer models can be quite similar to on-line models except that they use data collected previously, and the operating plans they develop are applied in the future. There is no guarantee that the traffic actually experienced by the control system is roughly equivalent to that represented by the data used to develop the timing plan.

The advantage of the off-line mode is that no time constraints are placed on the operation of the model. This allows the use of more refined and complex models and consequently improves the analyses that can be done. The primary difficulty with more complete and complex modeling of traffic situations is that it also requires more refined data for input to the models.

In typical on-line applications, the complexity of the models is significantly restricted. This is because the models operate under significant time constraints, in that they must be executed quickly, often with limited input data, in order to make control decisions for immediate implementation.

The two most important methods of on-line timing plan selection are

- time-of-day control, and
- traffic-responsive control.

In time-of-day control, a strategic timing plan is introduced into operation at a given time of the day. Often the timing plans only respond to three basic time periods, such as morning peak (7 to 9 AM), evening peak (3 to 6 PM) and off-peak. Modern signal controls can store and use many more time periods of signal information.

Where traffic demand is reasonably predictable, coordinated pretimed controllers operating under time-of-day control can provide satisfactory operations. Where one or more intersections requires multiphase operation, fully-actuated controllers can also be incorporated into such a coordinated system. Such systems, using state-of-the-are equipment, are available at present from traffic signal manufacturers. A typical system can provide up to four cycle lengths, three splits per cycle length, and three offsets per cycle length. Finally, present hardware technology permits engineers to change traffic signal patterns by downloading new patterns from a central system master into a microprocessor-based local controller.

In traffic-responsive control, the strategic timing plan is based upon traffic conditions that are measured through a traffic-detection system. With this approach, timing plan updates can occur as often as every cycle. Computer control systems of this type are referred to as either "table look-up" or "pattern generation systems."

On-line modelling efforts for real-time control purposes started in the late 1960s and continue today. Most of the work done in the U.S. on this type of model has been undertaken by researchers at KLD Associates, Inc., and at Sperry System Management, Inc.

Computation versus Simulation Models

Computational models involve the application of mathematical equations to calculate solutions to problems. These equations may represent fundamental mathematical truths, may be derived from basic principles (e.g., trigonometric functions) or may simply reflect an established relationship between several variables. In traffic modeling, a computational model attempts to solve for the "best" traffic control plan for a given set of available control options (i.e., signal plans). They do this by calculating which combination of control parameters (signal timings) minimize the delay, stops, or other criteria for traffic included in the model, using some mathematical algorithm.

A simulation model, on the other hand, is a mathematical representation of the sequence of events that comprise a process. In traffic modeling, a simulation model estimates what will happen to traffic flow given a set of control strategies. It does not solve for a specific answer. Simulation models do not select appropriate control strategies, they only describe the results of strategies selected outside of the simulation process.

The development of effective traffic control systems and strategies requires the ability to predict the response of the traffic environment to the implemented control. Simulation provides a powerful tool for gaining enough knowledge about this response for virtually all network configurations, traffic conditions, and control policies.

Simulations are used to test traffic control systems and strategies because

- an evaluation of the behavior of new control systems and/or operating procedures is needed before their actual implementation; and
- performing tests in the real world is often not practical because
    - implementation may be very expensive and/or time consuming,
    - experimentation with the real system may entail considerable risk (such as accidents), and
    - several alternative control systems or operating policies may need to be tested.

A variety of computerized traffic simulation models have been developed and refined for almost three decades. Several models have been tested, and each has its advantages and disadvantages. These are discussed in later chapters of this paper.

Empirical versus Analytical Models

The Highway Capacity Manual is an example of an empirical model. In this case, the basic relationships within the model were arrived at experimentally through extensive field studies. An empirical model uses results obtained previously with a variety of scientific observations. Analytical models take the form of equations developed by a purely computational process. In most traffic models, a combination of empirical and analytical methods are used.

Deterministic versus Stochastic Models

In a deterministic model, the fictitious sequence of events used within the model has a completely predictable outcome. For example, a bus passing through a toll plaza may be required to pay a specified fee and to use a specified lane that guarantees the bus precedence over automobile traffic upon entering the facility. The set of rules that govern the passage of a bus through the toll plaza under these circumstances would be described as a deterministic model. Deterministic models, by themselves, do not usually constitute the entire process being simulated, since they offer little potential for problem solving under repeated application. They are more commonly incorporated as sub-models within the overall program structure.

In a stochastic model, the outcome of a given sequence of events is not completely predictable, but depends on something that happens during the course of the process. In the toll plaza example, vehicles may pay a variable fee, depending on the number of axles on the vehicle, and each vehicle may be assigned to different lanes depending on whether the driver has the correct change available. (The availability of correct change will be a probability function.) The vehicles may experience further delay as a result of the driver missing the coin basket or by having to yield right of way to other traffic (buses for example) before entering the facility. The passage of vehicles through a toll plaza under these conditions would be described by a stochastic model, since the outcome of the process depends on a number of events, each of which can be described only in terms of the probability of its occurrence.

Microscopic versus Macroscopic Models

A process such as the flow of traffic may be simulated either at the microscopic level, in which each vehicle would be treated as a separate unit, or at the macroscopic level, in which the characteristics of the stream as a whole would be examined. In general, microscopic models tend to be more accurate in their description of the processing being simulated, but they usually require considerably more input data, programming and debugging effort, computer storage and computer time for execution. They also tend to be more demanding in terms of the level of detail required in their assumptions and approximations. These assumptions can lead to problems of credibility in the results if they are not properly designed and selected. In some cases, such as in SCOT, a third level of detail, termed "mesoscopic," is used in models to represent platoon movement. The MACK model is an example of macroscopic simulation. INTRAS is an example of a microscopic model.

Event Scan versus Time Scan Models

A further distinction can be made between models in which the process being analyzed is updated at constant time intervals (e.g., one second) or upon each event that occurs. Time scan models are, in general, easier to develop because the time factor is advanced by a constant increment each time the process is examined. Event scan models, in which the process is updated as each event occurs, are usually more efficient in terms of computer time, since they only update the simulated process in response to a specified event.

Optimization versus Evaluation Models

The two main purposes of computer modeling are

- to determine the values of specific design parameters that will optimize the operation (e.g., cycle, splits, sequence and offsets at a traffic signal or a signal network), and
- to evaluate the operation as a "system" with specified design parameters in terms of measures of effectiveness (e.g., delay, stops, fuel consumption, etc.).

Simulation models do not, by themselves, have any inherent optimization capabilities, and consequently, most fall under the category of "evaluation" models. They simply reproduce the process as faithfully as possible and accumulate the results. To obtain an optimal solution using simulation, the model must be applied repetitively with different design parameters selected by some other program or the operator. The set of design parameters that yields the best results should be chosen as the correct solution. Simulation is therefore best suited to the comparison of a small number of widely differing strategies. Examples of simulation models that do not optimize by themselves are NETSIM, TEXAS, and PRIFRE.

A function of all optimization models is to seek the best solution. Such a model may or may not provide the required degree of evaluation, although most optimization models include realistic simulations within the modeling package.

CHAPTER 2
TRAFFIC CONTROL STRATEGY AND TECHNIQUES

This chapter provides background information on the strategies and techniques available for controlling traffic. Traffic control is defined as the regulation, warning, and guidance of traffic for the purpose of improving the safety and efficiency of moving people and goods. Implementing this control process involves the installation, operation, and maintenance of various traffic control devices such as signs and signals. This chapter describes the basic methods for each of four types of urban systems:

- urban street traffic control,
- freeway traffic control,
- freeway corridor control, and
- integrated freeway and area traffic control.

URBAN STREET TRAFFIC CONTROL

Through the use of signal timing plans, traffic flow can be controlled to minimize delay, stops, fuel consumption and/or vehicle emissions. A variety of computer models are available to help prepare timing plans and even control traffic systems in real time.

The control of signalized intersections may be grouped into five fundamental categories [92]:

- isolated intersection control,
- arterial intersection control,
- arterial network control,
- areawide system control, and
- diamond interchange control.

Each of these control concepts is discussed below.

Isolated Intersection Control

Isolated intersection control is used for a signalized intersection when the flow of traffic through adjacent signalized intersections is not considered. Traffic signal control concepts for isolated intersections include two basic categories: pretimed signal control and traffic-actuated signal control.

Pretimed control assigns the right-of-way at an intersection according to a predetermined schedule. The sequence of right-of-way assignments (phases), and the length of the time interval for each signal indication in the cycle is fixed and is repeated with no consideration of the current traffic demand on the intersection. The major elements of pretimed control are fixed cycle length, fixed phase lengths, and number and sequence of phases. The timing plans for this type can be prepared manually or by an off-line computer model based on historical traffic patterns.

Traffic-actuated control techniques attempt to adjust green time, and, in some cases, the sequence of phasing (through skipping of phases with no traffic demand) continuously. These adjustments occur in accordance with real-time measures of traffic demand obtained from vehicle detectors placed on the approaches to the intersection. The basic timing parameters in a traffic-actuated phase are

- yellow change,
- red clearance intervals,
- minimum green passage time interval (vehicle interval), and
- maximum interval duration.

Intersection signal timing using either of these techniques can be prepared with manual design methods. However, these methods are relatively inefficient for conducting comprehensive analyses or for evaluating alternative timing plans for more than one set of geometric and/or traffic conditions. Consequently, computer models for intersection signal timing were among the earliest computer models for traffic control. The SOAP and TEXAS models are currently the most well documented and applied models for isolated intersection signal timing analysis, although the new Highway Capacity Manual software is also used for this purpose. SOAP is an optimization model, and TEXAS is a simulation model.

Arterial Intersection Control

Arterial intersection control is used along a signalized arterial when major emphasis is given to the provision of progressive traffic flow along the arterial at the expense of traffic movement on streets crossing that arterial. The signals along the arterial are analyzed as one system.

The basic approach to arterial signal control assumes that vehicles traveling along the arterial street are released in platoons from a signal, and hence, travel in platoons to the next signal. A timing plan that permits the continuous progression of these platoons along a street reduces the delay experienced by vehicles in those platoons.

To achieve this efficient coordination, three main signal timing parameters are selected at each intersection: cycle length, split and offset. The correct settings of these parameters in a series of intersections often results in the so-called "green wave" in which certain motorists can travel along a given route without stops or serious delay. Each of these parameters is discussed below.

Cycle length is the length of time a signal takes to rotate from the beginning of a green signal in a particular direction, through that signal lamp's yellow and red phases, and back to green again. In a coordinated network, all signals must operate on the same cycle time, or an even multiple (usually 2) of that cycle time.

Phase splits are the length of the various signal phases. That is, how long each green, yellow and red phase will be for each direction of travel. This must be determined for each individual intersection in the system. Phase lengths (splits) may vary from intersection to intersection.

An offset value is the lag between the times when adjacent signals turn green in a particular travel direction. For example, if the first signal on an arterial turns green in the southbound direction at time A, the offset for the second signal that turns green southbound at time B, will be A minus B. One offset value must be determined for each intersection in the system. This offset is based on the timing pattern of one "master" intersection and direction.

Developing timing plans for an arterial street is a complex task, especially if the street carries two directions of traffic. The problem of providing progression in both directions is not always solvable, and in most cases, the optimization of progression in both directions is mathematically complex. In addition, the traffic engineer must contend with the variation of traffic flow throughout the day (i.e., what is optimum at 10:30 is not optimum at 11:00). Thus, at least one timing plan must be developed for each allowed set of traffic conditions.

Computer models can be used to deal with traffic signal timing either using off-line or on-line techniques. Off-line computer models use data collected at some previous point in time to make the calculations necessary for determining "optimum" timing plans. Those plans are then implemented without direct interaction with the actual traffic flow on the arterial.

On-line computer models can be controlled either by time-of-day or traffic-responsive control, as noted in Chapter 1. On-line systems utilize a computer and various vehicle detectors to

- collect data on traffic flow conditions,
- make calculations to determine a desired timing plan (or changes to an existing timing plan), and
- implement or adjust the timing plan at planned, short time intervals.

Regardless of whether the model works on-line or off-line, one of two approaches for calculating the appropriate timing parameters is used. These approaches are

- minimizing overall delay and stops, or
- maximizing the width of the progression band.

The TRANSYT model uses the first of these approaches. It attempts to optimization signal offsets in a pretimed signal system by minimizing a linear combination of stops and delay. This specific method does not necessarily minimize the number of stops or provide uninterrupted progression but instead tries to balance both the number of stops required and the total amount of delay experienced by vehicles within the system.

Maximizing bandwidth has been the most popular way of developing arterial timing plans. Several computer models have been developed for arterial signal coordination design using this technique. These models are discussed in detail in Chapter 3.

Arterial Network Control

Arterial network control is a form of signal control for a group of adjacent signalized intersections in which coordination among all movements in the network is considered. A typical example of network control is in the central business district (CBD) of most large cities, where cross street traffic is too heavy to be ignored.

For the most part, the prevailing concept of signal timing for arterial networks has been to provide the best possible progression along all streets in the network. However since the scale and complexity of arterial network control are much greater than those of arterial street control, perhaps only selected streets should be considered for timing peak period flow so that those streets would form an open network like an arterial street system.

Areawide System Control

Areawide system control is a form of traffic signal control that treats all of the traffic signals in a city or metropolitan area (or major portion thereof) as a total system. The major difference between areawide control and network control is in the number of signals included in the analysis. The individual signals within the areawide control system may be controlled using isolated, arterial, arterial network, or a combination of these strategies.

In the past, traffic engineers tended to view the above three types of control separately, and relatively little thought was given to the concept of controlling all of the traffic signals within the limits of an urban area or political jurisdiction with a single control system. However, advanced computer and communication techniques have allowed engineers to recognize the surveillance and control of all traffic signals in an urban area as a feasible and necessary development in many cities [75].

The areawide system control can be designed to operate with either centralized control or distributed control.

In centralized control, all control logic and surveillance capability reside in one location, while in distributed (decentralized) systems, the control logic and surveillance capability are decentralized and placed at various levels in a hierarchical organization of surveillance and control functions.

The control strategies of most of the current generation of area traffic control systems are based on maximizing mobility in a network by minimizing total stops and delays. Timing plans for signal splits and offsets are usually determined from off-line traffic signal optimization programs such as TRANSYT or SIGOP. Since these programs cannot handle saturated or over-saturated conditions, the tendency is for traffic to be moved more quickly into the CBD, or to an urban freeway, than the rate at which it can be absorbed under peak demand conditions. This results in congestion.

This problem creates a need for the development of control strategies that would systematically delay (i.e., manage) some network traffic during certain periods to control congestion buildup in the CBD and/or major freeways and to ultimately reduce total delay. Such control, which would use the available storage in the network to maximum advantage, could be optimal from an overall system point of view and could remove the need for freeway ramp metering or other politically difficult steps. The system would be, in fact, equivalent to ramp metering but on a network-wide basis. Several European countries have implemented these techniques. These systems, variously known as "flow metering," "throttling" or "gatekeeping," deliberately restrict traffic at groups of lights at certain times to control congestion.

Diamond Interchange Control

Diamond interchange control is used for independent, freeway to arterial diamond interchanges. Diamond interchange control strategies are designed to improve the operational efficiency of these intersections. Since the geometric capacity of most diamond interchanges is basically fixed, without major capital investment cost-effective improvements in signal design and operations are the only viable alternative.

Several operational problems can occur with signalization of diamond interchanges. One problem occurs when traffic backs up from one of the ramp intersections through the other intersection. When this happens, traffic at the upstream ramp intersection may be partially or completely halted, thus reducing capacity of that intersection. When this happens at both off-ramps simultaneously, the entire diamond interchange becomes locked up, and no vehicles can move forward.

Another congestion problem that can severely influence traffic operation is when the left-turn pocket overflows and backs up into a through lane, thus reducing the capacity available for through traffic. A third type of songestion occurs on the off-ramp when a long queue of vehicles on the off-ramp stretches onto the freeway, creating a dangerous situation between high-speed freeway vehicles and vehicles stopped on the ramp waiting to move through the interchange.

Depending on the need for interconnection with adjacent traffic signals, diamond interchange signals can operate according to one of the following three principle modes of control [92]:

- isolated signal control — in which signalized intersections at the ramp terminals operate independently from adjacent signalized intersections,
- interconnected signal control — in which the two signalized intersections at the ramp terminals are interconnected to improve the interchange's operational efficiency, and

- signal system control — in which major consideration is given to providing progressive traffic movement along the frontage road and/or the crossing roadway.

Each of the above modes of diamond interchange control can be implemented with either pretimed controllers or fully-actuated controllers.

FREEWAY TRAFFIC CONTROL

Freeway congestion is usually discussed and treated as recurring congestion and nonrecurring congestion. Each type of congestion has a slightly different set of traffic control remedies. Recurring congestion prevails during particular times of the day at certain locations. Recurring congestion is caused either by too many vehicles entering the freeway within a given time period via unrestrained entrance ramps, or by localized reductions in freeway capacity due to geometric design deficiencies such as lane drops, steep grades, sharp horizontal curves, short weaving sections, or poor ramp design.

Nonrecurring congestion occurs unexpectedly as a result of some type of incident. The incident causes a temporary reduction in capacity by blocking lanes or disrupting drivers' concentration, which in turn creates congestion.

In a comprehensive overview of freeway management, Ahmed [2] in which he grouped potential control strategies into capacity management and demand management techniques. That is, congestion may be relieved by either expanding the capacity available or by better managing the demand for the available capacity so that the demand exceeds the capacity less often. The two major types of freeway control strategies that perform these functions are entrance ramp control and mainline control [90].

Entrance Ramp Control

The intent of entrance ramp control is to limit the number of vehicles entering the freeway so that the demand on the freeway itself will not exceed capacity. Optimum traffic densities and speeds can thus be attained and maintained on the freeway and consequently, the maximum number of vehicles can use the facility. With ramp control, drivers on the freeway mainline experience less congestion during peak periods and more orderly merging movements. To achieve these improvements, drivers on the ramps experience some delays waiting to enter the freeway. Theoretically (and usually in practice) the delays experienced on the ramp are smaller than those that would have been experienced on the freeway had there been no ramp metering.

Over a long period of time, drivers may avoid congested ramps, choose alternative routes and or faster transit modes (e.g., carpools) or change the time of day during which they travel. This diversion of traffic may either make better use of available capacity on parallel routes, or it may cause congestion to be transferred from the freeway to surface streets if the capacities of these alternative routes are not sufficient. The potential for significantly increasing congestion elsewhere is the most significant potential drawback to ramp control.

Ramp closure and ramp metering are the two most common techniques of entrance ramp control. Ramp closure during peak traffic conditions is the simplest form of entrance ramp control, although it is implemented infrequently because of political considerations. The benefit of ramp closure is that it eliminates an entire merge/weave process from the freeway system, significantly increasing functional capacity on that section of freeway. The ramp closure may also reduce total demand for the freeway by making parallel routes more attractive than the freeway.

Ramp metering is the most widely used technique to limit the rate at which traffic can enter a freeway. There are several ramp metering control philosophies, including

- local pretimed metering,
- local traffic responsive metering, and
- integrated metering.

With local pretimed metering, metering rates are predetermined from historical data on ramp and mainline traffic conditions near the controlled ramp. The ramp signal operates at a fixed cycle length based on the metering rate selected for a particular time of day.

For traffic-responsive metering, rates are based on real-time measurements of traffic variables (volumes, lane occupancies, speed) that indicate the current relation between upstream demand and downstream condition. Usually, a mathematical algorithm is used to estimate the appropriate ramp discharge rate for the ramp, given the measured conditions and the predicted effect of that metering rate on those conditions.

For the integrated system, a series of entrance ramps feeding a designated freeway section are controlled as one system. A pretimed, or traffic-responsive, metering plan for the controlled ramps is determined to provide a desired level of traffic service on the freeway section, subject to local as well as systemwide constraints.

Mainline Control

Mainline traffic control involves the regulation, warning, and guidance of traffic on the freeway mainline in order to improve safety and increase operational efficiency. The most common techniques of mainline control are motorist information systems, variable speed control, lane closures, and reversible lane control. Each of these are discussed below.

Motorist information systems are meant to provide drivers with relevant, (sometimes real-time) information concerning downstream freeway traffic conditions. Messages might indicate the presence of a capacity-reducing incident, location and duration of congestion, or alternative routes available for diversion. The goals of providing such information are to make drivers aware of hazardous downstream conditions, to help them avoid potentially hazardous situations and to help them avoid congestion whenever possible. Some systems also help redistribute traffic demand to other available routes when necessary.

Message systems can use any number of techniques for delivering information. The most common types of information systems in the U.S. are variable message signs, commercial radio messages, and roadside radio messages. A considerable amount of research is currently being done in this area to improve the information provided to motorists.

Message systems require information so that meaningful, useful messages can be developed and displayed. To provide this information, many urban areas are implementing or upgrading freeway surveillance and control systems.

Variable-speed control is based on the principle that maximum traffic flow occurs when the speed of traffic equals a certain optimum value. This is true when traffic operations are stable. Variable speed control attempts to create these stable conditions by changing the posted (or advisory) speed limit to match the existing conditions on the roadway. Variable speed control has been used in Belgium with some success, but experiments with it in the U.S. have had mixed results. Variable-speed control by itself can postpone the onset of congestion but not prevent it.

Lane closures are implemented during periods of reduced capacity to provide advanced warning of downstream lane blockages and to improve merging operations at downstream entrance ramps. They are most often used to improve flow through construction areas and other temporary capacity reductions.

Reversible lane control is used to change the directional capacity of a roadway when peak-period traffic demand exhibits significant directional imbalance. It is particularly effective for meeting imbalanced traffic flows but requires large amounts of traffic control and usually geometric improvements.

Each of the above systems requires some knowledge of existing congestion to be used effectively. For the most part, this information is obtained from a freeway surveillance system. The functions of a freeway surveillance system are to monitor traffic conditions, detect the occurrence of incidents, and report on the status of the control hardware. To achieve these functions, several surveillance techniques have been used, including moving police patrols, citizen-band radio, motorist call systems, and aerial surveillance. In recent years, the term freeway surveillance has also been used to refer to computer-based systems that use vehicle detectors, as well as closed-circuit television surveillance [11].

Traffic information is typically obtained by means of loop detectors embedded in the pavement at one-half mile intervals. A digital computer processes the output from these detectors and computes real-time estimates of flow rate, lane occupancy, and traffic speed at each detector station. The computer then uses these data for incident detection and traffic-responsive control applications.

Incident detection is accomplished through the execution of certain analytical and logical expressions known as incident detection algorithms. Because of false alarms and the inability of the computer to identify the nature of the traffic incident, closed-circuit television has supplemented computer-based incident detection by providing a human operator with some form of visual validation and assessment of incidents. When an incident is detected, the computer estimates the available capacity at the incident site, determines ramp metering rates, and provides the operator with specific incident-related information, which may, in turn, be displayed on variable message signs.

Delays in providing the needed services at the incident scene and in removing the incident off the freeway can endanger the life and property of those directly involved, result in rear-end collisions of arriving vehicles, and cause the formation of long queues of cars. As a result, in recent years incident management teams have been formed to coordinate the activities of the different agencies responsible for incident servicing. When the computer flags the occurrence of an incident, the control center personnel work with the incident management team to expedite the needed response [1]

FREEWAY CORRIDOR CONTROL

To control the traffic flow within a freeway corridor more effectively, the control concept can be extended to streets related to the freeway. A freeway corridor includes the freeway mainline and its ramps, the freeway service roads (if there are any), the network of parallel arterial streets that can be used as alternate routes, and cross streets that link the freeway ramps with these parallel routes. The purpose of corridor control is to optimize the utilization of the corridor capacity by diverting traffic from overloaded facilities to those with remaining capacity.

The traffic engineer has many tools available to help improve freeway corridor operations. These include, for example,

- preferential high occupancy vehicle (HOV) treatments,
- traffic signal systems,
- ramp metering,
- changeable message signs, and
- incident detection and management [2,3].

Preference given to HOV is intended to encourage the use of transit and the formation of carpools and vanpools. Vehicular demand on the freeway can thus be reduced, while at the same time, more people and goods can be moved on that facility. The remaining strategies have been discussed elsewhere in this chapter. The prime difference in their use in corridor control is that consideration is given to the effects of a control strategy on streets surrounding the freeway being controlled. The goal is that all the best operating plans for the system are implemented, rather than the best plan for each individual facility.

Many large urban centers have installed both area traffic control systems and freeway ramp metering control systems to alleviate congestion on their street networks and urban freeways. The problem is that most of these systems operate independent of one another, even if they are studied as a group. In fact, they often operate at cross purposes, particularly under severe congestion. This occurs because no integrated overall control strategy attempts to simultaneously optimize the performance of traffic flow both on the street and on the freeway. Control of on-ramp queue spillover on the local street network and prevention of off-ramp queuing on the freeway are obvious examples of times when integrated on-line control would be highly desirable [21, 22].

Virtually all area traffic control systems in use today employ so called First Generation Control, in which pre-stored plans are called up on either a time-of-day or traffic responsive basis. The timing plans for the splits and offsets are usually determined from an off-line traffic signal optimization program such as TRANSYT or SIGOP.

Many freeway control systems also use pre-stored time-of-day plans (metering rates, in this case), although some use traffic responsive control. The demand responsive plans are based on human observation or data from detectors that monitor basic freeway flows. However, there are some off-line programs that help determine optimum metering rates (e.g., FREQ3C). What is not clear is how, if at all, these two approaches could be combined for integrated control. Most likely, a new approach is needed.

A current goal of many freeway operators is the development of an on-line computer program for integrating freeway and arterial traffic control. In such a system, both parallel and perpendicular arterial traffic must be considered for integrated control of the freeway and the local network. At this time, only the integrated motorist information system (IMIS) approaches this type of integration.

CHAPTER 3
OFF-LINE SIGNALIZED INTERSECTION TRAFFIC CONTROL MODELS

A number of algorithms and computer models have been developed to aid traffic engineers in the design of signal timing plans for cycle-based, pretimed control of traffic signal systems. These algorithms and programs were developed to function off-line (as opposed to in realtime or on-line).

An overview of off-line computer models for urban streets is presented below. The models presented are the state-of-the-art models in this area of traffic control systems. The information provided below is taken from the extensive literature on the subject.

Computer models for traffic operation analysis have been comprehensively reviewed by Lieberman [53], Courage and Wallace [27], Byrne, et al. [13], and Ross [79]. A summary of these reviews is shown in Table 3-1. In addition, paired comparisons have been made in several studies. For example, SOAP was compared with NETSIM [65], MAXBAND with TRANSYT [24], TRANSYT with SCOOT [76, 74], and TRANSYT with SCAT [56].

The Signal Traffic Control models to be discussed in this chapter include the following:

- SOAP 84,
- TEXAS,
- MAXBAND,
- PASSER II (84),
- TRANSYT-7F,
- SIGOP III,
- NETSIM,
- NETSIM/BPS, and
- PASSER III, Diamond Interchange model.

Table 3-1. Summary of Off-line Models

	Location Application		
Model	Isolated Intersection	Arterial	Arterial Network
SOAP	opt. det. TS macro		
TEXAS	sim. TS. micro		
MAXBAND		opt. det. TS. macro	
PASSER - II	opt. det. TS. macro	opt. det. TS. macro	
PASSER - III		opt. det. TS. macro	
TRANSYT-7F		opt. det. TS. macro	opt. det. TS. macro
SIGOP III		opt. det. TS. macro	opt. det. TS. macro
NETSIM	sim.stoc. TS. micro	sim. stoc. TS. micro	sim. stoc. TS. micro

opt.	optimization model
sim.	simulation model
det.	deterministic model
Stoc.	Stochastic model
TS.	Time Scan
macro	macroscopic model
micro	microscopic model

SOAP-84 MODEL

Development and Description

The Signal Operations Analysis Package (SOAP) was developed in 1977, with updates in 1982 and 1984, by the State of Florida and the University of Florida Transportation Research Center. It is a computer model for developing signal timing plans for isolated intersections with specific attention paid to demand actuated signals. A wide range of control alternatives can be evaluated, including pretimed or multiphase-actuated control. The typical physical condition analyzed by SOAP is a four-legged intersection with left turns, through traffic, and right turns.

The program can evaluate the effect of a signal in an interconnected system by specifying a "platoon concentration factor" that results from signal progression, but the model is not really intended for use with signals in coordinated groups. SOAP is programmed in FORTRAN and can be run on both mainframe and microcomputers. The most recent development of SOAP is that it is included as part of a newly integrated system — The Arterial Analysis Package.

SOAP is an optimization model. It determines optimum signal timing and phasing in a three-step process of design, analysis, and evaluation. The "design" function of SOAP examines all legitimate phasing sequences for a given intersection configuration and traffic conditions and selects the one that can be executed with the least amount of green time. This design is then returned to the user. The next step is controller dial assignment and timing. The user must decide on the number of traffic patterns to be analyzed and can assign them to the appropriate dial. SOAP can also assist with this function. If traffic-actuated control is implemented, no traffic pattern assignments are made, and SOAP makes its computations accordingly.

For fixed time signals, optimum cycle length is determined by SOAP's optimization logic to minimize total intersection delay, subject to constraints on the amount of queuing that can be tolerated. For actuated signals, SOAP maximizes saturation. In most cases the cycle selected is the minimum cycle time permissible by traffic volumes on the target traffic flow plus minimum green phases on minor movements. The green is then allocated among conflicting movements based on the critical traffic movement.

SOAP computes measures of effectiveness that are common to traffic-control systems analysis (primarily delay). This allows the user to quantify the effects of different signal control strategies.

A comparison study between SOAP and NETSIM by Nemeth and Mekemson [65] concluded that results from both models were almost identical. In the case of a two-phase pretimed signal, NETSIM delays were somewhat higher, as expected, and the relative changes in delays corresponding to relative changes in volumes and left turns were similar.

In the case of a two-phase, fully actuated signal, the difference between NETSIM's and SOAP's average delay was higher than in the other two cases, but the difference can be explained by a unit extension specified in NETSIM that was too long.

The patterns of NETSIM and SOAP delays were similar enough to indicate that with additional research the correlation could be further improved. After differences in definitions had been accounted for, NETSIM and SOAP fuel-consumption estimates were found to be identical for all three cases.

Model Input/Output

SOAP requires three types of input cards. These are

- instruction cards that tell SOAP what to do,
- parameter cards that tell SOAP how to do it, and
- data cards that supply the input variables for the intersection being studied.

The input formats are standardized so that all cards have an identical format. This permits the use of a standard coding form, although all fields are not always used. Each card is identified by a single word in the first field that indicates to the program the meaning of the data contained in the subsequent fields. This simplifies the preparation of inputs considerably by eliminating the need for a specific sequence of cards, and SOAP will accept the cards in any order in which they are presented. This scheme has also been employed in the Arterial Analysis Package and MAXBAND programs, both of which are discussed later.

In the analysis function of SOAP, several measures of effectiveness are computed, including delay, stops, fuel consumption, volume-to-capacity ratio, and left-turn conflicts. This allows the user to quantify the effects of either the designed timing or any other timing scheme. The evaluation function of SOAP produces comparisons of the different design schemes.

Three primary types of outputs are available from SOAP:

- Input Report — echoes the input data and prints warning and error messages as appropriate;
- Design Recommendations — prints phase sequences and lengths, cycle lengths, and dial assignments; and
- MOE Report — provides infomation on delay, degree of saturation, maximum queue length, percentage of stops, excess fuel consumption, and left-turn conflicts.

Other supplementary outputs are available in both tabular and graphical forms to aid in detailed analysis.

TEXAS MODEL

Development and Description

The Traffic Experimental and Analytical Simulation (TEXAS) model is a simulation evaluation tool for isolated intersection signal analysis. It can be used to evaluate existing or proposed intersection designs and assess the effects on traffic operations of changes in roadway geometry, driver and vehicle characteristics, flow conditions, intersection control, lane control and signal timing plans [73]. TEXAS was developed by the Center for

Highway Research at the University of Texas at Austin for the Texas Department of Highways and Public Transportation in 1977. TEXAS-II is the most recent version. It is programmed in FORTRAN and can be operated on both mainframe and minicomputers [50].

The TEXAS model is a microscopic, time-scan simulation model. The model does not recommend design decisions; rather it provides rigorous analysis of a signal system based on input conditions. The user can evaluate alternative designs by performing several simulations with varied input parameters or data.

The model has three main subprograms. The geometry processor, GEOPRO, translates the user input data and program default values. The driver-vehicle processor, DVPRO, randomly generates the individual driver-vehicle units based on a variety of user data and program default values. The driver characteristics and vehicle generation are treated stochastically. The simulation processor, SIMPRO, microscopically processes each driver-vehicle unit through the intersection in a fixed, discrete time increment and accumulates data on vehicle performance and traffic interaction.

In TEXAS II, a new subprogram, EMPRO, was added [Lee]. EMPRO, incorporates algorithms to predict vehicle emissions of carbon monoxide, hydrocarbons, nitrogen oxides, and fuel flow for both light-duty vehicles and heavy-duty vehicles. EMPRO uses information from SIMPRO about the speed and acceleration of each vehicle to compute the vehicle emissions and fuel consumption at all points along the vehicle path.

One significant limitation to the application of TEXAS discussed in the literature is that the model does not calculate the interaction between pedestrians and vehicles moving simultaneously.

Model Input/Output

The TEXAS model will produce estimates of fuel consumption as well as vehicle exhaust emissions when the Center for Highway Research finishes work on adding these measures of effectiveness to the traffic simulator. This work was scheduled for completion in 1985 but was not reported in the literature reviewed for this report. Also scheduled are changes in input structure to provide interactive data input and program changes to make the model transportable between computer systems.

MAXBAND

Development and Description

MAXBAND is a bandwidth optimization model that develops signal timing plans for arterials and triangular networks. The optimization algorithm used in MAXBAND is based on a mixed-integer programming formulation which has been described by Little. MAXBAND was written in FORTRAN IV for use with IBM mainframe computers and has been recently transfixed to microcomputers. It was developed by the Massachusetts Institute of Technology for FHWA, and can handle as many as 20 signals efficiently.

The MAXBAND program uses as its traffic model the maximum green bandwidth principle. This is combined with a powerful mathematical programming algorithm, mixed integer linear programming, to obtain a cycle length, offsets, and left-turn phase sequences that maximize the weighted sum of bandwidths in both directions on an arterial. The program also has the capability to allow small deviations from the progression speed on individual links. This process is referred to as a speed search.

Some of the features of MAXBAND are as follows:

- cycle length is treated as a continuous variable within a specified range,
- design speed can vary within specified limits,
- the best phase sequence at each intersection is automatically selected from a specified set,
- queue clearance time is allowed to permit secondary flow accumulated during red phases to discharge before the arrival of the next platoon, and
- the model accepts user-specified weights for the green band in each direction.

Unlike TRANSYT, the MAXBAND program obtains a global optimum, requires no starting solution, and optimizes cycle length and phase sequences. However, according to the literature, MAXBAND has the following deficiencies [23].

- The traffic model is oversimplified. It does not consider flows turning from side streets, platoon dispersion, turning traffic, or platoon shape. Because of these factors, it is not generally true that maximizing bandwidth minimizes measures of effectiveness such as stops or delay.
- Green phase times are not optimized. This is because bandwidth provides no criteria for setting green times on the side street.

Model Input/Output

The basic inputs to MAXBAND include the range of permissible cycle lengths, the geometry of the different links, traffic flow rates, saturation flow rates, permitted phase sequences, queue clearance times, and ranges of speeds. Outputs include a data summary report and a solutions report that contains cycle length bandwidths, selected phase sequencing splits, offsets, link speed, and travel time [6, 16].

PASSER II (84) MODEL

Development and Description

The Progression Analysis and Signal System Evaluation Routine (PASSER) is a macroscopic, deterministic optimization model for calculating progression along an arterial street with various multiphase sequences. Improvements in the original processing algorithms and measures of effectiveness have been made by the Texas Transportation Institute, and the current version of the model is known as PASSER-II (84). Passer is incorporated into the Arterial Analysis Package (AAP).

PASSER-II (84) can handle up to 20 signalized intersections along a single arterial street, with up to four phase sequences per intersection. It is written in FORTRAN IV for use on 16/32 bit computers and most microcomputers. The program is currently maintained by the Texas Department of Highways and Public Transportation [78].

PASSER II is essentially a time-series, search-and-find optimization routine. The PASSER-II model combines Brook's Interference Algorithm with Little's Optimized Unequal Bandwidth Equation, and extends them to multiphase arterial signal operations. The model calculates phase intervals, offsets, and movement demand/capacity ratios to evaluate the level of service at each intersection. It finds green times by proportioning time according to the volumes plus lost time (subject to the minimum required greens).

Model Input/Output

The model inputs include turning movements, saturation capacity flow rates, distances between intersections, average link speeds, and the minimum green times that must be provided at each intersection. The program first uses these inputs to determine splits. Trial cycle lengths, phase patterns, and offsets are varied to determine the optimal set of timings that maximizes the progression bandwidth.

Inputs to PASSER II involve three types of data records:

- arterial header records that specify the global system parameters;
- intersection header records, each of which specifies the operating parameters for one of the intersections in the system; and
- intersection data cards that provide the traffic volume, saturation flow, and minimum green time for each approach to every intersection.

Three types of outputs are available from PASSER II:

- the input data report, which plays back all input data;
- the design recommendation report, which includes cycle length, offsets, phase sequences and splits, and MOE values for bandwidth efficiency and degree of saturation; and
- a time-space diagram.

TRANSYT-7F

Model Development and Description

The Traffic Network Study Tool (TRANSYT) is one of the most widely used models in the United States and in Europe for network signal timing design. It was developed in 1968 by Robertson of the Transport and Road Research Laboratory (TRRL) in England, and since then, the TRRL has released 8 versions of this model [93]. The version discussed below is TRANSYT-7F, in which the "7" denotes the seventh TRRL version of TRANSYT, and "F" symbolizes that this is the Federal Highway Administration's version of TRANSYT-7, which uses North American nomenclature and units on input and output. The TRANSYT-7F model is written in FORTRAN and is available for most microcomputers. FHWA has also had a comprehensive user's manual written to serve as an instructional guide for traffic engineers who desire to use the model.

TRANSYT-7F optimizes signal timing on coordinated arterials and grid networks. The mode of signal control considered by TRANSYT-7F is pretimed, with two to seven phases and fixed-phase sequence. More specifically, the model aims to identify the optimal offsets and phase splits that result in the minimization of a performance index — a linear combination of stops and delay. Weights may be supplied within the performance index process by the user on an individual link basis to favor either stops or delay and to increase the importance of certain links if desired (e.g., to prioritize movement on an arterial). The phase sequence at each intersection must be supplied and is not optimized.

TRANSYT is used extensively around the world and has gained considerable popularity in the U.S., mainly because it has been adapted at the University of California, Berkeley, (1977) and at the University of Florida (1983) to make it more suitable for American cities. TRANSYT has been incorporated in some newly developed integrated systems, such as AAP, ITDS, and TRAF. Also, TRANSYT has been used to test the performance of on-line, traffic-responsive traffic control systems. The most recently released version of the model is TRANSYT-8. Major enhancements of TRANSYT-7 in this upgrade include the addition of gap acceptance features and the implementation of a cycle search routine.

The TRANSYT model is macroscopic and deterministic, using a fixed time scan analysis process. The platooning effect caused by signals is modeled in a simulation subroutine. The volume of traffic entering each intersection per cycle is assumed to be a known constant, supplied by the user, as is the proportion of traffic turning either right or left. The mean flow leaving each signal is modeled by a histogram, which determines the number of stops, the delay, and hence the performance index. These calculations are based on link length and a mechanism to capture the random element in driver behavior and link saturation (which contributes to platoon dispersion). The TRANSYT simulation model has been used extensively in the field by researchers in many countries, and the literature indicates that it accurately reflects observed behavior.

TRANSYT-7F consists of two main parts:

- a macroscopic, deterministic traffic flow model, which computes the value of a specified performance index for a given signal network and a given set of signal timings, and
- a hill-climbing optimization procedure that makes changes to signal timing (splits and offsets) to determine whether the performance index can be improved by that change.

The model's minimization subroutine first calculates the performance index for the initial set of signal timings (which can be supplied either by the user or by the program). The offset of one signal is then altered by one step and the effect on the performance index is determined. If the index is smaller, the offset is altered by another "jump" in the same direction. Similar jumps in the offset are made until no further improvement in the index is found. If the initial step does not produce a reduction in the index, the offset is altered in the opposite direction by a series of jumps. The magnitude of successive jumps is given as part of the TRANSYT program but can easily be altered by the user. Once a local minimum (which may or may not be a global minimum) has been found for the index with respect to this first offset, the offset is held fixed at the "minimum" value and another offset is altered in the same way. Each offset is eventually altered in this way, in a specified order. This whole process is repeated a number of times in order to avoid minima which are local but not global. Phase splits may also be altered in the same way, if desired, to attempt to minimize the index. The user also has the option of asking the model to examine specified cycle lengths. TRANSYF-7F compares the results of alternative cycle lengths in its output results.

Model Input/Output

There are up to 20 types of input records for TRANSYT (depending on the version). The inputs fall into five functional categories, including data that

- are common to the entire network,
- control the optimization process,
- specify traffic data,
- specify signal timing, and
- specify plots.

TRANSYT input data are based on a link-node structure. This structure is considerably more complicated conceptually than the single intersection orientation of the non-network models such as SOAP-84. User training is therefore a significant factor with TRANSYT. Training has been addressed through a series of courses sponsored by FHWA. TRANSYT-6C and TRANSYT-7F have both been covered in these courses.

Five output reports are available from TRANSYT-7F:

- Input Data Report — a structured echo of input data, including any errors or warning conditions detected;
- Performance Table — a listing of significant data and MOEs, including (by link) volume, saturation flow, degree of saturation, total travel and travel time, delay, stops, fuel consumption, maximum back of queue, and green

times (subtotals are given by intersection and aggregated for the entire network);

- Signal Timing Tables — for each intersection the offset (or yield point) is given along with the signal timing in terms of individual interval lengths;
- Flow Profiles — graphically show the arrival and departure flow patterns; and
- Time-Space Diagrams — available for any number of routes desired.

The TRANSYT-7F postprocessor converts signal timing from the unfamiliar British scheme originally used in TRANSYT to conventions commonly used by engineers in the Americas and Canada. In addition, TRANSYT-7F has a number of options that the user can control. These options include the following.

- Buses can be modeled separately by including bus links. These can either be separate lanes or shared lanes.
- Right/left turn delays caused by pedestrians can be reflected.
- Large networks can be subdivided into sections that the program can handle (within 50 nodes and 250 links). The boundary nodes are fixed from section to section so that their timings are not changed in the subsequent analysis. Another alternative is the expansion of program dimensional arrays to accommodate a larger network.
- Unsignalized intersections controlled by stop signs on the cross-streets and bottlenecks can be modeled.
- An estimate of network fuel consumption can be computed based on total travel, stops, and delay. The fuel consumption value includes fuel consumed for cruise, idle, and acceleration or deceleration. Fuel consumption estimates are calculated for each link and then summed for the entire network.

Potential TRANSYT-7 Improvements

The minimization procedure just described is essentially the rectangular, or one-at-a-time, search, which uses a hill climbing process to search over each variable in order to minimize the value of a multivariable discrete function. The function is the performance index, and its variables are the offsets and phase splits. A possible improvement would be the use of more efficient multivariable optimization strategies to increase the efficiency of the minimization procedure. Timmermans, et al.(1979), and Nelder and Mead (1965) concluded that, due to the nature of the performance index function, the rectangular search is superior to other methods. However, these authors do suggest that, within the strategy of the rectangular search currently employed by TRANSYT, a slight modification of the Fibonancci search would be more efficient than the hill climbing, one-dimensional search currently used [33].

Use of TRANSYT

During 1981, TRANSYT-7F was tested in 11 cities in the United States, where it was used to produce optimal signal timing plans for 520 intersections in coordinated signal systems. These cities contracted with the FHWA to undertake a project to optimize the signal timing in a portion of their street networks and to evaluate the effectiveness of the optimized signal timing plans. Although the evaluation data are limited, the cities reported unanimous agreement on the positive value of the TRANSYT-7F model for use in retiming traffic signals. In addition, some extensions were developed to enhance TRANSYT-7F's capability [98].

In a comparison between MAXBAND and TRANSYT, Cohen made several conclusions. These points are listed below.

The TRANSYT program includes an excellent traffic model that uses network geometry and traffic flows estimate two measures of effectiveness (MOEs) — delay and stops. The hill-climbing optimization procedure adjusts offsets and green times separately to minimize the value of a performance index (PI), which is equal to the weighted sum of stops and delay.

Although field tests and simulation tests indicate that TRANSYT produces good signal timing plans, it also has a number of deficiencies.

1. The hill-climbing optimization algorithm does not generally guarantee that a global optimum for the PI will be achieved and therefore does not guarantee that the "best" signal timing plan will be found. This is because the signal timing problem in general has a solution space for the PI, which consists of a number of local optima. It is computationally infeasible, when using the hill-climbing technique, to search through all local optima to find the best one.
2. TRANSYT requires a signal timing plan as a starting solution. Because of item 1 above, the quality of the final signal settings often depends on the starting solution.
3. However, because of item 1 above, there is no way of knowing whether the selected cycle length is the best one or whether, for that cycle length, a solution was found that was closer to the global optimum than the solutions found for the other cycle lengths scanned.
4. The sequence of left-turn phases and through phases is not optimized. At signalized intersections where left-turn phases are used, there are four possible combinations for the left-turn phases and through phases in both directions: (a) left-turn phases in both directions preceding the two-directional through phase (lead-lead), (b) left-turn phases in both directions preceding the two-directional through phase (lag-lag), (c) a left-turn phase in the inbound direction preceding the two-directional through phase and a left-turn phase in the outbound direction following the two-directional through phase (lead-lag), and (d) a left-turn phase in the inbound direction

following the two-directional through phase and a left-turn phase in the outbound direction preceding the two-directional through phase (lag-lead) [23, 24].

Other comments on TRANSYT in the literature are that the performance of the model obviously relies heavily on the quality of the input data and the efficiency of the network model. The choice of plan change times and the duration to be covered by a single plan is left to the user of TRANSYT. Without an enormous effort and extensive field validation the optimal plan change times and durations cannot be determined.

While TRANSYT usually produces green splits that appear sensible to the experienced traffic engineer (given the ability to model accurately), the same cannot be said of the offsets it produces, which at some times appear inappropriate. The temptation to trim these offsets in the field must be tempered by the knowledge that the alteration may well be sub-optimal. It is difficult to validate, by remodeling, changes to any offsets that appeared to be inappropriate. In summary, TRANSYT is regarded as a logical and theoretically sound program. Its success has been demonstrated in a number of research studies. The strength of the program lies in its traffic-flow model, which accurately treats flow patterns from signal to signal and calculates the effects of platoon dispersion by means of a platoon production model.

SIGOP-III

Model Development and Description

SIGOP-III is an acronym for Traffic Signal Optimization Model, Version III. It was developed by KLD Associates, Inc., as an outgrowth and refinement of the original SIGOP model developed in the mid-1960s. The similarities between TRANSYT-7F and SIGOP-III are (1) both models are macroscopic signal timing and analysis tools, and (2) both models contain a traffic flow submodel and an optimization submodel that minimizes a user-specified "disutility" function. TRANSYT-7F considers delay and stops, whereas

SIGOP considers delay, stops and "spillover," a term for queuing. SIGOP is written in FORTRAN IV originally for use with IBM 360/370, CDC 6600, and Amdahl 470 computer systems. SIGOP can optimize a network of up to 50 intersections and 130 links, and a single link can have up to three movements.

SIGOP extends the underlying principles of TRANSYT while reducing the effort required to use the model. Furthermore, the following additional considerations were desired in the design of SIGOP:

- a faster optimization procedure,
- explicit representation of turning bays,
- explicit consideration of queue buildup and prevention of spillover, and
- production of estimates of fuel consumption and time-space diagrams.

SIGOP uses a periodic time scan as part of its optimization procedure. The optimization method used is similar to TRANSYT; however, at each gradient search step, only the intersections adjacent to the "current" intersection are re-analyzed for impact. This technique is referred to as the "method of successive approximations." Although this procedure results in significantly reduced execution time, the simplification may possibly sacrifice some confidence in the model's "optimal" solution. A major improvement over TRANSYT is the explicit inclusion of a queue length term in the optimization objective function. This term is designed to prevent spillover, which is not assured in TRANSYT. Although similar to TRANSYT, the traffic simulation model has also been simplified. All platoons are assumed to be either "main street" or "cross street," and differences in departure times from multiple upstream sources are not explicitly considered.

SIGOP contains a split calculation routine in which green times at each signal are computed in proportion to their respective critical-lane flows, total approach flows, or a combination of both. It also contains an offset optimization algorithm that minimizes the discrepancy between the actual signal offsets and a set of given or calculated ideal offsets. The resulting optimized settings are evaluated in terms of delay, stops, and cost values.

Model Input/Output

SIGOP requires more extensive data than arterial and single intersection models, but not as much input data as TRANSYT. Thirteen types of input data records are available to SIGOP. They fall into the same functional categories as discussed in the TRANSYT section of this chapter. The significant differences between SIGOP and TRANSYT inputs are as follows:

- SIGOP does not require link-to-link flows as does TRANSYT;
- signal phase sequences are coded from preset tables, which reduces the coding effort but which also reduces flexibility. SIGOP also limits the number of phases to four, while TRANSYT does not;
- SIGOP requires input nodes for external links, while TRANSYT does not; and
- in SIGOP, diagonal approaches may be coded, but their movement must be coincidental with another normal movement (in TRANSYT, all movements may be modeled explicitly and independently).

The basic inputs to SIGOP include flow rates, saturation flows (in terms of headways), minimum green times, yellow times, special phase times, and passenger car equivalent factors for trucks, buses, and turning vehicles.

Outputs include the data summary report, the signal timing report (which contains offsets and splits for each phase), the performance analysis report (which shows the value of the disutility function for each iteration of the model, including the optimal value and detailed performance measures, for each link and for the network as a total), and user-specified, time-space diagrams.

Like many signal timing computer models, SIGOP has advantages and limitations. One of the major advantages of this model is the multiple cycle-length evaluation capability. This can save the designer a considerable amount of time that would ordinarily be spent in preparing and running several jobs.

The major limitations of SIGOP include the following.

- Each signal cycle can accommodate a maximum of four phases, which does not adequately serve some users.
- The model does not consider buses explicitly.
- Permissive and unprotected turns are not addressed explicitly by SIGOP. However, these conditions can be included to some extent by manipulating model parameters.
- The model does not explicitly consider nonsignalized intersections (stop sign control, for example).
- The model lacks extensive field testing and evaluation.

<u>Comparison of SIGOP and TRANSYT</u>

Four off-line traffic signal optimization techniques (SIGOP, TRANSYT, Combination Method, and a preferential-street program that is SIGOP-based) were evaluated in both a suburban area and central area networks within metropolitan Toronto [Rach, Leonard]. The Toronto experiment concluded that both SIGOP and TRANSYT did not result in statistically significant improvements over the existing pretimed system in terms of travel time. Thus, neither TRANSYT or SIGOP was judged superior in the Toronto study. The lack of success of both models was due perhaps to the use of the models without prior calibration of some of the program parameters for local conditions (such as the smoothing factor used in the platoon dispersion model), and simply as a result of the difficulty in statistically proving travel time benefits in signal networks.

The Toronto study did indicate that TRANSYT needed more computer time and person hours for input preparation than SIGOP. SIGOP needed 12 minutes per computer run and 710 person hours, whereas TRANSYT needs 165 minutes per computer run and 926 person hours.

NETSIM MODEL

Model Development and Description

NETSIM (NETwork SIMulation Model) is an extension of the Urban Traffic Control System (UTCS-1) model, which incorporated and expanded on two earlier models, DYNETG and TRANS. NETSIM is written in FORTRAN IV for both mainframe and microcomputers. It can handle as many as 99 nodes, 160 links, and 1,600 vehicles at any instant. The model has been the subject of extensive field testing, evaluation, and validation, and it is one of the most widely used network simulation models available today. NETSIM is maintained by the Federal Highway Administration. It is designed primarily to provide the engineer with a powerful tool for analyzing and evaluating a wide range of traffic control and surveillance concepts for complex street networks [19, 51, 85].

The NETSIM model is a microscopic, stochastic simulation model. NETSIM treats a street network as a series of interconnected directional links and nodes. Each link represents a particular approach to a node, and changes in link characteristics can be modeled by inserting artificial mid-block nodes. A link may contain up to five lanes of traffic plus left and right turn pockets. Traffic generators, such as parking lots and minor streets, may be modeled as sink or source nodes. Along links, vehicles are processed in a time-scan format subject to alternative traffic controls imposed at the nodes (yield or stop sign, signal with pretimed or traffic-actuated controllers, etc.). Vehicle motion is governed by a series of car-following, queue-discharging, and lane-changing algorithms.

Some enhancements to NETSIM are currently being developed, including an interactive graphic capability called NETGRAF. NETGRAF will enable the engineer to use the postprocessor function of NETSIM more effectively and will provide on-line graphical comparisons between the measures of effectiveness obtained from each NETSIM execution run. Other enhancements have also been reported by Liberman. KLD Associates, Inc., performed an informal survey of NETSIM users to determine the desire for enhancements to NETSIM. The survey results indicated four possible enhancements:

- easing the input data preparation effort,
- reducing the computational cost of NETSIM execution,
- extending the output capabilities, and
- adding more features that have been introduced by the users.

NETSIM/BPS

NETSIM/BPS is the basic NETSIM model modified to simulate bus preemption at intersections. This logic provides the means to advance or extend signal phases that service oncoming buses. Two sets of logic are provided: one set for a far-side bus stop, the other set for a near-side bus stop. The program can simulate strategies for green extension, red truncation, cycle-skipping, and the cross-street/mainstreet preemption selection. At this time, the model will process 20 nodes, 40 detectors, and 40 links operating in a bus-preemption environment.

While NETSIM/BPS incorporates bus preemption in the evaluation process, this version of the program lacks the ability to simulate preemption systems with either single or dual ring actuated controllers.

Model Input/Output

Inputs to NETSIM include network geometry, traffic flow rates, saturation flow rates, turning movements and counts, traffic composition, type of signal controller, mode of operation, and timing settings. Output includes a variety of measures of effectiveness such as delay, stops, cycle failures (for pretimed signal controllers), fuel consumption, and emissions. NETSIM does not perform design optimization; rather it evaluates the effectiveness of different alternative designs specified by the user's input.

The input data for NETSIM/BPS consists of the basic parameters of the original NETSIM program plus other parameters related to the preemption logic, such as

- green extension durations,
- red truncation durations,
- cycle length,

- detector surveillance parameters, and
- bus trace parameters.

Program output includes measures of effectiveness normally of interest to researchers (speed, delay, etc.), estimates of fuel consumption, emissions for each vehicle type, and bus stop statistics.

DIAMOND INTERCHANGE MODEL — PASSER III

Model Development and Description

The Progressive Analysis and Signal System Evaluation Routine, Model III (PASSER III) was developed by the Texas Transportation Institute at Texas A&M University in the early 1960s [Messer]. The basic concept of the model is to develop a signalization strategy for diamond interchanges that increases capacity of the system by allowing several potentially conflicting movements at the separate intersections to occur simultaneously for a short time. The model is used for analyzing signal phase needs and determining the optimal traffic signal timings at signalized diamond interchanges. PASSER III considers two basic diamond interchange configurations:

- the isolated interchange mode (signalized diamond interchange with or without frontage roads), and
- the progressive frontage road mode (series of interconnected interchanges with progression on the frontage roads).

PASSER III was originally written in FORTRAN IV for use with IBM 370 computers. The interchange analysis portion was adapted in 1984 for microcomputer application [Courage]. The model has been tested extensively and is maintained by the Texas Department of Highways and Public Transportation.

PASSER III is a time-based, macroscopic, deterministic optimization model. In the isolated interchange mode, PASSER-III can analyze five identified phasing patterns, including the four-phase, two-overlap sequence and different variations of the three-phase sequence patterns. The model examines all possible combinations of phasing patterns and offsets to minimize the total delay in the interchange. In the progressive mode, the model determines optimal cycle length by maximizing the progression bandwidth. Progression may be one-way or two-way (with or without preference to one direction). Both the isolated interchange mode and the progressive frontage road mode can be run together or separately to examine alternative solutions. The optimal progression design is that which provides the largest bandwidth efficiency [60].

Although PASSER-III is designed primarily to study fixed-time and fixed-sequence control, the delay-offset analysis built into the model can be used to study fully-actuated phasing and to determine the effects of different interchange approach lane configurations, left-turn configurations, and U-turn provisions.

Model Input/Output

The inputs of the model include turning movement volumes, saturation capacity flow rates, distances between intersections, average links speeds, queue clearance intervals, phasing sequences, and minimum green times that must be provided at each intersection.

Output from the model includes a data summary report, phase-interval report, optimal progression solution report, general signalization information, frontage road progression information, and time-space plots. The output varies somewhat according to the mode (isolated and progressive) being analyzed.

CHAPTER 4
FREEWAY SIMULATION MODELS

INTRODUCTION

Numerous freeway simulation models have been developed since the 1960s. This chapter will summarize the capabilities of the most useful freeway system models. Most of the information provided in this chapter comes from work previously done by Hsu [40] and May [58]. Fifteen early freeway simulation models were reviewed by Hsu and Munjal in 1974. The 15 reviewed models were as follows:

(1) Arizona Transportation and Traffic Institute Traffic Simulation Model

(2) Midwest Research Institute Freeway Simulation Model

(3) Midwest Research Institute Mountainous Terrain Model

(4) Northwestern University Lane-Changing Model

(5) Sinha Freeway Simulation Model

(6) Connecticut Department of Transportation Expressway Simulation Model

(7) Texas Transportation Institute Freeway Merging Model

(8) System Development Corporation Diamond Interchange Model

(9) System Development Corporation Freeway simulation Model

(10) Mikhalkin Freeway Simulation Model

(11) Georgia Model

(12) SCOT Corridor Model

(13) Priority Lane Model

(14) Aggregate Variable Model

(15) Aerospace Corporation Freeway simulation Model

Among the 15 reviewed models, the Midwest Research Institute Mountainous Terrain Model dealt specifically with mountainous terrain, and the System Development

Corporation Diamond Interchange Model dealt specifically with diamond interchanges rather than conventional freeway operation.

Hsu compared the remaining models against a baseline of eight features that he regarded both desirable and independent of specific simulation purpose. The eight baseline features were

- realism for representing freeway flow phenomena,
- features built into the models to handle anticipated applications,
- logic complexity,
- computer running efficiency,
- extent of model evaluation,
- flexibility and expandability,
- suitability for incident detection and ramp control, and
- completeness of program documentation.

A more recent review of freeway models was made by May [58] in 1982. May grouped the currently available freeway corridor models into five model families, CORQ, FREQ, INTRAS, MACK, and SCOT. As a matter of fact, many of the models reviewed by Hsu were early versions or parts of current models. For example, INTRAS was based on the Northwestern Model, and the SCOT Corridor Model was the early version in the SCOT family. The Priority Lane Model was an early priority-lane version in the FREQ model family, and the Aggregate Variable Model was an early version in the MACK model family. On the basis of May's review, the current five families of freeway simulation models are described below. A summary of each model family's characteristics is shown in Table 4-1.

CORQ-CORCON MODEL

CORQ and CORCON are the two packages in this model family. Freeway models are usually either microscopic or macroscopic. The macroscopic models deal with freeway

Table 4-1 Freeway Simulation Models Characteristics

Model	Model Type
CORQ-CORCON	macro. det. sim.
FREQ	macro. det. opt.
INTRAS	micro. sim. det.
MACK	macro. det.
SCOT	macro. (platoon), det. sim.

opt.	optimization model
sim.	simulation model
det.	deterministic model
macro	macrorigin model
micro	microrigin model

operations and characteristics in an aggregate manner by ignoring detail in favor of analytic simplicity and efficiency. Microscopic models treat the freeway and the street network in greater detail and thereby achieve greater precision. The CORQ-CORCON models conceptually lie somewhere between micro and macro.

CORQ System Development and Description

The CORQ model was developed by Yagar between 1968 and 1975 for simulating time-varying traffic demands in a freeway corridor [Yagar 54]. CORQ is written in FORTRAN IV. It is intended to help the traffic analyst assess the systemwide effects of any traffic-control strategies proposed for a network, as long as the total system's demand remains invariant or, at least, responds predictably to the controls. (That is, it is meant to work when the controls don't change traffic demand during the time period modeled.) This model addresses in detail the critical elements of a corridor in terms of traffic flow, capacity, queuing and delays. It is related to another model, FREQ, which places more emphasis on the modeling of freeway queues.

The CORQ model process divides the peak period to be modeled into a set of short, uniform slices of time so that the rates of demand between the various O/D pairs can be considered constant. This allows time-varying demand to be expressed as a set of constant O/D matrices that represent the respective time slices, with each slice having stationary demands. The O/D matrices are assigned to the network sequentially in time.

This technique allows temporary oversaturation of network links. That is, in any time slice, certain network links may have more demand assigned to them than they can serve. Excess vehicles queue on upstream links and are reassigned to their destinations in the succeeding time slices from the points at which they queued. The assignment is based on the principle of minimum individual travel cost, and the minimum cost path may include some time in queue. CORQ does not account for changes in corridor O/D demands because of changing operating conditions caused by traffic controls. The user is left to alter the origin-destination matrices to account for those demand changes that he or she feels are required.

CORQ has the capability of changing network characteristics at the beginning of each time slice. In this manner, both capacity variations and demand variations (for example, those that simulate transient traffic controls such as time-varying, ramp-metering rates) can be accounted for by the model.

Three assumptions are inherent in CORQ's methodology [99, 100]:

- <u>Queue Dissipation</u> — A queue that dissipates in a certain time slice is assumed to decrease at a constant rate over the entire length of that time slice and thus disappear at the end of the slice.
- <u>Queue evolution</u> — A queue that exists on a certain link at the end of a time slice is assumed to have been taken out of the network and is fed back in as new demand originating at the downstream end of that link. This new demand is fed in at a constant rate over the duration of the following time slice.
- <u>Driver's knowledge of travel times</u> — The model assumes that the driver knows the unit travel times of all the links for the present time slice but not for the next time slice. This means that the present best path can be chosen for the driver, but if that path leads to a queue, he or she will select the remainder of the path (i.e., a revised path) based on new information when he or she is ready to leave the queue.

<u>CORQ Input/Output</u>

The data required for CORQ are of the type generally collected when freeway control is studied. These include capacities, volume counts, queue sizes, travel times as a function of flow, and Origin-Destination (O/D) information on users who could or should be affected by controls. Unless these data are already available, the task of collecting the data solely for use in the model may be prohibitively expensive. The outputs of CORQ are predictions of the link flows, queues and the travel time for the entire freeway corridor.

Application of CORQ

The CORQ model was tested on Ottawa's Queensway corridor [Yagar 55]. The flows and queues that it initially predicted were reasonably close to those measured in the field. It was then calibrated to actual flows and queues and applied in testing alternative traffic-control schemes. It was further validated against observed flows when it demonstrated its sensitivity in modeling the effects of various strategies and its power in suggesting alternative paths around bottlenecks.

CORCON System Development and Description

On the basis of Yagar's work in developing the CORQ model, CORCON (the freeway CORridor assignment and CONtrol model) was developed by Allen and Easa [Easa, Allen]. CORCON is an analytical procedure for predicting traffic volumes, queuing conditions, and travel times in a freeway corridor. Essentially, the model structures of CORCON are similar to CORQ's, except that more functions were added. In particular, a new method of link-node representation was incorporated into CORCON that allows more than one directional roadway link to have common upstream and downstream nodes. This feature is particularly advantageous in simplifying the representation of complex merging, weaving, intersection, and interchange network sections [4, 5, 29].

Like CORQ, CORCON divides the modeled period into equal lengths of homogeneous demand called time slices. The demand in each time slice and the queued demand of the previous time slice are assigned to the network on the principle of minimized individual travel cost (time). The flow versus travel time relationship for each link is an increasing function and is approximated by three linear components. Network features are represented by a simplified link-node method. The model incorporates a procedure for turn prohibitions, a traffic diversion procedure, and a method for calculating turning volumes without the need for turning links.

During the development phase, particular attention was given to establishing a simple and efficient method of network representation. The selected link/node process minimizes the input data requirements and considers traffic diversion and queuing characteristics.

CORCON Input/Output

The data requirements for CORCON are similar to those for CORQ, namely, O/D demands, link volumes and queues (15-minute basis), flow versus travel time relationships, capacity information, and network turn prohibitions. Outputs are the predictions of the link flows, queues and the travel time for the entire freeway corridor.

Application of CORCON

CORCON was calibrated and validated on the Queen Elizabeth Way freeway corridor, southwest of Toronto. It was used to predict traffic operating characteristics in the corridor for peak-period traffic before and after the implementation of a freeway-access control strategy in 1976 [5, 29].

The validation showed that CORCON was capable of predicting traffic volumes, travel times, and queuing characteristics on a freeway corridor. The average difference between predicted and observed characteristics was 10 percent. In the study, this level of correspondence between predicted and actual traffic operating conditions was deemed sufficient to recommend CORCON for regular use as a planning tool for assessing alternative freeway-corridor control plans.

One limitation of the model is that because of dimensioning constraints the model cannot be easily applied to very large corridors (5 by 25 miles). Researchers must accommodate large areas by either increasing the dimension sizes, analyzing two or three subsections of the corridors separately, reducing the amount of coded network detail, or doing some combination of the preceding.

FREQ MODEL FAMILY

General Model Development and Description

The FREQ models are the most widely used freeway corridor models in the U.S. Among the users of FREQ is WSDOT, which used FREQ6PE in 1981 to evaluate TSM strategies for I-5 in Seattle and FREQ6PL to evaluate priority lanes on I-405.

Work on the first model of the FREQ family of models began in 1968 at the University of California at Berkeley. Since the creation of the first FREQ model, several extensions and refinements have been made, with particular attention directed to shock wave analysis, computer efficiency, emissions estimation, and output format. All programs are written in FORTRAN and can be run on either CDC or IBM mainframe computers. In the early 1970s, FREQ3CP, FREQ3D, and FREQ3C were developed. That incorporated, respectively, priority-entry control, design improvement, and normal-entry control optimization submodels.

The eighth version of FREQ is the latest released. In addition, a microcomputer version of FREQ8PE, FREQ8PC, can now be run on IBM AT or XT compatibles. All FREQ models are supported by ITS, University of California, Berkeley [7, 8]. The PC model requires a math co-processor and at least 512k of memory and is available on one high density diskette or three double density backup version diskettes.

Also recently developed to become available soon is the FREQ9PE model. The difference between these versions is that version 9 handles four times the capacity of version 8 (i.e., 80 on and off ramps instead of 20 and 160 subsections instead of 40).

The FREQ#PE model (where # is the version number of the software) is a macroscopic decision model of a freeway corridor. It is used primarily for the evaluation of priority-entry and normal-entry control on a freeway. The model can also be used for evaluating design improvements with or without freeway-entry control. The model predicts a time stream of impacts and travel responses.

FREQ#PE is a time-scan model. It progresses, by time slice, from roadway subsection to subsection, performing the following analyses:

1. Ramp analysis is performed to determine if a ramp queue exists, develops or dissipates, and the appropriate delays are calculated.
2. Volume calculations are performed using the input demands, O-Ds and any existing queues. If capacity is exceeded, the freeway and not ramps are queued.
3. Ramp merging analysis is based on the ramp volume inputs and estimated right-lane volumes on the freeway. Again, if the right lane exceeds capacity (because of ramp inputs), the freeway is queued.
4. Weaving analysis is confined to on-off ramp maneuvers, and capacity reductions are computed using techniques from the Highway Capacity Manual. Weaving effects in the area of the HOV lane entrance and exit must be accounted for by adjusting the mainline capacities in these roadway subsections.
5. Queuing analysis on the mainline takes into account the propagation of shock waves, whether traffic moves upstream or downstream, and adjusts volume versus demand accordingly.
6. The speed-flow analysis uses the Highway Capacity Manual curves to determine travel time related impacts based on the flow characteristics computed earlier. Additionally, users can input up to nine of their own curves, which may be specified for use in any subsection(s).

The latest versions of FREQ also allow the user to input metering plans, queue length limits, requests for congestion optimization, and metering over-control protection. In addition to the above features, FREQ8PE has the capability of generating, at the user's request, synthetic O/D matrices from ramp counts, based on a computer model called SYNPD2, also developed at ITS. It can also optimize fixed time ramp metering plans.

The FREQ#PL model is a macroscopic model of a freeway corridor used primarily for evaluating reserved lane(s) on freeways for carpools and/or buses. The model can also be used to evaluate design improvements with or without priority operations and to develop fixed time ramp metering strategies. The model automatically modifies the demand and supply sides of the model based on travel conditions and predicts traffic impacts and traveler responses by time period [20, 61].

A complete run of the FREQ#PL model consists of the following sequential steps:

1. The freeway simulation submodel is executed for the existing condition and impacts are reported.
2. The optimization submodel is executed to determine the optimal metering system that will maximize the user's selected objective function. The optimal design is output.
3. FREQ#PE is executed again with the results of the optimization submodel to compute the impacts of the specified control strategy, and the results are reported.

This sequence provides the user with impacts for the "before" condition and the "after" effects of the control strategy.

Model Input/Output

As with the other models, the data requirements to operate the FREQ models are quite extensive. The freeway is broken into homogeneous subsections that are usually stretches of freeway between on-ramps and off-ramps, or where changes in freeway characteristics (number of lanes, widths, etc.) occur. The physical characteristics of the freeway necessary for input are based on these subsections. For each subsection, the length, number of lanes, lane width, capacity, subsection truck factor, grading, length of grade, and design speed are needed.

In addition to the physical characteristics of freeway subsections, ramp characteristics are also considered. They are

- location of on-ramps and off-ramps,
- characteristics of special ramps (multilane ramps, left-hand side ramps, etc.),
- ramp metering limit and/or capacities.

The FREQ models also require demand data, which consist of the following:

- an origin/destination table for each time slice, which may be in the form of traffic counts or hourly rates and may be either measured or simulated; and
- passenger volume data. In some cases, it is desirable to measure passengers throughout (as in priority entry control simulation). The user may vary the average vehicular occupancy during each time slice.

For the FREQ#PE model, data are also required on the capacity and flow rates of parallel arterials. The model uses an "artificial" arterial which represents the sum of the parallel capacity and the average operating characteristics of alternative surface routes, as a means of measuring the impact of freeway operation on alternative routes.

The output reports for FREQ include estimates for the system as a whole and for individual freeway links. For the total system the calculates travel times, fuel consumption, emissions, and delay. For individual links and ramps, the model calculates average speed by time slice, delay, link volumes, and queue lengths (in vehicles). Additional data are output on spatial and modal traveler responses to congestion by time slice.

Application

The FREQ models [61] have been applied by a number of investigators analyzing freeway corridor traffic. Table 4-2 shows some selected examples of FREQ model applications.

Table 4-2 Selected Examples of FREQ Model Applications

Year	Principal Investigator	Agency/Firm	Site	FREQ Model	Description	Ref. No.
1970	Allen	University of California	Oakland Bay Bridge	3	Evaluate design and control alternatives	55
1971	Aidoo	University of California	I-80, Berkeley	3	Evaluate design and control alternatives	56
1972	Capelle	Voorhees	I-90, Cleveland	PRIFRE	Determine feasibility of priority lanes	57
1976	Gabard	French government	Nord RTE, Paris	3C	Evaluation of access control	58
1978	Ritch	Texas A&M	I-10, Houston	3CP	Projected future operations improvements	59
1978	English	Texas DOT	US59, Houston	3CP	Compare operations and design improvements	60
1979	Schneider	University of Washington	I-80, Berkeley	GRAF	Developed graphical output	61
1980	Michalopoulos	University of Minnesota	I-394, Minneapolis	6PE/6PL	Design and control strategy evaluation	62
1980	Immers	Delft University	A-12, Hague	4CP	Measured Impacts of design and operations	63
1980	White	New Zealand government	North Freeway, Auckland	3CP	Develop and evaluate ramp control plan	64
1980	Anderson	CALTRANS	I-10, Los Angeles	6PE	Estimated metering impacts on city streets	65
1981	Torres	JFT Associates	RTE 11, Los Angeles	6PE/6PL	Evaluated fuel conservation strategies	69
1981	Meyer	Colorado DOT	I-25, Denver	6PE	Analysis of optimized metering and geometrics	66
1981	Howard	Bartholomew	I-95, Miami	6PE/6PL	Evaluated feasibility of TSM techniques	67
1981	Deakin	Lockner	I-95, Miami	6PE	Determined feasibility of ramp metering	67
1981	Berg	Parsons	I-5, Seattle	6PE	Evaluation of TSM-type strategies	68
1981	O'Neill	Washington DOT	I-405, Seattle	6PE/6PL	Priority lane evaluation	68

INTRAS

Model Development and Description

On the basis of the simulation procedure used in the NETSIM UTCS-1 model, Lieberman and Associates developed INTRAS (INtegrated TRAffic Simulation) in 1977 for use in studying freeway incident detection and control strategies. It is based on knowledge of freeway operations and surveillance systems and incorporates detailed traffic simulation logic developed and validated for use in the model. The model is written in FORTRAN and can be run on both CDC and IBM mainframe computers. INTRAS is supported by FHWA [95, 96, 97].

INTRAS is a stochastic, microscopic model that uses car-following and lane-changing algorithms to simulate the movement of individual vehicles. For this reason, it can examine both traffic control and geometric alternatives, even in complex freeway design situations such as interchanges. INTRAS allows the user to simulate an incident at any location on a freeway link and for any length of time. The incident may block one or more lanes or be confined to the shoulder.

To allow simulation of freeway control policies, including ramp metering and diversion, the capacity to model the off-freeway environment is included in INTRAS. This surface traffic modeling is patterned after the logic of the NETSIM (UTCS-1) simulation model.

To facilitate the simulation of closed-loop incident detection and control, as well as off-line traffic analysis, the INTRAS model contains a realistic surveillance system simulation capacity. The ability to graphically detail vehicle trajectories and the contours of measures of effectiveness (MOEs) in time and space is included in INTRAS through a digital plotting module. INTRAS also contains a statistical analysis module that uses standard parametric and nonparametric tests to compare MOEs from different simulation runs and/or field data.

INTRAS can simulate three types of traffic detectors: Doppler radar detectors, short inductance loops, and coupled short inductance loops. Finally, a fuel-consumption and vehicle-emission evaluation model is built into INTRAS, patterned after a similar module developed for the NETSIM (UTCS—1) simulation model.

INTRAS Input/Output

The data required for input to INTRAS include geometric data to describe each link length, the number of lanes, lane channelization, type of link (mainline, ramp, arterial), grade, radius of curvature (for freeway links), percent of superelevation, and pavement type. Operational data required for input include entry-link flow rates; the percentages of intercity buses, heavy single-unit trucks, trailer trucks, high-performance passenger cars, and low-performance passenger cars; turning percentages at intersections; discharge headways; lost time; and free-flow speeds. Also required for input to INTRAS are control data to identify the type of control at intersections, i.e., stop sign, pretimed or actuated signal; the actual signal timing; ramp-control operation; location and type of detectors present; and location, type, and time of any incidents occurring during the simulation.

The standard output (consisting of such MOEs as vehicle-miles, vehicle-minutes, volume, density, speed, delay per vehicle, lane changes, etc.) will normally be reported at the end of each simulation subinterval, on both a link-specific and network-wide basis. The user may also generate these reports at specified time intervals within each subinterval. These statistics are cumulative either from the start of the simulation or, optionally, from the beginning of each subinterval. The output of the surveillance detector data may be restricted to individual links or inhibited altogether. Furthermore, data may be output to tape or disc file for later processing by a plotting module. Through this option, vehicle paths and/or MOE contour plots may be created for selected freeway links or groups of links.

Application of INTRAS

Both the traffic control simulation model and the incident detection algorithm of the INTRAS model were validated against a number of data sets covering a range of flow conditions on freeways with and without incidents [97]. A demonstration of INTRAS was conducted on a section of the northbound Shirley Highway (I-395) located in Virginia. For the sake of computing efficiency, the network was truncated into approximately 15 lane-miles of freeway links plus the associated ramps and surface streets. The flow conditions simulated were representative of A.M. peak period traffic. Entrance volumes were spread equally over all lanes. The simulation run consisted of a series of 30 one-minute subintervals. During this time, an incident was triggered. Two runs were performed: the first had no ramp control strategy in effect, and the second implemented a speed control metering strategy. The results showed that statistically significant improvement could be attributed to the presence of the ramp control strategy; however, during and after the incident, the ramp control strategy improved speeds consistently on this link.

The most recent INTRAS application found in the literature was a test of the model's performance in examining several demand-responsive strategies. The study was conducted by VERAC Inc. in 1984 [Payne]. In this study, INTRAS provided realistic representations of detector data when compared to actual data produced by current surveillance systems.

Recommended INTRAS Improvements

In a review of the model, the following recommendations were made by KLD Corporation to improve the usefulness of the model:

- Implementation of the "Dead-Start" capability should be undertaken. This would allow the simulation to be halted and then restarted without the loss of any data. This capability would be useful for debugging problems and in running parametric studies.

- Outer weaving flows should be simulated. The model could be modified to allow on-ramps to have separate turn percentages. This would allow a different proportion of vehicles entering an on-ramp to exit at an associated off-ramp than the proportion specified for freeway vehicles.
- Lane changing logic should be modified to allow vehicles to "look" ahead or behind for appropriate gaps. At present a vehicle can only "see" the lane immediately adjacent to it.

MACK FAMILY

Model Development and Description

The MACK model and its later refinements are deterministic, macroscopic models that consist of a set of conservation equations and a corresponding set of dynamic speed-density equations. Payne and Associates began work on the model in the late 1960s, and models in this family include MACK I, MACK II, MACK III, FREFLO, and TRAFLO. TRAFLO is the latest version in this model family [70]. The MACK models are programmed in FORTRAN and are well documented.

In the MACK I model, the dynamics of traffic are described by the numerical solution of fluid-flow differential equations, appropriately modified to represent the traffic environment. This model generates results that exhibit all the global dynamic responses to freeway traffic flow, with a minimum of computational costs. It is capable of simulating the response of the system to incidents that block one or more lanes on a section of roadway. The model does not distinguish flow by lanes.

In the MACK II model, a new equilibrium speed-density relationship and a structural change in the dynamic speed relationship were introduced. Other changes to the model were minor. The MACK II and INTRAS models were applied to a segment of the Shirley Highway (I-395) outside of Washington, D.C., with incident-free and incident scenarios. Because of the qualitation argument between the MACK II and INTRAS results and the large difference in cost between execution of the macroscopic MACK model and the microscopic INTRAS model, the study recommended that MACK II be adopted to make preliminary evaluations of control strategies for responding to incidents.

The FREFLO model was a successor to the MACK II model [70]. The FREFLO model was designed to provide a basic freeway simulation model, by

- providing input data diagnostics,
- representing incidents,
- modeling on ramps,
- representing different time-of-day control periods,
- representing surveillance systems,
- representing traffic-responsive metering schemes,
- providing standard measures of travel and travel time,
- providing fuel consumption estimates, and
- estimating air pollution emissions.

Model Input/Output

The geometric data required by FREFLO consist of the number of lanes in each section, section lengths, on-ramp and off-ramp locations, and nominal section capacities. The traffic data required are densities and speeds of each section for the initial state, upstream freeway volume for each time period, on- and off-ramp volumes for each time period and an O/D model. FREFLO offers the output options of diagnostics only or simulation and a choice of detailed outputs. Travel time, queue waiting time, diverted volume, emission and fuel consumption all are included.

Application of FREFLO

Because of the dynamic nature of the model and the anticipated short control intervals, the FREFLO model was chosen by the Institute of Transportation Studies at the University of California, Berkeley, to evaluate control strategy for a research project examining on-line freeway entry control in 1980. In a recent study of demand responsive strategies for interconnected freeway ramp control systems, FREFLO was used to test the evaluated control system's performance [67, 68, 69].

Experimentation revealed a serious deficiency in the model's ability to simulate congested flows in a realistic fashion. The problem was traced to the model's transition from the continuous to the discrete domain. The properly discrete model required an excessive amount of computer time for a real freeway simulation. Hence, two adaptive schemes that reduced the computing time to a manageable level were presented. The resulting model, FRECON, was then calibrated and validated with five peak-period data sets from the Santa Monica Freeway in Los Angeles [7].

The first adaptive scheme is a heuristic scheme in which flow levels and geometry are examined and compared with a library of required subdivision patterns. In this way the spatial steps are adjusted at regular intervals so that the maximum number of allowed steps is always used. When these spatial steps have been established, the corresponding maximum number of temporal steps that ensure stability is known. With the use of these asynchronous time steps, it is possible to reduce the number of times each subdivision state needs to be integrated.

The second adaptive scheme transforms the model from a stationary reference frame to a moving one. Regions that move with the flow of traffic can be defined. Each region is aggregated into a "box" that contains a constant number of vehicles and a length that is inversely proportional to its density. This natural variation of length with density provides the spatial adaptation. A scheme similar to the heuristic adaptation is used to reduce the frequency of the integration steps required. Both of these adaptation schemes have proven useful in reducing the computing time required for a simulation.

Another development of FREFLO is that it is an integral part of the TRAFLO model system, which is a subsystem of a larger integrated TRAF. TRAF will be described in more detail in Chapter 6.

and the large difference in cost between execution of the macroscopic MACK model and the microscopic INTRAS model, the study recommended that MACK II be adopted to make preliminary evaluations of control strategies for responding to incidents.

The FREFLO model was a successor to the MACK II model [70]. The FREFLO model was designed to provide a basic freeway simulation model, by

- providing input data diagnostics,
- representing incidents,
- modeling on ramps,
- representing different time-of-day control periods,
- representing surveillance systems,
- representing traffic-responsive metering schemes,
- providing standard measures of travel and travel time,
- providing fuel consumption estimates, and
- estimating air pollution emissions.

Model Input/Output

The geometric data required by FREFLO consist of the number of lanes in each section, section lengths, on-ramp and off-ramp locations, and nominal section capacities. The traffic data required are densities and speeds of each section for the initial state, upstream freeway volume for each time period, on- and off-ramp volumes for each time period and an O/D model. FREFLO offers the output options of diagnostics only or simulation and a choice of detailed outputs. Travel time, queue waiting time, diverted volume, emission and fuel consumption all are included.

Application of FREFLO

Because of the dynamic nature of the model and the anticipated short control intervals, the FREFLO model was chosen by the Institute of Transportation Studies at the University of California, Berkeley, to evaluate control strategy for a research project examining on-line freeway entry control in 1980. In a recent study of demand responsive strategies for interconnected freeway ramp control systems, FREFLO was used to test the evaluated control system's performance [67, 68, 69].

Experimentation revealed a serious deficiency in the model's ability to simulate congested flows in a realistic fashion. The problem was traced to the model's transition from the continuous to the discrete domain. The properly discrete model required an excessive amount of computer time for a real freeway simulation. Hence, two adaptive schemes that reduced the computing time to a manageable level were presented. The resulting model, FRECON, was then calibrated and validated with five peak-period data sets from the Santa Monica Freeway in Los Angeles [7].

The first adaptive scheme is a heuristic scheme in which flow levels and geometry are examined and compared with a library of required subdivision patterns. In this way the spatial steps are adjusted at regular intervals so that the maximum number of allowed steps is always used. When these spatial steps have been established, the corresponding maximum number of temporal steps that ensure stability is known. With the use of these asynchronous time steps, it is possible to reduce the number of times each subdivision state needs to be integrated.

The second adaptive scheme transforms the model from a stationary reference frame to a moving one. Regions that move with the flow of traffic can be defined. Each region is aggregated into a "box" that contains a constant number of vehicles and a length that is inversely proportional to its density. This natural variation of length with density provides the spatial adaptation. A scheme similar to the heuristic adaptation is used to reduce the frequency of the integration steps required. Both of these adaptation schemes have proven useful in reducing the computing time required for a simulation.

Another development of FREFLO is that it is an integral part of the TRAFLO model system, which is a subsystem of a larger integrated TRAF. TRAF will be described in more detail in Chapter 6.

SCOT MODEL

Model Development and Description

The DAFT, SCOT, and SCOT-Q models are classified as members of the SCOT family of models [May]. Work began in this modeling family in the late 1960s and continues today. The initial work was undertaken by Lieberman and his associates. The SCOT (Simulation of COrridor Traffic) model is a combination of two simulation models: DAFT (Dynamic Analysis of Freeway Traffic) and NETSIM(UTCS-1). It has the capacity of simulating traffic patterns within an integrated freeway corridor. It is macroscopic on the freeway elements, microscopic on the ramps, service roads, major arterials, and city streets. Traffic performance measures may be computed for a wide variety of control strategies. SCOT-Q is a faster version of SCOT.

The DAFT model is a macroscopic simulation of traffic along a network of freeways, ramps, and arterials. Vehicles are grouped into platoons and lose their individual characteristics. The platoons are moved along the freeway according to a single, prespecified, speed-density relation that applies to all freeway links. Along the non-freeway links, they travel at the specified free-flow speed for each link and are delayed at the downstream end for a time related to the ratio of green time to signal cycle time and the volume of traffic on that link. Minimum-cost paths are calculated frequently by the model, based on current conditions. Whenever a platoon reaches a network node, its turning movement there is dictated by its minimum cost path as it exists at that instant of time. Hence, the model produces a dynamic assignment of traffic as a by-product of the simulation.

The microscopic logic of NETSIM is applied to those components of the network characterized by signalized, at-grade intersections. Here, the traffic mechanisms are complex, because many conflicts are common to the patterns of urban traffic. Each vehicle is treated individually, and a small time step must be applied to obtain an acceptable level of accuracy in the replication of global traffic flow. Traffic along the freeway, however, has been modeled successfully with the use of fluid-flow equations. The model designers believed that applying a microscopic approach to freeway traffic would yield little in the way of additional global accuracy. Furthermore, a microscopic treatment of freeway traffic would greatly magnify computing costs and storage requirements.

SCOT Input/Output

The information required by SCOT to model a network is generally available to traffic engineers. These data fall into four sets: geometric, traffic demand, control system, and bus schedules.

Typical geometric data needed include

- the configuration of each street or freeway section,
- grades,
- details of ramps and turning lanes,
- the number and widths of travel and parking lanes, and
- the locations of internal traffic generators, such as shopping centers and parking lots.

Data that describe the volume and character of traffic must be provided by the user. Speed-density data are needed for each freeway section. Also required are flow rates at the periphery of the network, traffic mix (i.e., ratio of number of passenger cars to trucks), turning movements and pedestrian activity at each intersection, street free-flow speeds and their rate of queue discharge.

All parameters of the control strategy are needed. These parameters include the following:

- traffic signal timings and synchronizations,
- parking restrictions and lane use policies, and
- traffic-actuated logic.

Bus service data needed include bus routes, station locations and capacity, frequency of bus service, and mean station dwell times.

The SCOT output is a dynamic history of the state of traffic on each street and intersection during the simulation period. In addition, averages for the entire network are computed. Thus, a local and a systemwide view is obtained.

The locations and durations of spillback (the extension of a queue into the upstream intersection blocking cross traffic) and the instances of cycle failure (the inability of a green interval duration to discharge the entire queue) are recorded. The number of vehicles currently on each street is given. Cumulative statistical data for each street, as well as the network as a whole, are also available. These data include the number of stops made on each street, average speed, average delay per vehicle and mean occupancy on each street and for the network as a whole, vehicle miles, vehicle-minutes, delay per vehicle-mile, and travel time per vehicle mile. Bus statistics are compiled separately and include the number of stops made on each street. Bus route data, such as average speed, total bus station dwell time, and total delay time for each bus route, are given.

SCOT Application

SCOT was applied to the central business district of Minneapolis and to a 1.2-mile test network of the Dallas North Central Expressway [May]. In the Minneapolis application, the SCOT model was used to predict the effect on bus service and general traffic performance of implementing candidate bus priority strategies. In both applications, tests showed no significant differences between field and simulation results for the basic parameters of traffic speed, flow, and saturation. However, a demonstration of the O/D traffic assignment capacity of the model indicated that the minimum time-path criteria used were not conclusively correct for traffic assignment.

Sperry Corporation researchers have also used SCOT for their applications and model development [58].

CHAPTER 5
DEMAND-RESPONSIVE URBAN TRAFFIC CONTROL SYSTEMS

As an expression of human behavior, urban traffic is variable in time and space. Thus, a high degree of adaptiveness is required in the control of such traffic to provide a suitable response to this variability. Since the development of modern traffic signal controls, traffic engineers and signal system designers have attempted to make them as responsive as possible to prevailing traffic conditions. Designers have believed that increased responsiveness leads to improved traffic performance. However, the extent to which responsiveness to traffic is achieved depends on a variety of factors, including control hardware, software capabilities, surveillance equipment, and operator qualifications. A significant amount of research in the development of efficient and effective real-time, demand-responsive traffic control systems is still being done [34, 35, 36].

To date, many generations of computerized urban traffic control systems, also termed area traffic control systems, have been developed and implemented. These systems gather data on traffic volumes and network link occupancies from vehicle detectors (sensors). On the basis of this data, they then alter the signal settings at intersections in the network in order to control traffic flow.

The configuration of the controlling computer system can be implemented in either a central-processing format or a distributed-processing format. Development of the three most well known demand-responsive traffic control systems, namely, UTCS in the United States, SCOOT in England and SCAT in Australia are described below. These three systems are fully supported by their respective governments, and their experiences have been widely discussed and documented in the literature.

URBAN TRAFFIC CONTROL SYSTEM (UTCS)

The Urban Traffic Control System (UTCS) experimented project was initiated by the Federal Highway Administration in the early 1970s to provide computer-supervised control of 200 signalized intersections in Washington, D.C. The software package was converted from assembly language into FORTRAN in 1973. This project has had an enormous impact on the traffic engineering community in the United States [Stockfisch]. The UTCS research project on traffic control strategies has been divided into three generations of traffic control techniques, usually referred to as 1-GC, 2-GC, and 3-GC. The "Traffic Control Systems Handbook" has a detailed description of this system [92]. (Recent work in the U.S. with UTCS systems has produced a hybrid system sometimes refferred to as generation 1.5.)

The First-Generation Control (UTCS 1-GC)

First-generation control uses prestored signal timing plans developed off-line and based on historical traffic data; the system is capable of storing up to 40 timing plans. The plan controlling the traffic system can be selected on the basis of time-of-day, by direct operator selection, or by the match of a plan from the existing library to recently measured traffic conditions (traffic-responsive mode). The matching criterion is based on a network threshold value that incorporates traffic volumes and occupancies. The mode of plan selection is determined by the operator [30].

In the traffic-responsive mode, timing plans are usually updated once every 15 minutes. Smooth transition between different timing plans is provided by a transition routine which is part of 1-GC. This routine evaluates the magnitude of the changes, determines the amount of time required for a smooth transition, and then controls signal timing settings until the transition is complete. The same procedure of control is used for transitions between computer control and operator or time of day control.

The pattern in effect is enhanced by a critical intersection control (CIC) feature, which is used to fine-tune the system at intersections that saturate frequently, and which adjusts the allocation of green time (split) based on fluctuations in local traffic demand.

Certain intersections may also be instrumented for bus priority. The decision to grant additional green time to buses is a function of passenger volumes, vehicular queues in and around the intersection, and the time of the arrival of a bus on the approach to the intersection. The program logic that makes this decision is contained within the bus priority system algorithms in realtime and adjusts the signal splits in response to actuated demands.[8, 30] The TRANSYT model was used to generate the timing plans for testing the UTCS 1-GC research project in Washington, D.C.

The Second-Generation Control (UTCS 2-GC)

As an alternative to the stored library of signal timing plans used by 1-GC software, 2-GC was developed as an on-line model that computes and implements signal timing plans in realtime based on surveillance data gathered from vehicle detectors and predicted traffic volumes. The 2-GC technique can be termed a cycle-based control strategy, since a common cycle length is used with variable groups of intersections. The software of 2-GC contains an on-line optimization routine (which employs the optimization algorithm of the SIGOP model discussed earlier) to determine the control parameters that will minimize the total delay and stops within the network under current traffic conditions. The optimization process can be repeated at 5- to 10-minute intervals or whenever traffic conditions have changed enough to justify the computation of a new set of control parameters.

Additional sophistication is also provided for the second-generation critical intersection control. This routine will vary not only the signal split, as in the case of the first-generation software, but the relative green times between intersections (offsets) are also adjusted on a cycle-by-cycle basis to provide additional responsiveness to short-term fluctuations in traffic volumes.

Second-generation software also has the capability for dynamically decomposing the network into sections (subnetworks) on the basis of traffic conditions. The optimum timing patterns are then computed for individual sections, and the sections are interfaced to assure smooth traffic progression across the subnetwork boundaries.

A number of other routines are contained in the second-generation software that are designed to improve flow within the network, including an improved transition routine, a prediction capability, and software to estimate traffic conditions at intersections that are not instrumented with vehicle detectors. The second-generation control package is known as the Traffic Adaptive Network Signal Timing Program [44].

Prediction of traffic volumes, occupancies, and speeds on each link of the street network is critical to the effectiveness of the timing plans determined by the 2-GC technique. The prediction models must be capable of providing accurate forecasts of traffic variables that will exist when the new timing plan is in use, and these forecasts are based on data that have been collected during previous time periods. Errors in the prediction process can seriously degrade the performance of the implemented timing plan, particularly in the case of 2-GC, when the advanced time for prediction is greater than 5 minutes.

The Third-Generation Control (UTCS 3-GC)

This control strategy is significantly different from those of 1-GC and 2-GC, in that it was conceived to implement a fully traffic-responsive, on-line control system. To accomplish this goal, 3-GC was designed to allow the signal timing parameters to change continuously in response to real-time measurements of traffic variables. The time period between timing plan revisions were intended to be in the magnitude of 3 to 5 minutes. Numerous problems associated with instrumentation (and associated maintenance) and transition-related deficiencies forced the discontinuation of this work, and no further work is currently planned in this area [92].

The UTCS First Generation Enhanced and Extended

In order to minimize the costs and increase the reliability and operational performance of computer-based traffic control systems for the states and municipalities by standardizing the software and improving and supporting the FHWA-developed control programs, the Federal Highway Administration decided to incorporate the proven functions of the original 1-GC software into an enhanced software package that would be compatible with minicomputer traffic control systems.

The major functional components comprising the enhanced 1-GC software are detector telemetry processing, measures of effectiveness algorithms, operator interface language, communications processing, controller operations, manual plan selection, time-of-day/day-of-week operation, traffic responsive operation, and critical intersection control.

The features of the enhanced and extended UTCS first generation are as follows:

Extended New capabilities. The new capabilities incorporated into the extended 1-GC involve the following.

- Database Management. The new software provides complete database management capability. Operator interface is implemented with an operator interface language, which uses traffic engineering terminology. The software will enable the operator to change any parameters that are dynamic in nature.
- Controller Assemblies. Commands to actuated controller assemblies are expanded and an interface with microprocessor-based controller units is provided. Controller modes include flashing operation and special function commands. Intersections that are preempted locally are accounted for, and intersections with similar traffic patterns can be grouped automatically into sections or subnetworks.

Enhanced capabilities. The basic enhancements to the 1-GC capabilities include the following.

- Controller Monitoring and Control. The software's capability of monitoring and controlling actuated controller operations are increased.
- Signal Timing Plans. The software associates timing plans with individual controllers, thus allowing any combination of intersections and controllers to operate as a subsystem or section.

The enhanced software is capable of assigning an operator command to a specific time of day/day of week file, thus allowing commands to be issued automatically. Also, it is capable of providing rapid transitions from one timing plan to another within minimum or maximum constraints imposed on the transition cycle length. A transition begins at a designated main street green with the implementation of the new timing plan's cycle length and split. The offset transition is accomplished by either the contraction or expansion of phase times until the current offset equals the new intended offset.

UTCS 1.5-Generation Control

The cities that use 1-GC are faced with the problem of developing a full library of suitable timing plans to accommodate the various traffic patterns that may arise in each control section within an areawide control system. Several hundred timing plans are often required when the number of control sections and the number of different traffic patterns are properly enumerated. Many cities use versions of the TRANSYT and SIGOP models as timing plan generation software, but the task of assembling and inputting data is complex, burdensome, and requires considerable staff time. Therefore, it is not uncommon to find that many computer-supervised signal control systems are functioning with a small number of relatively crude timing plans. Most often these plans were developed shortly after system implementation and have not been revised for a long time. [Handbook]

To overcome the problems resulting from an inadequate number of improperly maintained timing plans, a simple procedure must be developed for generating and maintaining an adequate library of appropriate timing plans. One way to accomplish this objective is to use an on-line timing plan generator that automates the loading of the generated plans into the traffic control system. A computer program, FORECAST, was developed to handle this situation, and the result of this development is the UTCS 1.5-Generation Control.

FORECAST

FORECAST is a computer software model that was developed by Computran Systems Corporation for the purpose of generating timing plans for signalized intersection networks. It is designed for on-line usage with a combination of UTCS 1-GC and a proprietary software traffic control package to facilitate the generation and maintenance of the timing plans library.[101]

The structure of FORECAST permits an interactive search for optimum timing plans over a range of cycle lengths. Phase sequences, green times, change times, and clearance times are stipulated in the FORECAST input in accordance with their settings on the controllers. The optimization logic FORECAST uses involves sequential threading of prescribed movements through the network with a priority list of demands to be accommodated. During this threading process, FORECAST adjusts in a priority manner the individual splits and offsets of each intersection in order to accommodate the defined movements throughout the network. The optimization criterion is a function of stops and delays. Once FORECAST identifies the optimum solution, the corresponding timing plan for this solution is output for the user in the form of cycle length, offset, and interval time for the intersections. Time-space diagrams for the selected timing plan can also be produced if requested by the user.

The UTCS 1.5-generation control was initially implemented in Winston-Salem, North Carolina, in 1981. In this implementation, optimal timing plans can be generated for sections of 50 intersections while the system is on line and operating a total of 209 intersections. The system operator evaluates the suitability of the generated plan based on the output measures of effectiveness and time-space diagrams. If the resultant output is acceptable, the timing plan can be loaded into the system library.

The overall impact of using FORECAST along with 1-GC in the Winston-Salem signal control system is that the city staff are more able to contend with the normal operating problems inherent in the long-term operation of the control system [101].

The UTCS Experience

At the present, many computer-supervised control systems operate using the first-generation UTCS control concepts. The UTCS 2-GC generation is operational in Overland Park, Kansas [46]. No other operational 2-GC installation in the United States could be found in the literature. The third UTCS generation has not been fully implemented or evaluated because its extensive surveillance and computation requirements exceeded the capabilities of the available hardware and software when it was developed. For this reason, third generation control requires additional research and development before it can receive serious consideration as a candidate for operational implementation.

Experimentation with 1-GC has indicated that, in its various modes of operation, it performed well and demonstrated that it can provide some measurable reductions in total travel time over that which could be attained with a well-designed pretimed system. The pattern-matching mode of 1-GC plan selection is generally more effective than the time-of-day mode. The UTCS second generation showed mixed results during its limited evaluation. Overall it was judged inferior to 1-GC It demonstrated some small improvements on the arterial street level but degraded traffic flow in the network [92].

A comparison of the basic features of the first two UTCS generations is shown in Table 5-1. Each control generation is designed to provide an increased degree of traffic responsiveness. Note that very few 2-GC systems have been installed in the U.S. Apparently the expected benefits of the system do not warrant its costs.

Because of declining funds available for urban traffic control projects, an apparent lack of a vocal constituency for new development in this area, and problems with

Table 5-1 Comparison of UTCS Control Techniques

Feature	First Generation	Second Generation
Optimization	Offline	Online
Frequency of update	15 min	5-10 min
Number of timing patterns	Up to 40	Unlimited
Traffic prediction	No	Yes
Critical intersection control	Adjusts split	Adjusts split and offset
Hierarchies of control	Pattern selection	Pattern computation
Fixed cycle length	Within each section	Within variable groups of intersections

competing with the private sector, FHWA's future role will likely be to concentrate on testing and evaluation of existing software and to provide software support and maintenance to existing programs rather than to develop new on-line programs [59, 89].

SPLIT, CYCLE AND OFFSET OPTIMIZATION TECHNIQUE — SCOOT

System Development

The SCOOT (Split, Cycle and Offset Optimization Technique) system is a UTCS 2-GC equivalent, demand-responsive, urban traffic control system developed by the Transport and Road Research Laboratory (TRRL) in England beginning in 1973. Research was carried out in Glasgow by a team from TRRL along with the Ferranti, GEC and Plessey companies, and assistance from the Strathclyde Regional Council. A development project between the Department of Transport and Industry and the three traffic companies was initiated in 1976 and included a full-scale trial of SCOOT in Glasgow in 1979. With the cooperation of the West Midlands County Council, SCOOT was then installed in Coventry. The latest version of SCOOT is 2.3. Operations of SCOOT systems have been reported successful. So far 15 SCOOT systems have been installed, including one in Hong Kong. The key persons in the SCOOT system's development are Hunt and Robertson, and material presented in this section are mainly from their papers [10].

The objectives of the Bristish SCOOT research project are described as follows [Robertson].

- To reduce vehicle delay, stops and congestion to below the levels achieved by the best fixed time system. Previous attempts to realize the potentially better performance of adaptive UTC systems have failed to prove that this minimum objective can be met.
- To remove the need for updating fixed time plans. Maintaining a library of fixed time plans is expensive. Estimates are that one person-year is needed to produce new plans for a 24 signal network. Though it may be cost-

effective to update plans annually, in practice plans are usually several years old.

- To provide information for traffic management purposes. Real-time data on vehicle flows and congestion levels are valuable for incident management and for longer term planning and monitoring of road improvement schemes.

The SCOOT Concepts

The lack of success of previous traffic responsive systems in both US and UK led to an investigation of the reasons for their failure. Some of the problems of the earlier adaptive systems are thought to be

- frequent plan changing,
- inadequate prediction of traffic flows,
- slow response, and
- effects of poor decisions.

The SCOOT system design attempts to avoid these problems as follows.

- <u>Minimize Transients</u> — SCOOT uses frequent, small incremental alternations to split, cycle time and offset. These alternations minimize transients but can create new patterns of coordination.
- <u>Short Term Prediction</u> — Most SCOOT decisions are based on the current situation and longer term predictions are seldom necessary.
- <u>Fast Response</u> — In SCOOT, every red/green transition is optimized. The split, offset and cycle optimizers are designed into a hierarchy to ensure compatibility and minimize the potential for degrading traffic performance.
- <u>Faulty Detectors</u> — Detectors are monitored continuously for faults. Suspect detectors are automatically ignored and cause a local reversion to fixed timings that have a minimal effect on the normal operation of SCOOT.
- <u>On-line Traffic Model</u> — SCOOT uses data from the detectors to predict queues at signal stoplines. This traffic model provides information

from which the signal optimizers make their decisions. In addition, the traffic engineer can use the model to obtain insight into the operation of the traffic network.

- No Background Plans — SCOOT needs no initial fixed time plans and can, if necessary, start from any traffic signal settings.

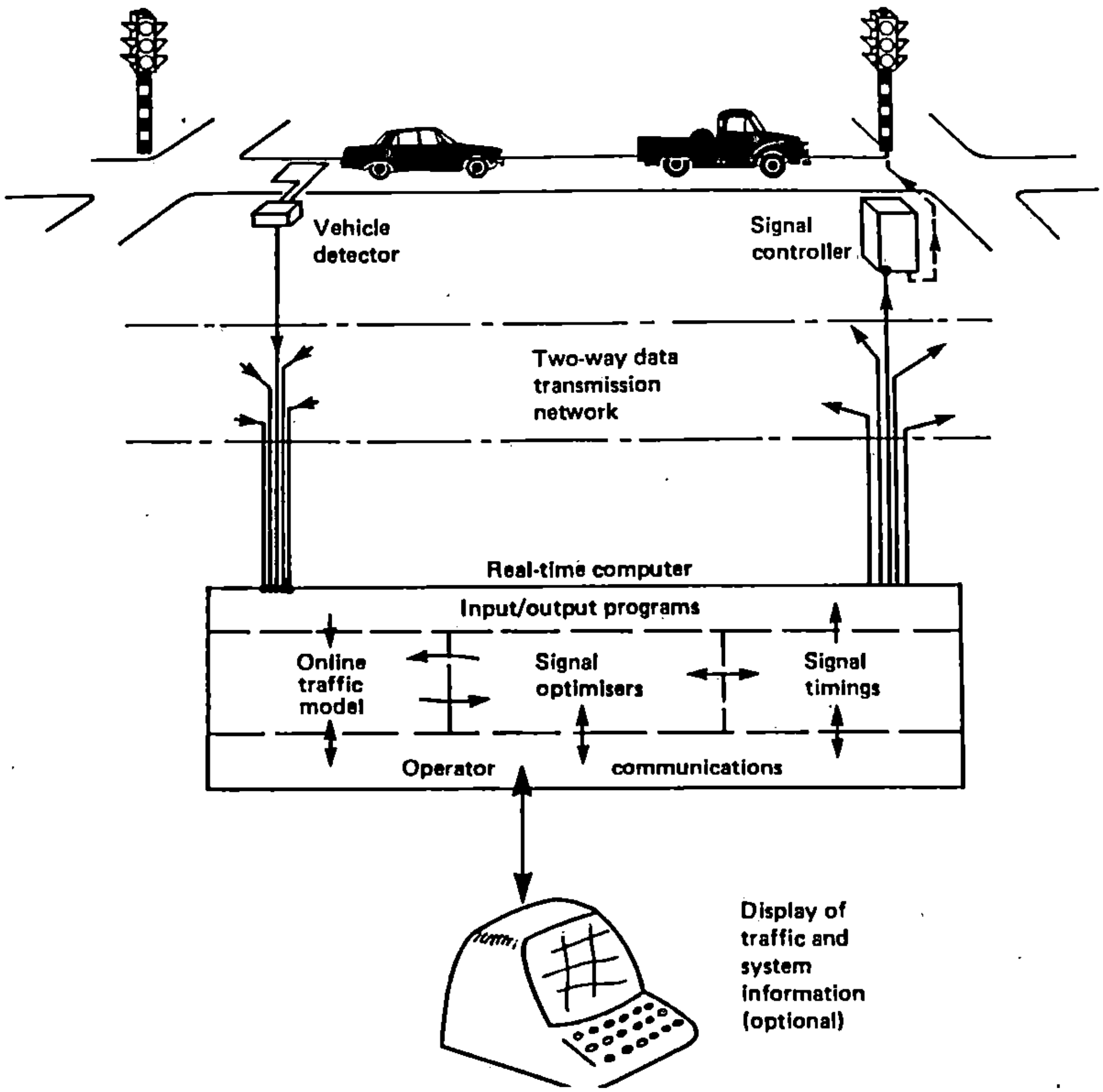
The SCOOT Model

Figure 5-1 shows the structure of the SCOOT system. Inductive loop vehicle detectors are located on the approaches to all signalized junctions that are under SCOOT control. The detectors are positioned as far upstream as possible from the signal stopline. Data are collected, processed and stored in link cyclic flow profiles. The cyclic flow profiles contain information needed to decide how best to coordinate adjacent pairs of signals, as well as information on the demand for green time. The values in the profiles are affected by vehicle presence as well as vehicle flow, but the term "flow" is used for simplicity. Figure 5-2 shows an example of cyclic flow profiles. [41, 42]

For each link, the SCOOT traffic model predicts the current value of the queue at the stop line. The detected vehicle is assumed to travel at a fixed cruise speed to the stop line. The state of the signals is known, and with the use of a preset saturation flow value, the length of the queue and the back of the queue is estimated. The position of the back of the queue is used to provide congestion information for the signal optimizers. Figure 5-3 shows the principle of the SCOOT traffic model.

The SCOOT Optimizers

Incremental optimization is the feature of SCOOT's optimization function. The key idea is that the coordination plan should be able to respond to new traffic situations in a series of frequent but small increments. This is necessary because research has shown that traffic flows in the next few minutes are very difficult to predict; hence, any fixed coordination plan may be out of date before it is calculated or inappropriate after it is



Data source: SCOOT - A Traffic Responsive Method of Coordinated Signals, by P.B. Hunt, D.Z. Robertson, R.D. Brethertion and R.Z. Winston, TRRL Laboratory Report 1014, 1981.

Figure 5-1. The Flow of Information in a SCOOT Urban Traffic Control System

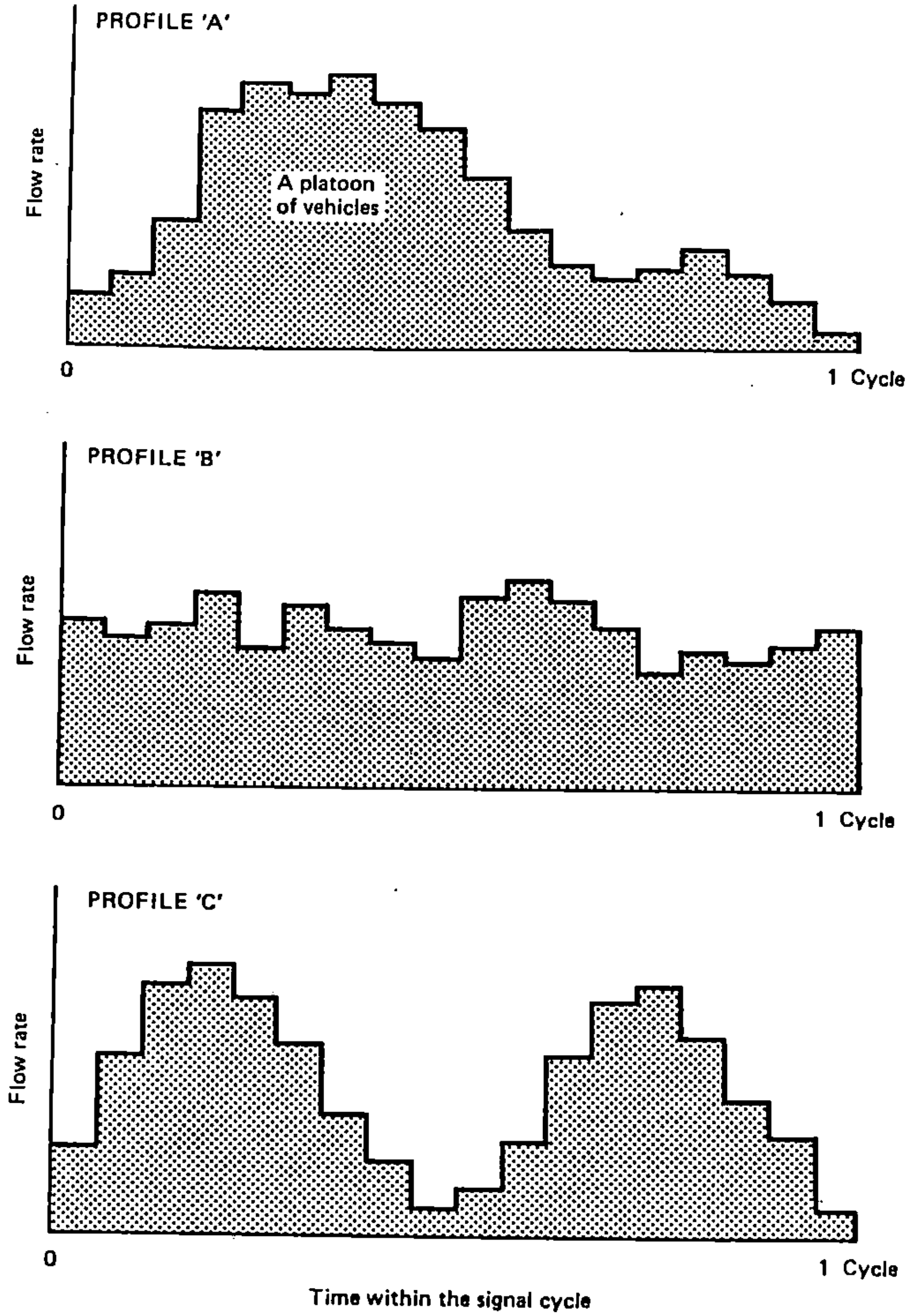


Figure 5-2. Three Examples of Cyclic Flow Profiles

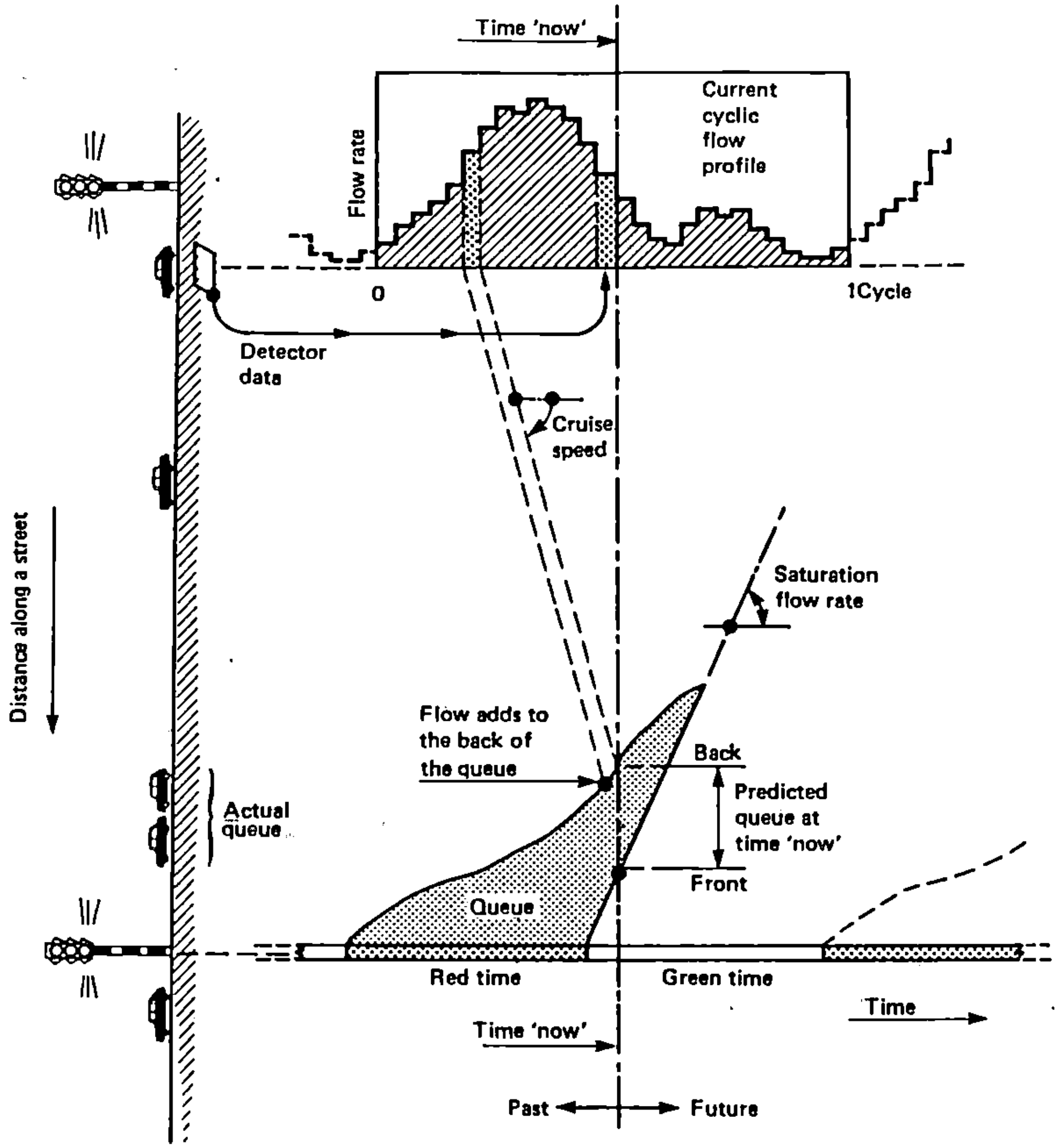


Figure 5-3. Principles of the SCOOT Traffic Model

implemented (and implementation is likely to cause extra delay during the transition from the old timings to the new).

The SCOOT on-line system has a set of signal timings which, if unaltered by the optimizers, would effectively be a fixed time plan. By frequent small alternations, SCOOT controls the signals on a plan that evolves through time. SCOOT makes a very large number of small optimization decisions - more than 10,000 per hour in a network of 100 intersections. Of course, some decisions are wrong, but this is not important provided the large majority are correct.

The offset optimizer operates on each junction during each cycle. The information in the cyclic flow profiles is used to estimate whether an alternative to the offset will improve the overall traffic progression on the streets that are immediately upstream or downstream of the junction. This decision is based on minimizing a calculated performance index which uses inputs of delay, stops and congestion.

SCOOT operates sub-areas of signals on a common cycle time in order to maintain coordination between signals. The cycle time optimizer can vary the cycle of each subarea in increments of a few seconds at intervals of not less than 2.5 minutes. Each sub-area is varied independently of other sub-areas between preset upper and lower bounds. The cycle time is varied by SCOOT to ensure that the most heavily loaded intersection operates, if possible, at a maximum degree of saturation of 90 percent. The cycle time of a sub-area may also be changed if SCOOT calculates that by switching from single to double cycle operation, a net saving in delay is possible.

The amount of congestion, measured in the on-line model for each link, is used to modify the decisions of the split and offset optimizers. Green time can be increased to reduce congestion and the offset on a link can be improved to reduce the risk of blocking an upstream intersection. Thus SCOOT uses an "elastic" coordination plan that can be stretched or shrunk to match the latest situation recorded by the vehicle detectors. SCOOT changes the coordination plan by optimizing splits, offsets, and cycle time in the following way.

- A few seconds before every phase change, the SCOOT split optimizer calculates whether it is better to advance or retard the scheduled change by up to 4 seconds, or to leave it unchanged.
- Then, once a cycle, the offset optimizer assesses whether the Performance Index on streets around each intersection can be reduced by altering the offset of that intersection by up to 4 seconds either way.
- Favorable split and offset alterations of a few seconds are implemented every few minutes.

<u>New System–SCOOT Version 2.3</u>

While minimizing overall delay and stops is the normal criterion for choosing signal settings, the traffic engineer may wish to adopt an alternative strategy in certain situations. A new version of SCOOT, designated SCOOT 2.3, was developed by TRRL, and it has several features to give more flexibility to the system. This new version of SCOOT contains additions to the offset and split optimizers that allow preference to be given to particular links relative to others in the network. These link weightings can be used to give better progression along chosen routes or to reduce delay on particular links.

This version of SCOOT contains several measures that allow the traffic engineer to constrain the SCOOT optimizers to implement his/her chosen traffic control policy decisions. The main new features are

- offset weighting,
- split weighting, and
- fixed and biased offsets.

The methods used to weight links in the split and offset optimizers are different because the methods used by each optimizer to reduce delay and stops are different. Each of these features is described below.

Offset Weighting. The offset optimizer uses a performance index that is a combination of delay and stops. The offset weighting is therefore a multiplier on the PI of the chosen link. Weighting one link more heavily than others will cause the offset optimizer to reduce the delay and stops on that link in order to reduce the total PI.

Split Weighting. Normally the split optimizer chooses green times that equalize the degree of saturation (%) on conflicting links. It would be incorrect to simply factor up or down the percentage of saturations in order to weight links, as this might completely overload disfavored links. The method used allows any specified link to run at a higher degree of saturation with any spare green time being allocated to other phases.

Fixed and Biased Offsets. In some locations where intersections are closely spaced, there may be advantages in fixing the offset between adjacent signals. The offset optimizer can be commanded to fix any particular link offset. This will fix the offset between the start of the green phase on the main upstream and downstream stages of the link. Another case in which offsets can usefully be chosen to favor certain movements concerns the preferential coordination of a turning movement rather than the main through movement.

The new commands in SCOOT 2.3 are all link-specific and must be entered individually by the operator or must appear in the time-of-day schedule. The main use or these facilities is for routine policy measures such as favoring routes or main roads. The weighting values are a normal part of the data structure. Such facilities may also have a role in managing short-term traffic conditions and in mitigating the effects of traffic incidents.

Experiences with SCOOT

During the research stage of SCOOT development, TRRL conducted "floating car" traffic surveys in Glasgow and Coventry. These surveys made use of four or five specially instrumented cars to measure journey times throughout the working day. Measurements were made for two weeks before and after installation, with TRANSYT-based, up-to-date, fixed time plans as the before condition. The analysis took account of differing traffic flow levels [94].

These studies concluded that SCOOT reduced the average queue lengths by about 12 percent and was most effective where traffic was congested and flows were variable. Further benefit may occur over time because fixed time plans age by a few percent each year, thus degrading their performance. Thus it seems likely that in many applications SCOOT should reduce congestion by at least 20 percent compared with conventional, fixed time control. The general conclusions on the effectiveness of SCOOT have been born out by subsequent trials in Southampton and London [26, 71].

SCOOT is the most widely applied, traffic-responsive traffic control system in the world today. Fifteen SCOOT urban traffic control systems have been set up so far; the first application of SCOOT outside the United Kingdom is in Hong Kong [14, 77].

The cost of SCOOT largely relates to installation, connection, and maintenance of the vehicle sensors and computer control hardware. Both theoretical estimation and the end users' experience indicate that the extra cost of a traffic-responsive SCOOT system over a fixed time system is likely to be recovered in traffic benefits within the first year of operation.

Comparing SCOOT and TRANSYT

The research on SCOOT was launched as a result of the need for a system that automatically generated signal settings to match the current situation on the road. SCOOT adopted and adapted some of the TRANSYT concepts, as well as lessons learned from traffic-responsive systems that failed to work well in earlier trials in Glasgow and other cities. Essentially, SCOOT is an evolution of TRANSYT toward real-time control. The models' similarities are described below [74].

Optimization of signal coordination is meaningless unless the objectives are clearly identified and seen to be sensible. In both TRANSYT and SCOOT, the prime objective is to minimize the sum of the average queues in the network. This objective is expressed as a performance index (PI), which can be easily translated from its physical significance of vehicles delayed to financial terms.

A user of TRANSYT or SCOOT can choose a weighting factor that balances the relative importance of queues and stops. Most of the time, signal settings that minimize queues also do a reasonable job of minimizing stops, although a tendency exists to favor longer cycle times if a heavy weighting is given to stops.

Neither model uses "bandwidth" calculations. Many traffic engineers coordinate signals by maximizing bandwidth on a time-distance (TD) diagram. The great merit of the bandwidth method is that traffic flows do not have to be known in detail, and TD diagrams are visually informative. But bandwidth measures cannot be translated into financial terms. Furthermore, when serious congestion occurs, the bandwidth concept starts to break down because the growth of queues disrupts the bands in significant ways. In addition, in central urban areas, where many complex traffic movements intersect, bandwidth has little meaning.

Both TRANSYT and SCOOT contain traffic models that are able to estimate queue sizes. These models are used by "optimizers" to evaluate alternative signal timings and so help find the best settings. The TRANSYT and SCOOT models are both based on the concept of cyclic flow profiles (CFP). A CFP is a measure of the average one-way flow of vehicles past any chosen point on the road during each part of the cycle time of the upstream signal. Once a CFP is known, a computer can easily estimate how many vehicles will reach the downstream signals when they are red, and hence, calculate both the queue's size, how long it takes to clear and the effects of alterations in the offsets, and splits can be predicted. TRANSYT and SCOOT carry out these calculations in a similar manner.

The shape of the CFP has to be calculated for each one-way flow along all streets in the area. The calculation is made in an upstream-to-downstream direction. Splits and offsets can be optimized simultaneously by a "hillclimbing" procedure that mimics the way traffic engineers search for good timings.

At present, signal phase order has to be manually preselected for both SCOOT and TRANSYT, but both TRRL and FHWA are investigating ways of incorporating automatic selection of phase order into the optimizer.

SYDNEY COORDINATED ADAPTIVE TRAFFIC SYSTEM (SCAT)

System Development

The Sydney Coordinated Adaptive Traffic system (SCAT) has been under development by the Department of Main Roads, New South Wales in Australia since the early 1970s. SCAT is intended to provide a traffic adaptive UTC system that coordinates the greater proportion of Sydney's traffic signals. SCAT is a UTCS 2-GC equivalent urban traffic system that features a distributed control format. In 1982 the system comprised a central monitoring PDP11/40 computer, twelve remote regional PDP11/34 or PDP11/40 computers and 600 microprocessor local controllers. All software has been programmed in assembly language [54, 86, 87].

SCAT Computer Control System

The SCAT computer system is organized in a hierarchical manner. Figure 5-4 and Figure 5-5 show the system structure. Each regional computer maintains autonomous traffic responsive control of up to 120 local controllers without reference to the central computer, which is provided only for centralized monitoring of system performance and equipment status. The local controllers within each region are grouped into "systems" and "sub-systems." Systems of signals do not interact with each other, since they are typically geographically unrelated. Sub-systems, however, do interact and may link together to form systems.

The sub-system may be considered the basic element of control at the "strategic" or multi-intersection level and typically consists of between one and ten intersections, which compose a discrete traffic entity. In response to detected variations in demand and capacity, the strategic control algorithms select the appropriate green splits, offsets and

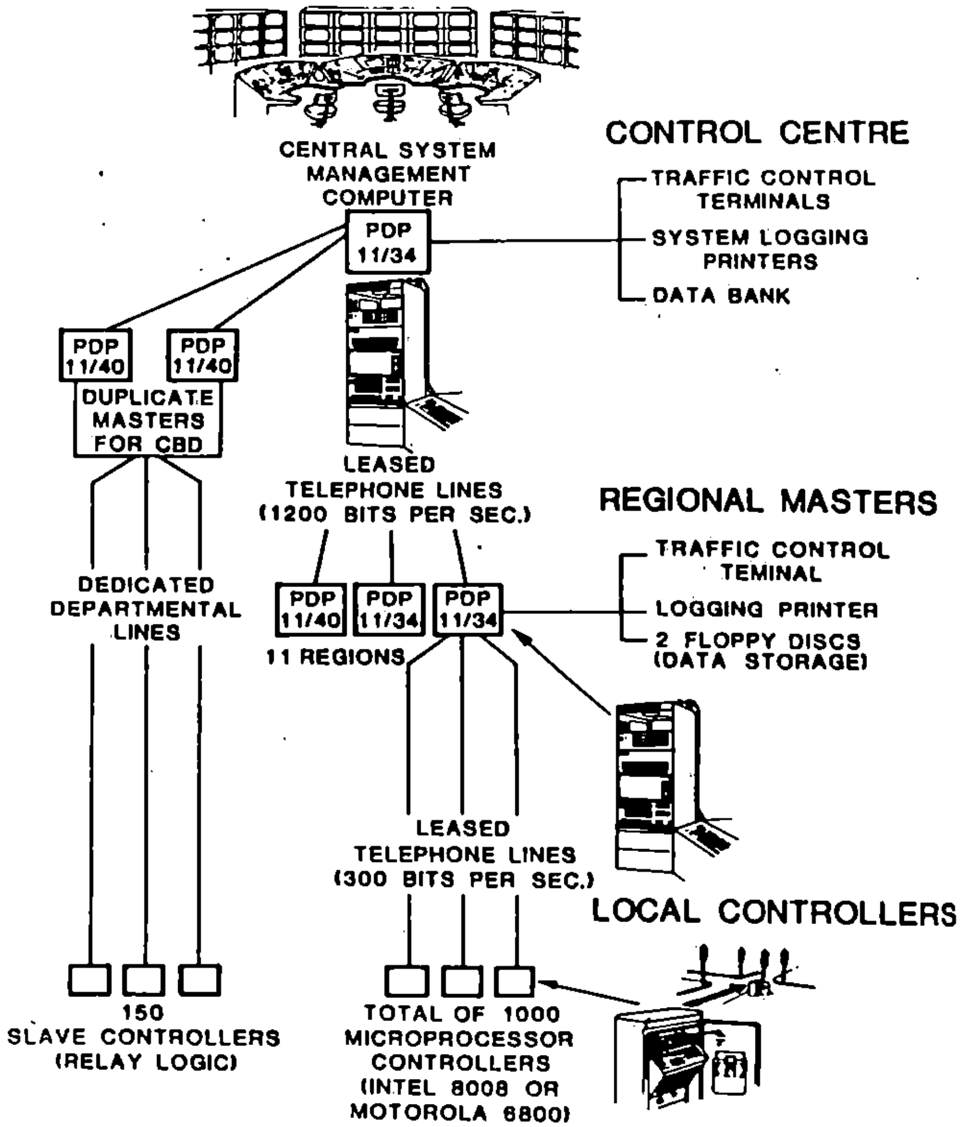


Figure 5-4. S.C.A.T. Computer Hierarchy

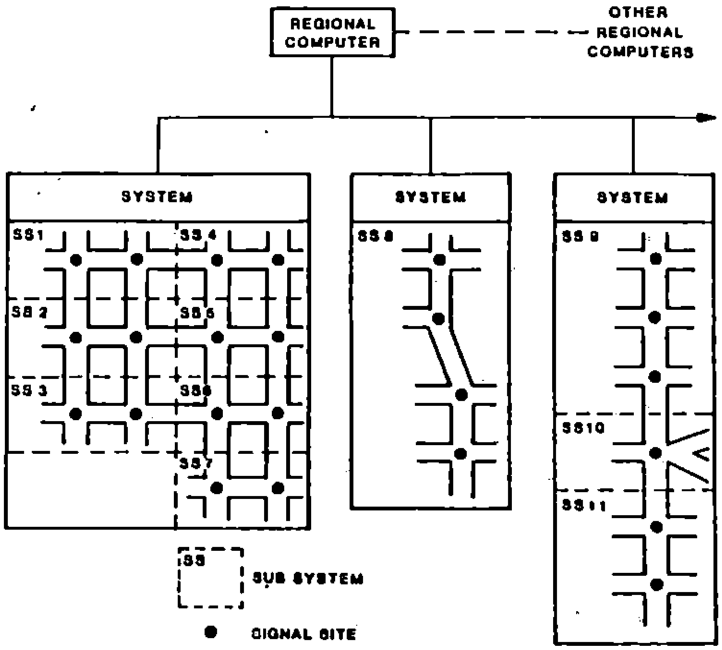


Figure 5-5. Regional Computer Control of Traffic Systems and Subsystems

cycle length for each sub-system and the offsets that are to apply between sub-systems. Four "background" green split plans are provided for each intersection, and the selection of split plans is made at the sub-system level, based on the requirements of the critical intersection in the sub-system. A subsystem contains only one critical intersection. Five background internal offset plans are provided for each subsystem. These plans determine the offsets between intersections within the sub-system. Five external offset plans are used for linking adjacent sub-systems. All intersections within a sub-system operate on a common cycle length.

All intersections are equipped with inductive loop vehicle detectors on all approaches. These are located in each lane immediately in advance of the stop-line and perform the dual functions of providing traffic flow data for strategic control and local or "tactical" vehicle actuation. For strategic operation, the local controller passes the number of vehicles counted during the green time on the approach and the total time that the loop was unoccupied during the green to the regional computer for each detector defined as a "strategic detector."

SCAT Algorithms

A great deal of effort has been expended in the development of algorithms to optimize traffic operation in terms of delay and number of stops. These algorithms control the three basic parameters of a coordinated signal system: green splits, offsets and cycle lengths. SCAT does not utilize a mathematical traffic model coupled to a signal timing optimizer. Instead it uses the available data, describing traffic flow conditions as they occur crossing intersection stop lines.

The model's algorithm use strategic detector information to calculate, on a cycle by cycle basis, the phase split plan, internal offset plan, external offset plan and cycle length to apply to the sub-system for the next cycle, together with an incremental modification to the splits and offsets.

The most important traffic parameter the SCAT algorithms use is one analogous to the degree of saturation. It is defined as the ratio of the effectively utilized green time to the total available green time. To determine the effectively used green time, the time during which unused space is "crossing" the stop-line is subtracted from the total available green time. However, not all the space crossing the stop line is unused because each vehicle has associated with it a space which is a function primarily of speed and which cannot be zero.

SCAT System Operation

The normal mode of coordination is a real-time adjustment of cycle length, split and offset in response to detected variations in demand and capacity. Maximum freedom, consistent with good coordination, is given to local controllers to act in the traffic-actuated mode. The system is designed to automatically calibrate itself on the basis of monitored data. This minimizes the need for manual calibration and adjustment and reduces the amount and importance of pre-prepared data.

For control purposes, the total system is divided into a large number of comparatively small sub-systems varying from one to ten intersections. As far as possible the sub-systems are chosen to be traffic entities, and for many traffic conditions they run without relation to each other. As traffic conditions demand, the sub-systems "marry" with adjacent sub-systems to form a number of large systems or one large system. This "marriage" of sub-systems is calculated in much the same way as are the inter-relationships between intersections within a sub-system. Thus, there is a hierarchy of control, distinct from the hardware hierarchy.

The data for each sub-system specifies minimum, maximum and geometrically optimum cycle length. Cycle length and the appropriate plan are selected independently of each other to meet traffic demand. As noted above, various system factors are calculated from strategic detector data, which are used to decide whether the current cycle and plan should remain or be changed.

Each sub-system has four linking plans, which define the conditions for "marriage" with other sub-systems and which use strategic data in much the same way as sub-system plans. When a number of sub-systems are linked together, the cycle time becomes that of the linked sub-system with the longest cycle time. The combinations of sub-system plans, link plans between sub-systems, variable cycle length and variation of offsets provide an infinite number of operating plans.

Strategic options are available that provide for the operation by minimum delay, minimum stops or maximum capacity. These may be either permanent options or dynamically changed at threshold levels of traffic activity. During normal operation of the system, the regional computer notifies each local controller of a default mode, which can be a "lamps off," "flashing," "isolated" or "cableless link" operation. The cableless link operation is the normal default mode, since it provides an effective linking system without the need for a master, through built-in software programs ("cableless link" is equivalent to "time-based coordination" in the U.S.).

<u>Detector Data Processing</u>

SCAT makes more use of detector data than most other traffic control systems. In addition to measuring volumes and detecting vehicle presence, SCAT uses detector data to estimate real-time saturation flows and intersection capacities. This is significant, because the capacity of any traffic lane is not a constant. The flow at saturation levels varies due to many factors such as weather, time of day, parking, pedestrian friction, downstream conditions and types of vehicles.

Simple volume and headway information cannot show the difference between these variations and changes in actual demand and can lead to gross errors in operation. Data from presence detectors can also be evaluated to obtain the information essential to describe all of the flow parameters. To do this, the detector locations must be close to the intersection so that a high correlation exists between the signal timing and traffic measurements. In other words, the information will only directly relate to the intersection's capacity if the measurements are made when the traffic should be moving at saturation flow with a green signal. Remote detector locations do not provide this direct correlation and, consequently, assumptions about intersection capacities must be made.

In the SCAT detector database, the highest flow rate recorded is stored for calibration purposes. The occupancy that occurred when that flow rate was attained is also recorded. The following data relating to maximum flow are also collected by the system:

- headway (time),
- loop occupancy time,
- space-time between vehicles, and
- speed.

Maximum flow is assumed to occur when only cars are present. Loop length is known and the length of cars is assumed, so speed can be calculated from detector data.

The above data are compared against the cyclic data in the various model algorithms to determine traffic flow status. The speed/spacing relationship defined by Wardrop is used to determine whether speed and hence actuation flow have varied (i.e., if the actual average space/time is less than the reference space/time, then speed is less than optimum and flow rate has decreased).

Where the flow has deceased due to lower speeds, the saturation flow of the lane can be assumed to have been reached due to intersection factors. If the decrease is excessive, then assumably it is due to downstream conditions. If the actual average space/time is larger than the reference data, it is interpreted as a reduction in demand.

<u>Experience of SCAT</u>

The SCAT method has been evaluated in Newtown (a suburb in Sydney) and in Parramatta. The first evaluation was carried out in an arterial road. An initial trail on a length of arterial road showed advantages in journey time over optimized, fixed time signal coordination of 35 to 39 percent in peak periods [Sims]. Later, researchers discovered that the fixed-time plans against which it was tested may not have been optimal. The second evaluation took place in Parramatta, where both open and closed networks were included and fixed-time plans were prepared using TRANSYT. The sites chosen included the CBD of Parramatta and its adjacent arterial road, the Great Western Highway. Both evaluations used the floating-car technique, which involves the use of survey vehicles driven along prescribed routes during predetermined time periods.

The results for the CBD and Great Western Highway studies are shown in Table 5-2. Four time periods were used: 12:00 to 8:00 p.m., 12:00 to 2:30 p.m., 3:00 to 6:00 p.m., and 6:30 to 8:00 p.m. As shown in this table, the performance of SCAT and TRANSYT did not differ greatly in the CBD. This confirmed that the arterial network was highly constrained. SCAT was found to be 6 percent worse than TRANSYT in journey time in the lunch period and 9 percent better in stops in the late evening period. Over the full survey period, SCAT was 2 percent better than TRANSYT in journey time and 1 per cent better in stops, but these results were not statistically significant at the 5 percent level. However, cycle lengths were constrained by pedestrian movements, preventing use of more optimum cycle times for vehicle progression. It was therefore not possible, either with SCAT or TRANSYT, to achieve good progression on most of the major routes through the CBD [56, 57, 64]

The less constrained arterial part of the network (Great Western Highway) showed marked performance improvements as a result of coordination. SCAT consistently performed better than TRANSYT, although the 4 percent difference in travel time was not statistically significant at the 5 percent level. SCAT reduced stops by 25 percent compared with TRANSYT.

The overall result including both the CBD and the Great Western Highway and all four survey periods showed that SCAT caused travel times similar to TRANSYT, but 9 percent fewer stops.

Table 5-2. Comparison of SCAT and TRANSYT in Parramatta

Sub-Area	PERCENT DIFFERENCE†							
	Journey Time				Stops			
	Total Period	Lunch	p.m.	late p.m.	Total Period	Lunch	p.m.	late p.m.
CBD	-3	-6*	0	0	1	-2	2	9*
GWH	4	0	2	9	25*	27*	21*	26*
Church Street	-26*	-31*	-31*	-9	-43*	-66*	-34*	-50*
Total Network	-3	-7*	-2	2	4	0	5	9
Total - Church St.	0	-4	2	3	9*	8*	10*	14*

* Significant at the 5 percent level

† Calculated as : $(x_2 - x_1)/x_2$, where x_2 is the journey time (or stop) of Transyt and x_1 is that of SCAT

By April 1986, most major cities in Australia and New Zealand use the SCAT system. At that time 49 planned or completed signal system installations controlled 4,400 sets of traffic signals.

SCATSIM

A traffic simulation model, SCATSIM, was developed based on SCAT. SCATSIM is able to simulate traffic responsive control systems. It is made up of two parts, a modified version of the SCAT software, which maintains all control functions of the on-street version, and a vehicle-by-vehicle simulation model, which can represent vehicles travelling through a road network with signal controlled or priority intersections. These two parts are interfaced by tables that represent the detector inputs to SCAT and signal output to the traffic. In a validation study, SCATSIM showed acceptable results [31].

A Comparison of SCAT and SCOOT

Both SCAT and SCOOT are successful demand responsive traffic control systems in operation today. Evaluations of both systems showed that they perform better than TRANSYT (although not always to statistically signficiant levels), while UTCS in the U.S. achieved mixed results. A comparison of SCAT and SCOOT will further the understanding of these two systems. In the absence of a field evaluation directly comparing SCAT with SCOOT, it is only possible to study the differences in the philosophies adopted by the two methods in optimizing the control elements. Material in this section is taken mainly from Luk's papers [55].

All the SCAT software has been developed in an assembly language specific to the Digital (PDP/11) computers. On the other hand, the SCOOT software is written in a high-level language and is implemented on a variety of computer systems. As already mentioned, detectors in a SCAT system are located at stoplines, whereas SCOOT detectors are located far upstream of the stopline. It is therefore not possible to implement SCOOT software on a signal system with detectors placed for use by SCAT.

The discussion below centers around the differences between the philosophies adopted by SCAT and SCOOT for optimizing the three control elements, cycle time, phase splits and offset.

Cycle time

Although both SCAT and SCOOT vary cycle times according to the level of congestion, the two methods are different in the following aspects.

- The frequency of cycle time change in SCOOT is restricted to, at most, once every 2.5 minutes, whereas SCAT can vary cycle times once every cycle (although both methods adopt small increments of a few seconds in each update of the cycle time).
- The level of congestion in SCAT is indicated by the degree of saturation measured at the stoplines of the preselected approaches in a network. In SCOOT, the degree of saturation of an approach is estimated from measured flow upstream of the stopline, and a pre-determined value of saturation flow. SCAT has an advantage in that it self-calibrates saturation flows according to changes in intersection geometry, lane utilization, weather conditions or driver behavior.
- SCOOT has a double-cycling facility that operates in both directions, i.e., an intersection can operate at half of the network cycle time when the network cycle time is too high, and it can return to single cycle when there is less traffic demand at that intersection.

In summary, the two cycle time optimizers operate differently. SCAT directly measures the degree of saturation, but SCOOT has the advantage of automatic double-cycling. SCOOT can only estimate the degree of saturation and has a slower response in varying cycle time. The two methods appear to be comparable in the accuracy of choosing an appropriate cycle time for the prevailing traffic conditions.

Phase splits

The four predetermined phase split plans in SCAT stipulate different vehicle actuation tactics for four different flow patterns that usually correspond to different time periods of the day. These vehicle actuation tactics are facilitated by the location of vehicle detectors at the stoplines in a SCAT system. SCAT therefore responds to fluctuating traffic demands at a local intersection without the usual delay of one cycle time.

On the other hand, SCOOT relies on modeling the queue length and estimating the degree of saturation to optimize phase splits in small steps of a few seconds. Phase changes may not be frequent enough to meet the cycle-by-cycle variation of traffic demand. SCOOT is further limited by what it can model. For example, a vehicle actuation tactic such as phase skipping cannot be modeled in the current version of SCOOT. A signal system using SCOOT would not be able to utilize the vehicle actuation facilities available from a modern microprocessor controller. This is in contrast with the SCAT method, in which phase optimization is largely performed in its local controllers and which is in line with the trend toward the decentralization of control to the local controllers. In summary, the SCAT method, together with its system configuration, appears to be more capable of quickly changing phase splits to reduce delays and stops, but the SCOOT model provides a more flexible number of alternative phase settings.

Offset

An offset is closely related to cycle time and phase splits, and all three control elements should ideally be optimized simultaneously. In a SCAT system, the offset plans are predetermined and selected to match the current cycle time, phase plans or the directional splits of traffic flow. Special care is required for preparing offset plans for CBD or grid-type networks. The offset can also be modified, as an option, according to the level of congestion. The objective of the SCAT offset selection algorithm can be broadly treated as the maximization of the bandwidth for platoon progression. With detectors located at stoplines, it is difficult to monitor platoon progression, and there is currently no feedback information regarding the performance of the offset adopted.

The SCOOT optimizer has the benefit of using a traffic model to optimize all control elements, including offsets. The flow profiles are used to predict the effect of small offset changes on queue lengths in a network. The optimization objective function is the sum of queue lengths, which can be displayed according to selected approaches or sub-areas. The SCOOT detector loops are located close to the upstream intersection and offer the benefit of indicating when a queue fills the space between the two intersections. This allows the system to take special action to prevent blockage of the upstream intersection. The SCOOT offset optimization philosophy therefore appears better defined than the SCAT offset selection method.

CHAPTER 6
INTEGRATED TRAFFIC CONTROL MODELS

Due to budget constraints, the most significant trend in the field of transportation engineering and planning is the prevailing philosophy of "making the best use of existing resources." The realization of this philosophy in transportation planning is the concept "Transportation System Management (TSM)." Under TSM every effort is made to maximize the use of existing transportation facilities. Consequently in the area of traffic control there are no plans to develop new models for either on-line or off-line traffic control evaluation. Instead, resources are allocated to make the best use of existing computer models. Within such an environment, system integration is an effort to enhance the capabilities of existing offline computer models by making them work together.

System integration of traffic control models has been attempted using two different approaches:

- functional integration and
- database integration.

In functional integration, computer models are either horizontally integrated (integration of different computer models with the same basic application situation but with different analysis capability, e.g., AAP) or vertically integrated (integration of computer models with different application situations, i.e., TRAFLO). TRAF is an attempt to combine both functional integration structures.

In database integration, data requirements for different computer models are stored in a single database and managed by a database management system. The database management software produces an appropriately formatted input file for the transportation network and model specified. Both approaches aim at easing the use and operation of computer models as opposed to improving the analysis capabilities of the available models.

In the following sections the different integrated modeling packages will be discussed. It should be noted that not all of these modeling packages are currently available to the transportation professional. Some of the modeling packages to be discussed have been released, while others are still in final design, and others are somewhere essentially completed but are still undergoing final testing and revision.

ARTERIAL ANALYSIS PACKAGE (AAP)

The Arterial Analysis Package is a tool for timing traffic signals on arterial streets. It uses three of the most widely used design and analysis programs to provide a framework for solving signal timing problems using commonly available traffic engineering data [37]. It is an integration of three computer models for similar application situations but with different analysis capabilities (i.e., functional integration).

AAP consists of the Signal Operations Analysis Package (SOAP), Progression Analysis and Signal System Evaluation Routine (PASSER II), and Traffic Network Study Tool (TRANSYT). These models are used as independent programs with their own particular strengths and weakness. The AAP provides a common coding structure so that the component programs can be used from a common data file. In addition to the three component programs that form the engineering analysis core of AAP, the package also provides a group of interactive and batch support programs.

Traffic engineers coordinate signals by either maximizing "band width" or minimizing "total stops and delays", however, no computer model can optimize both objectives at the same time. In such a situation, solutions from optimization programs using different analysis strategies are informative to the traffic engineer, who may select one of the different solutions, or use one model's output as another model's input.

TRANSYT optimizes signal offsets and shifts for a given cycle length by minimizing the "performance index" (a linear combination of stops and delays). SOAP specializes in individual intersections, determining optimum signal timing (cycle length, splits, and phase sequence) and dial assignments for multiple time periods under either pretimed or actuated control conditions. SOAP does not select an "optimum" cycle length or phasing; the user must use the design and analysis capabilities of the model to set up and evaluate designs to determine the best mix of measures of effectiveness (MOE's). PASSER II determines cycle lengths, phase sequences, offsets and splits so that the bandwidth along an arterial is maximized. Detailed descriptions of these three models are given in Chapter 3 "Offline Traffic Control Models."

AAP Inputs

The AAP input coding scheme is based on the format used in SOAP 84. Each input record begins with an "identification" field that tells the AAP what kind of data to expect on the remainder of the card. The ID field is followed by a two-column numeric field (to identify time period or duration), a five-column numerical field (to identify beginning time or intersection number), and eight five-column fields for variables (usually data for each traffic movement). The last 25 columns are for alphanumeric variables — usually one-column directional indicators such as N, S, E, W, T(hru), and L(eft) and labels for use in headings. The coding format allows the same "card deck" to be used for both PASSER and TRANSYT with minimal changes. By using the same structure for all kinds of input records, it is easier for the user to learn the coding scheme and to spot errors. Typical cards — the SETUP card and the VOLUME card — are shown in Figure 1.

AAP Structure

Figure 6-1 shows the structure of AAP. The AAP structure consists of

- data input manager (DIM's) to create the input data card deck,
- the card deck,
- the preprocessor, which maps the card deck to the temporary data file, performs diagnostic checking, and then creates input card decks for one or more component programs,
- the three component programs,

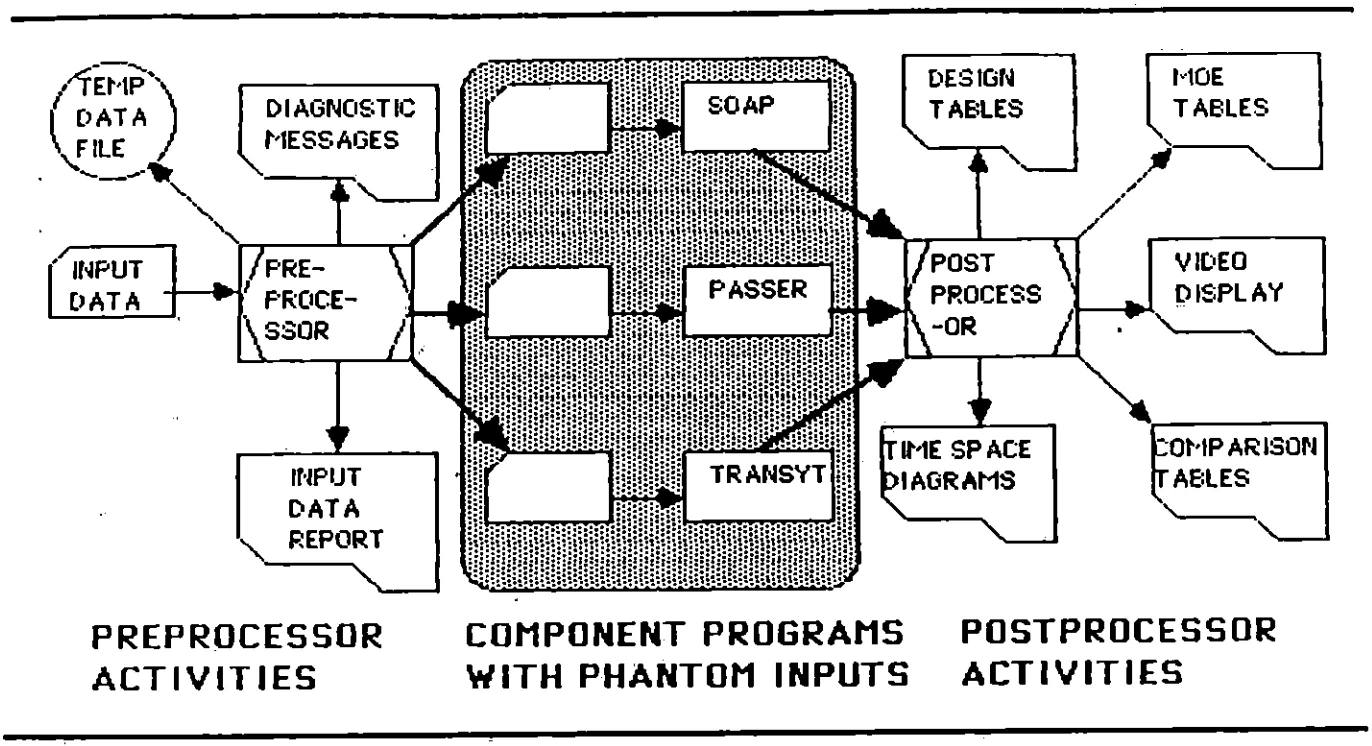


Figure 6-1. The Arterial Analysis Package

- the postprocessor, which creates files used in graphic displays on microcomputers and transmits information between PASSER and TRANSYT, and
- the microcomputer-based graphic display programs.

AAP Outputs

A major problem with using SOAP, PASSER, and TRANSYT before the AAP was developed was in mentally converting from one set of outputs to another while evaluating the results. A consistent set of 80-column output formats was developed for each of the three component programs. This allows the output from the alternative models to be spread side by side so that the results can be compared easily. Although the MOEs generated by the programs are not identical, they are now comparable, facilitating translation of the program outputs to traffic signal controller settings.

AAP-85 is the latest version of this system. It can be run on IBM PC and Apple Macintosh Computers.

INTEGRATED TRAFFIC DATA SYSTEM (ITDS)

System Development

Extensive application of computer models has demonstrated their potential as effective tools in developing traffic control strategies that reduce motorist operating costs; vehicle fuel consumption and emissions; planning, design, and implementation costs of new control strategies; and costly and inconvenient retrofits when problems in a strategy are detected only after implementation. However, differences in data requirements and input formats, the need to be comfortable working in a computer environment, and the perceived difficulty of using these computer models have deterred some traffic engineers and analysts from using these powerful tools. The Integrated Traffic Data System (ITDS) is being developed as solution to many of the problems associated with data availability and coding [82].

ITDS is a product of system integration by database approach. It is currently in final testing. It is a microcomputer-based system composed of hardware and software elements which jointly perform the following functions:

- provide for a centralized microcomputer database to store traffic data in a predetermined format and organization; and
- utilizes this database to generate input data sets for various traffic simulation models and signal timing optimization programs.

ITDS has the capability of being connected to a mainframe computer. This requirement is based on the fact that most current traffic models were designed to run in mainframe environments. Conceptually, ITDS is used to generate input data sets, transmit them to a host computer for processing, and retrieve and reformat (in the case of optimization programs exclusively) the results. A diagram of the ITDS concept is presented in Figure 6-2.

ITDS Features

The main features of ITDS include the following:

- menu-driven software with online assistance for easy, user-friendly access,
- networkwide traffic database storage on a hard disk that allows high storage capacity and access speed,
- database maintenance using a state-of-the-art CODASYL (Conference On DAta SYstems Languages) type database management system,
- input data sets for traffic models generated by either querying the database or supplied directly by users,
- job submission to and retrieval from a remote mainframe computer where the models can be executed by means of communication lines,
- ability to use an optimization model's output as input to other traffic models,
- database management security,

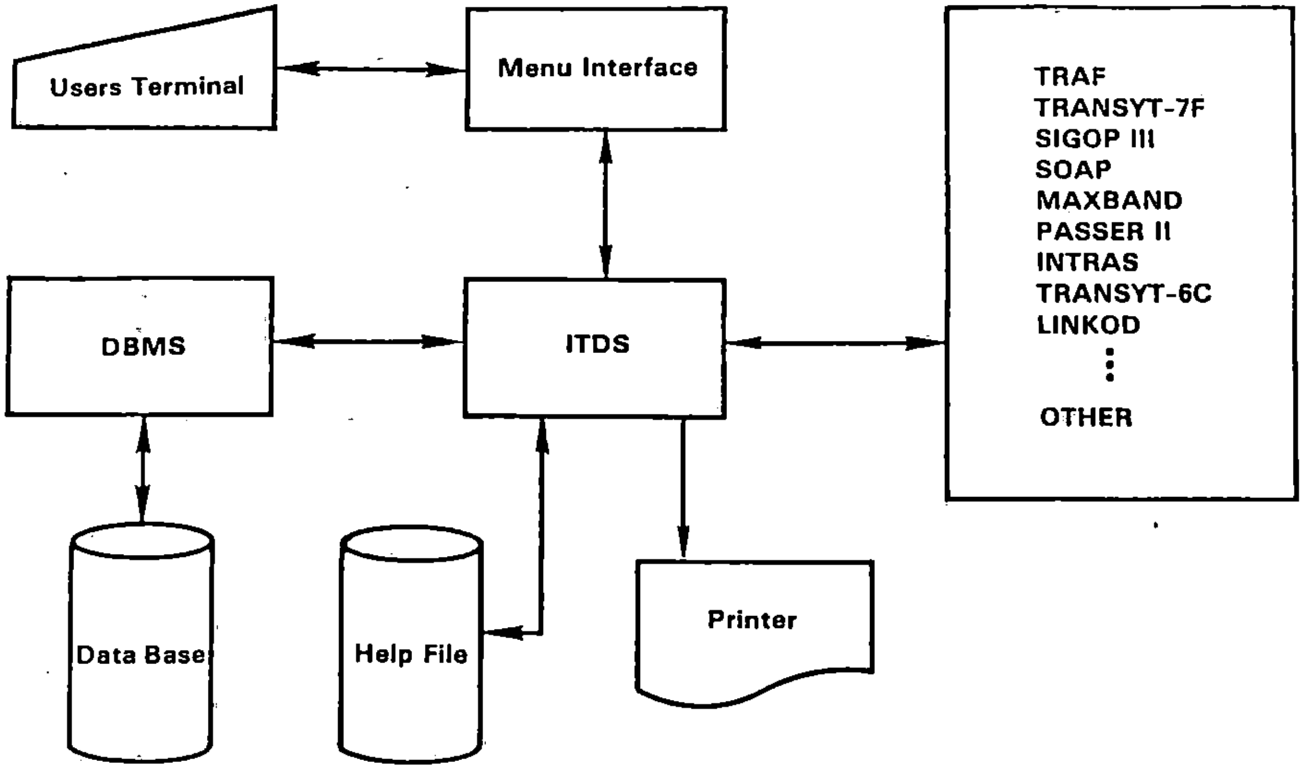


Figure 6-2. ITDS: A Data Base Driven Interface to Traffic Models Using a Microcomputer

- data requirements listing, on a per model basis, for all interfaced models, and
- adaptability to meet future needs including interfacing to other traffic models and the use of color graphics, light pens, and a "mouse."

ITDS — Database Approach

The most important task in developing ITDS was to design a file structure that could store any kind of traffic engineering data in a generic manner and still be readily accessible to any external application software (traffic models). This requirement led to the development of the database approach.

A database is a centralized collection or storage of data for use in one or more particular applications. In traffic engineering, these data include the physical and operational characteristics of a traffic network. In most traffic modeling applications, specific means of handing data storage requirements have been developed. Typically, this data storage has been in the form of fixed-length sequential or "flat" files, which are appropriate for most stand alone uses.

Because ITDS was designed to interface to a wide range of models, it has a more substantial data management problem than most similar applications (such as AAP). For this reason, the heart of ITDS is a formal Database Management System (DBMS) — a collection of software that organizes, stores, and retrieves data in a database. ITDS represents the first known formal application of database management theory to traffic engineering.

ITDS is built around the MDBS-III DBMS, an extended CODASYL network-type DBMS, which is designed as a library of software tools for interfacing to a user's application program. ITDS was developed using the Pascal MT+ language running under the CP/M-86 and MS-DOS environments.

The system design of ITDS is future oriented. Because ITDS's database structure is independent of any particular model's requirements, new and revised models generally can be interfaced to ITDS by adding only a new "deck" formatting module. Modifications to the database structure, if required, should be minor and would not affect the operation of the system's other components.

As more and more mainframe traffic engineering models are implemented on microcomputer systems, ITDS can provide a standardized, easy-to-use interface. This eliminates the problem of providing an easy-to-use "front end" to the program.

ITDS System Components

The initial release of ITDS includes four major programs — DBedit, DBprint, NETSIM interface, and TRANSYT-7F interface. The DBedit program allows the user to create and maintain the central database and provides an option for segregating the database for the entire network into user-definable subnetworks. The DBprint program creates hardcopies of the information stored in the database and provides for a limited database querying capability.

The NETSIM and TRANSYT-7F interfaces are the programs that create the input data sets (for the respective traffic models) by querying the database. The TRANSYT-7F interface supports both the mainframe and microcomputer versions of the model. The NETSIM interface was designed for the "TRAF" NETSIM. (This is currently a mainframe product but will eventually be converted to a microcomputer application.) Interfaces for additional traffic models, such as PASSER II, MAXBAND, SOAP, SIGOP III, and the computerized Highway Capacity Manual, are being developed as part of the ITDS support and maintenance activities. Graphics capabilities also are being developed, and the database querying capabilities are being enhanced by the Oak Ridge National Laboratory.

ITDS Benefits

ITDS is intended to fill the gap between mainframe and microcomputer technology as related to traffic modeling. Traditionally, traffic models have been developed to run on mainframe computers. ITDS is attempting to take full advantage of state-of-the-art microcomputer hardware and software, enabling the preprocessing of data to create input files offline, job submission and output retrieval, and the automatic use of optimization programs' output as input to other models. ITDS will thus be the middle step between the traditional approach of using a mainframe computer for processing and creating input files manually, and the newer approach of creating input files and executing the programs locally and interactively on a microcomputer.

Additionally, ITDS is intended to reduce the overhead costs associated with traffic models for training, coding and submitting input files for processing. It provides an accessible, manageable, and centralized database, and a simple, user-friendly work environment. It is also designed to provide information on input data needs to the user, so that data on roadway geometry need only be collected once for all models.

An important contribution of the ITDS project has been the development of a database scheme covering the data needs of a range of traffic engineering and network analysis models. The resulting data model has been designed to adapt to the changing needs of traffic engineering simulation and optimization tasks and also can be adapted to other transportation applications such as planning, mass transit, and safety, specifically in accident record management.

AN INTEGRATED SIMULATION MODEL (TRAF)

System Development

TRAF is an integrated system of various levels and varieties of traffic simulation models. It includes simulation models dealing with the various traffic situations (freeways, arterial networks, and two-lane two-way rural roads) in two different levels of detail (macroscopic and microscopic). Essentially, TRAF is a result of functional integration in both horizontal and vertical directions.

In 1975, FHWA investigated the feasibility of creating an integrated traffic simulation system. The system was to be able to represent traffic flow on any existing highway facility. It was concluded that such a system was feasible. A modeling system was proposed and guidelines for its development were established.

The most important finding of the FHWA study was that computer software maintenance and technical support to users consume more resources than any other activity in the life cycle of a computer program. Maintenance costs are estimated at 40 to 90 percent of the total cost of a system. Experiences with most large computer programs that are never fully "debugged" support these conclusions.

Current trends emphasize these findings. For many years advances in microelectronics have dramatically reduced the hardware-associated cost of data processing. In contrast, the cost of human time, and thus the cost of producing software to run on the newer and faster computers, increased steadily. Thus, emphasis in software development is being shifted from computational efficiency to human efficiency.

The integrated traffic simulation system (TRAF) is envisioned to facilitate maintenance and support operations in two ways:

- with only one simulation system to maintain and support, these operations could be centralized, and
- those operations could be automated largely by developing a separate computer program (called an operating system) that would aid in making changes and managing model capabilities.

The TRAF system is designed to represent traffic flow on any existing highway facility. Thus the system can deal with different traffic situations and should minimize the effort needed from the user in order to make the system easier to operate. From the user's standpoint, since TRAF will be a single source of traffic simulation programs, the user need be concerned with only one set of documentation and one input and output format. This standardization should put an end to the confusion caused by the diversity of simulation approaches and formats. It should also considerably reduce the overall learning effort required from users in connection with the application of traffic simulations.

The creation of TRAF does not involve new model development, but the enhancement of what is regarded by the FHWA as the best traffic simulation logic available. In TRAF this logic is in the form of modularized subroutines that are stored in a master file. A program tailored to a particular application is generated by an "operating system" that selects the needed subroutines, adjusts their dimensions, and integrates them. This flexibility should minimize necessary computer resources because the generated programs contain only the user's selected features and dimensions required by the desired applications.

TRAF Features

TRAF is not an integrated model, but rather an integrated simulation system. Traffic on urban networks, freeways and two-lane rural roads can be simulated at different levels of detail. Once the needs of a particular application are specified (levels of detail, types of roads to be simulated), the appropriate subroutines are selected from the master file and the TRAF operating system generates a simulation program tailored to these needs. Thus, the user is not overwhelmed with unnecessary features. The following procedure has been established to integrate the best available traffic simulation logic into TRAF:

(a) The candidate models to be integrated are reviewed, their basic functions identified, and the functional modules (or subroutines) defined and described.

(b) The modules that perform the same or equivalent functions are combined into a single module that performs the common function.

(c) The modules resulting from the two previous steps are arranged in logical structure according to the anticipated functions of TRAF.

The traffic simulation logic that is being integrated into TRAF is contained in the component models shown in Figure 6-2. The names of the component models denote their place of application and level of detail. The prefixes NET, FRE and ROAD indicate urban networks, freeways, and two-lane rural roads, respectively. The suffix -SIM means microscopic, and FLO macroscopic. It is believed that this logic represents the best traffic simulation techniques available and will satisfy most traffic simulation needs.

The current version of the TRAF system represents the integration of several different traffic simulation models. Each component model in TRAF is designed to represent traffic on a particular physical environment (i.e., urban streets, arterials, two-lane, two-way rural roads and freeways) at a specific level of simulation details (i.e., microscopic or macroscopic). Microscopic simulation models represent movements of individual vehicles including influences of driver behavior. The effects of very detailed strategies such as relocating bus stations or changes in parking restrictions may be studied with such models. The price for such detail, however, is in the computer resources of time and memory consumed. Less detailed strategies, involving changes in circulation patterns for example, may be studied with macroscopic simulation models. These models may also be used to gauge the repercussions of very detailed strategies outside the boundaries of the immediate area in which they are implemented.

The ability to combine these models in a single analysis run is a major feature of TRAF. In addition, a traffic assignment model is included in the system. This model is designed to expand the applicability of traffic simulation modeling to transportation planners in addition to traffic engineers. The traffic assignment model in TRAF internally translates the origin-destination data available to the planning community into a form suitable for use by the simulation models. All component models of TRAF are described in the following section of this chapter.

Another significant feature of TRAF is its flexibility in partitioning the analysis network into its component subnetworks. This feature allows the user to "tailor" the model to his needs. For urban networks, where a high level of detail is required, when the inputs are accurate, bus operations are important and extensive, and when resources are available, the user can select the most detailed urban arterial model — NETSIM. Where some of the above conditions are absent, the user can reduce computer requirements by specifying NETFLO, and still obtain results of satisfactory accuracy.

The subnetwork configurations resulting from partitioning can also reflect the complexities of the network and traffic congestion. For example, the CBD of a city could well be modeled with NETSIM, the peripheral arterials with NETFLO Level III and suburban grids or major arterials with Level I or Level II. Freeway "subnetworks" must be modeled with FREFLO or FRESIM, just as a rural road must be modeled by ROADSIM [TRAF user's guide].

TRAF Component Models

NETSIM is a microscopic model for urban networks simulation. It has been enhanced several times since its initial release. It has been reprogrammed to conform to TRAF standards. Details of NETSIM are described in Chapter 3. FRESIM, the microscopic freeway model, is primarily the freeway portion of INTRAS, a microscopic freeway corridor model that has been tested by FHWA. FRESIM is being enhanced and reprogrammed before becoming part of TRAF. Details of the INTRAS model are described in Chapter 4.

The macroscopic models for urban networks, NETFLO and FREFLO, form a subsystem called TRAFLO. NETFLO was developed according to TRAF programming standards, and FREFLO is essentially the existing MACK freeway model, reprogrammed and adapted to the TRAF environment. NETFLO is beginning its implementation phase, while FREFLO is going through enhancement and testing. NETFLO has three levels of detail and sophistication. Level 1 requires the greatest amount of input detail and provides the most precise calculations. Level 3 is the least detailed of the three levels.

Finally, ROADSIM, the microscopic two-lane, two-way rural road model is basically the TWOWAY model developed by the National Cooperative Highway Research Program. It is being reprogrammed and integrated into the TRAF system.

The component programs in TRAF are not simply combined together; instead, they are reprogrammed to fit the system's structure, especially TRAFLO and ROADSIM. These two programs will be discussed in more detail below.

TRAFLO

The objective of the TRAFLO model is to provide an efficient tool which can be used to test and evaluate traffic management strategies that are applied over a large area. This model is being developed in response to the need for the examination of transportation system management actions, and the need to examine the impacts of TSM actions on one facility on the functioning of other related facilities (i.e., what effect do freeway improvements have on parrallel arterials). The model was designed to satisfy the following requirements:

- the model must provide values of all relevant measures of effectiveness (MOE) which describe traffic operations on urban streets and freeways. The scope, accuracy and level of detail of these MOE must be adequate for the purpose of evaluating traffic management strategies;
- the model must exhibit the flexibility necessary to accommodate the widest possible range of such strategies, including those which affect route and model choice;
- the model must be able to represent a region of approximately 2,000 intersections, whose traffic environment includes networks of freeways, arterials, and other surface streets;
- the model must be designed to satisfy these requirements within a reasonable amount of computer resources. It should be operational on virtually any general purpose computer;
- the program must be easy to use, requiring as little information as possible so as to minimize the cost, effort and level of expertise needed for its implementation.

- the computer program of the model must be easy to understand, to maintain, to update, and to extend in scope.

The simulation is macroscopic in nature with three separate levels of detail:

- Level 1 is the most detailed level of traffic representation. It is designed to explicitly treat traffic control devices, including all channelization options, and to describe the traffic operations at grade intersections in considerable detail. Careful distinction is made between general traffic operations of private automobiles and mass transit vehicles servicing passengers at bus stops located along fixed routes. In addition, trucks and car-pool vehicles are explicitly considered. Other features include actuated signal control logic, right-turn-on-red, pedestrian interference, and midblock source/sink flow. A wide range of MOE are provided as output.
- Level 2 is an extension of the traffic flow model embedded in the TRANSYT signal optimization program [Yedlin]. Level 2 is computationally faster than level 1, is less detailed and includes fewer features. Nevertheless, the traffic flow patterns are carefully described in the form of statistical histograms. These histograms express flow rate as a function of time on each network link, stratified by turning movement; buses are treated in somewhat more detail. Platoon dispersion is treated explicitly and service rates at intersections are related to turn movements and the signal control. This level provides the same output MOE as level 1.
- Level 3, which is the fastest computationally, is the least detailed and is applicable only to arterials. The platoon structure of traffic is not represented; traffic flow and signal control are described in terms of aggregate variables. However, traffic is stratified by turn movement to reflect the differing service rates associated with turning vehicles. Bus traffic is treated explicitly, as is signal coordination and the time-dependent

behavior of traffic. Congested conditions are accommodated and spillback is considered. While the detailed behavior of traffic at intersections is not explicitly represented, the associated impedances are modeled.

A separate model treats freeway operations which can be partitioned into a number of subsystems to save computer costs.

TRAFLO also incorporates a traffic assignment model to extend the functions of the package to include transportation planning in addition to traffic engineering. An existing assignment model TRAFFIC is interfaced internally to the traffic simulation model to facilitate examination of the effects of traffic controls on route choice.

ROADSIM

Simulation of rural traffic on two-lane roads developed at a slower pace because two-lane flow is complicated by platooning and passing decisions and is therefore not easily modeled. Also, the low volumes on rural two-lane roads usually do not make simulation cost-effective. In addition, two-lane traffic simulation requires numerous computations, which require considerable computer time and memory, particularly for microscopic models. To date, most of the two-lane simulation models are microscopic. Microscopic models simulate and trace individual vehicles and are more accurate and realistic than macroscopic models, which simulate traffic using aggregate variables such as traffic volume and average speed.

Simulation models for two-lane roads have evolved over the past two decades. Most of the early attempts contributed little to the study of two-lane flow at a practical level. However, those attempts were stepping stones for other sophisticated simulation models currently available. ROADSIM, the latest product of the evolutionary process of two-lane simulation model development, is not a new model with new methodology and logic but rather a reprogrammed version of an earlier model (called TWOWAY) with modified routines and adaptations from other models [Morales].

The TWOWAY model can "move" individual vehicles in accordance with several parameters specified by the user. The vehicles are "advanced" through successive 1-second intervals taking into account roadway geometry, traffic control, driver preferences, vehicle type and performance characteristics, and passing opportunities based on oncoming traffic. Spot data, space data, vehicle interaction data, and the overall travel data are accumulated and processed. Several statistical summaries are reported.

TWOWAY logic was modified to include logic elements from two other simulation models — INTRAS and SOVT. INTRAS, a microscopic freeway simulation model developed in 1976 for FHWA, provided the basic car-following logic to TWOWAY. This logic is based on the premise that a vehicle that is following another will always maintain a space headway relative to its lead vehicle which is linearly proportional to its speed. This premise was much simpler than the one used in TWOWAY and thus easier to calibrate. SOVT, a microscopic two-lane simulation model developed in 1980 at North Carolina State University, provides its vehicle generation logic to TWOWAY. This logic emits vehicles onto the simulated roadway at each end. For low volumes, the Schuhl distribution used in SOVT provides a realistic approximation of vehicles generated. However, for high volumes where traffic density approaches queuing, a shifted exponential headway distribution is used. The new TWOWAY model was reprogrammed according to FHWA specifications, modified with new input and output subroutines, and renamed ROADSIM.

CHAPTER 7
FUTURE DEVELOPMENTS AND RESEARCH

This chapter discusses planned and desired improvements to existing traffic models and modeling capabilities. The chapter is divided into the following sections:

- Computer graphics,
- Ease of Use,
- Interactive Programs,
- On-line Simulation,
- Intelligent Traffic Control Systems, and
- New Theories of Traffic Control.

In the sections below, the future direction of each of the above topics is described. Needs for further research and the author's opinion on the requirements for WSDOT research support are presented. It must be emphasized that the ideas presented in this section are only the authors' opinions, which are based on the available literature and the project team's work experience. The opinions do not represent the offical policy of the Department, nor do they represent the stated goals of specific individuals within the Department.

Many of the above topics are interrelated. For example, improvements in graphics combined with the speed and capacity of microcomputers allows:

- programs that are easier to use,
- on-line simulation, and
- intelligent traffic control devices.

Because these interrelationships exist, some discussions may be appropriate for more than one subject heading. While this chapter will describe each subject area only once, it will indicate the interrelationships that exist and provide references to related discussions within the chapter.

COMPUTER GRAPHICS

The area of traffic modeling has a variety of uses for graphics. Graphics can improve the output capability of a model, can ease the development of input data files, and can even be used as part of the input process. The use of graphics for input is discussed below in the section on ease of use. This section only describes the use of graphics to display output from the model (both during and after the model executes).

Full animation technology has been available for only a few years and is relatively expensive. However, increases in the power of graphics displays, improvements in graphics software and decreases in the cost of computer hardware and CPU time (such as the use of microcomputers) make sophisticated use of graphics practical in the immediate future.

Graphic output can be an excellent tool for describing the effectiveness of a particular signal plan. Very soon simulation programs will begin to make greater use of graphics for displaying output. In a simplistic (but still future) application, graphic outputs might show queue lengths at designated intersections, average number of stops, total delay, and various other forms of output information. Such displays will allow the user to grasp the overall operation of a network more quickly and meaningfully than can be done with tabular output.

More sophisticated graphic output might display the simulation program in operation (i.e., simulate the vehicles moving through the network). Animated operating displays are useful for explaining the simulation program to non-experts but can be even more useful to users. For example, finding mistakes in an input data file is easily done by looking at how the computer represents the network's operation. Left turns coded as right turns or mistakes in signal phasing stand out immediately.

The first and most significant graphic improvement under development is the graphic output of NETSIM results, currently being readied by FHWA. This extension of NETSIM should allow the model's simulation results to be visually represented.

However, the needs for research in this area are many, and the NETSIM effort is only a small beginning. Existing models such as TRANSYT, FRESIM and TRAFLO could be significnatly enhanced with the addition of animation graphics if for no other purpose than the review of input data. The application of graphics to existing computer models is not simple and will require relatively large amounts of staff time. While most of this research is outside the scope of the normal WSDOT research effort, the Department should work to support its continued development.

EASE OF USE

One particular problem with existing modeling capabilities is the difficult and time consuming nature of running the models. In modern transportation agencies, resources are severely limited, and it is often politically difficult to spare the resources necessary to maintain and operate the appropriate traffic models. If the existing models required less time to use and were easier to learn and operate, their use (and therefore benefit) would increase significantly.

Considerable work has been done since the advent of the microcomputer to simplify the use of available models. Most of the work done to date has focused on simplifying the process of constructing input datasets for models. Graphically oriented and/or menu driven pre-processor programs have been written for several commonly used traffic models (e.g., EZ-7, EZ-Passer). The QRS II model includes a graphically oriented network creation procedure. ITDS and AAP are also attempting to ease this part of model use through using a database approach for creating input data files on request for a number of computer programs based on a single data file.

While modern microcomputers lack the mainframe's power and speed necessary for running large traffic simulations, they do offer significant improvements over most mainframes in the area of ease of use. One specific advantage of the microcomputer has been the standardization of an operating environment. This has allowed the distribution of computer programs in executable form and has eliminated problems caused by the user having to compile the program and determine its interaction within his/her computer system. Consistent computer platforms are necessary for the creation of add-on products such as EZ-7, and microcomputers appear to provide that stable environment, at least for the time being.

Work is needed not only to continue this effort but to extend the effort by making the model setup, operation and review of output more obvious and straightforward. WSDOT needs to support this in the development of any new software it sponsors. The Department should remember, however, that providing ease of use means additional expense in the development of the software.

INTERACTIVE PROGRAMS

One extension of the ease of use concept should be towards interactive program execution. With interactive programs, operators will be able to interrupt the models during execution to change various parameters. Increasing the use of graphic input and output of data (particularly in the area of animation) should improve the utility of interactive models and increase the desire for interactive programs. Programs may still run primarily in batch mode and generate output for later presentation and use, but interactive capabilities will slowly increase.

Once the operator can see at a glance from a graphic display how the network is operating, s/he will want to be able to control that operation. Fortunately, if animation capabilities are added to traffic models, the step to fully interactive programs will be reasonably simple. This is because with graphic displays, the computer commonly instructs the display device to draw some complicated picture and then awaits confirmation that the picture was indeed drawn correctly before proceeding with the computer application. Consequently, it is a small step to allow the operator to interrupt, change, or even restart the controlling program, because the computer is already waiting for a response from the operator before continuing. In addition, the use of interactive program execution should result in more efficient usage of computer resources, since the user can abort runs that contain obvious input errors or those that produce illogical results before they complete execution.

WSDOT could obtain significant benefits by using interactive programing, particularly in the area of on-line simulation, discussed below. However, WSDOT does not need to actively promote the interactive concept at this time, as true interactive programing requires more graphic capabilites than exist at present or in the near future.

ON-LINE SIMULATION

The "ultimate" ease of use situation is one in which the transportation network is already coded, the input traffic data are automatically collected and entered in the program, and only the controling parameters need to be provided by the user. These criteria describe an "on-line" system in which the model uses actual vehicle data collected from detectors in the field. Such a model would then simulate probable traffic flows based on the initially observed detector information.

Forerunners of on-line systems are already appearing in some of the SCOOT and UTCS 1.5 installations and the IMIS system in New York (and to a lesser extent in many SC and DI systems across the country). These systems collect and monitor real-time traffic information. This process includes aggregating the real-time information into various formats for use in computer programs. At this time, only SCOOT actually performs real-time traffic modeling with the data. The other systems store the data for later use, or simply present the collected data for information purposes.

On-line evaluation systems will likely be provided to operators of various computer traffic control systems in the future. The operator might, at the touch of a button, start an interactive graphic simulation. The program would start with conditions that are identical to those that are in the real network at the time the button is pushed. The program (running four or five times real speed) might allow the operator to predict and then forestall congestion by suggesting and testing alternative traffic control strategies. SCOOT does this to a very limited degree, but considerable improvements can still be made in this area. For example, the techniques have yet to be applied to freeways and integrated freeway/arterial systems.

Even when congestion could not be prevented, such a predictive capability would be useful in many cases. For example, even when a control algorithm was unable to provide an appropriate response, it would be useful to foresee how the traffic disturbances would propagate during congestion. With this information, police and transportation agencies could respond appropriately.

The ultimate goal in on-line simulation is a program that runs continuously, uses vehicle detector data as input, and checks itself against the real traffic so that the simulation program can adjust itself for changes in the vehicle mix and driver behavior without human intervention. The simulation would run faster than real time and provide signal control and VMS messages for the highway system automatically. Such systems are several years away, but could conceivably evolve from existing modeling programs. WSDOT could benefit greatly from such a system, initially in the Puget Sound region, but also in the other urban areas of the state as they experience growth. Development of such systems should be encouraged as part of FAME, but the bulk of this research must be done on a national level.

INTELLIGENT TRAFFIC CONTROL SYSTEM

A related but separate issue is to extend the sophistication of models to include the capability to "reason." Robertson has predicted that the late 1980s will see much greater use of the full potential of computers for "intelligent," real-time traffic control . SCOOT is one example of such a system. By "intelligent" Robertson means that machines will be able to infer likely future traffic patterns using reasoning processes similar to those of human experts. The machines would then select or develop the appropriate traffic control strategies that would mediate the expected congestion impacts as much as possible. Such systems would operate with little or no human input.

The potential for "intelligent" traffic control is enhanced by the development of expert system technologies and the ever decreasing cost of computing power and resources. The 1990s are likely to see the development of computer-based traffic management systems that incorporate in-car driver information and route guidance subsystems in the control loop. These systems would take O/D information provided by vehicles equipped with externally linked route guidance equipment, predict the expected travel paths of those vehicles, add the existing congestion and traffic volume information, and select the appropriate traffic signal control parameters before the traveling vehicles reached controlled intersections. Such systems will probably make use of on-line traffic models similar to those built into SCOOT. However, many theoretical and practical problems must be solved before road users will benefit from this generation of traffic management and control systems.

Work currently under way in California (PATH), Europe (PROMETHEUS), and Japan (ATSIC) will provide the impetus for these systems. Again, the FAME project needs to help promote and test the technologies developed in these programs. The cost of system development is, however, outside of the scope of the WSDOT research effort, unless WSDOT combines its effort with that of other agencies..

NEW THEORIES OF TRAFFIC CONTROL

The development of theories that support the computer models used by traffic engineers has slowed somewhat in recent years. The current emphasis is on the application of existing theories and on the refinement of the computation logic and data management aspects of the models. The addition of expert system capabilities on top of existing models (see above) could offer the most significant opportunities to improve the functionality of computer models.

With the exception of the expert systems area, the development of theories for describing and optimizing the traffic control appears to be in a fairly mature stage. Federally funded activities in this area have diminished somewhat in the past few years as the emphasis has shifted more to refinement and maintenance of existing models. Some work is progressing on the application of optimal control theory to oversaturated signal systems, and this theory may eventually find its way into operational models. Minimization of energy consumption due to stops and delay at traffic signals may be expected to generate further theoretical development if energy problems reemerge as signficant national concerns. Energy consumption in highway lighting systems has also attracted some interest. A linear programing model has been developed to examine traffic volumes throughout an illuminated network and to maximize the exposure of traffic to highway lighting under energy constraints.

Some interesting work has recently been published on the use of catastrophe theory for automatically detecting traffic accidents. Preliminary research in this area has indicated the possibility that more effective automatic incident detection procedures will be available for use within on-line traffic control systems. Research can be expected to focus on the problems of accurately detecting vehicle flows and estimating traffic characteristics. This area of research, although peripheral to traffic modeling, is important because of its impact on on-line simulation and the potential of new technologies to make intelligent traffic systems possible.

The WSDOT needs to remain aware of this ongoing research as part of its general research effort and its technology transfer work. Work in these areas is still primarily basic research, and the practical application of these ideas is still a number of years away.

REFERENCES

1. Ahmed, Samir A.; Cook, Allen R. "Discrete Dynamic Models for Freeway Incident Detection Systems," Journal of Transportation Planning and Technology, 1982, Vol.7, pp. 231-242.

2. Ahmed, Samir A. "Urban Freeway Traffic Management," Journal of Transportation Engineering, Vol.112, No.4, July 1986. ASCE.

3. Alderson, Stephen R.; Stephanedes, Yorgos. "Transportation Corridor Strategies and Land Use," Journal of Transportation Engineering, Vol.112, No.1, January 1986, ASCE, pp. 15-28.

4. Allen, Brian L., Easa, Said, M. and Case, E.R. " Application of Freeway-Corridor Assignment and Control Model," TRR 682, 1978, pp. 76-84.

5. Allen, Brian L. "Simplified Approach to Modelling Freeway Operations and Control," TRR 533, 1975, pp. 88-96.

6. Baass, K.G. and Allard, B. "Description of a Combined Approach for Arterial Signal Coordination Considering Bandwidth and Delays," TRR 957, 1984, pp. 32-46.

7. Babcock , Philip IV S., Auslander, D.M., Tamizuka, M., and May, A.D. "Role of Adaptive Discretization in a Freeway Simulation Model," TRR 971, 1984, pp. 80-95.

8. Benevelli, David A., Radwan, A. Essam, Hurley, Jamie W., Jr.; "Evaluation of a Bus Preemption Strategy by Use of Computer Simulation," TRR 906, 1983, pp. 60-67.

9. Bikowitz, Edward W. and Ross, Scott P. "Evaluation and Improvement of Inductive Loop Traffic Detectors," TRR 1010, 1985, pp. 76-80.

10. Bretherton, R.D., Bowen, G.T., Wood, K. "The Use of SCOOT for Traffic Management," paper in the Second International Conference on Road Traffic Control, Institution of Electrical Engineers, April 1986.

11. Bohnke, Peter and Pfannerstill, Elmer. "A System for The Automatic Surveillance of Traffic Situation," ITE Journal, January 1986.

12. Bolland, J.D., Hall, M.D., Vliet, D.V. and Willumsen. "A Model for the Simulation of Traffic Management Schemes," paper presented in Seminar on "Traffic and Environmental Management," PTRC Annual Meeting, 1977.

13. Byrne, A.S., de Laski, A.B., Courage, K.G., Wallace, C.E. "Handbook of Computer Models for traffic Operations Analysis," U.S. Department of Transportation Federal Highway Administration, Technology sharing Report, 1982.

14 Chan, C.C., Catling, I. "Area Traffic Control Systems and Eletronic Road Pricing in Hong Kong," IEE, papaer in Proceedings of the Second International Conferrence on Road Traffic Control," 1986.

15. Chang, Edmand C-P. "Select Traffic Analysis Software Using Expert Systems," paper presented at the 66th Annual Meeting, TRB, Washington, D.C., January 1987, Texas Transportation Institute, The Texas A&M University System.

16. Chang, Edmond C.P., Messer, Carroll J. and Cohen, Stephen L. "Directional Weighting for Maximal Bandwidth Arterial signal Optimization Programs," TRB 1057, 1986.

17. Chen, Leon and May, Adolf D. "The Use of Vehicle detectors for Freeway Traffic Management, Volume I: Detector Errors And Diagnosis," University of California at Berkeley, Insititute of Transportation Study, 1986.

18. Chin, Shih-Miao. "Interactive Computer-Graphics User Interface for Traffic Simulation Models," paper in "Application of Traffic Simulation Models," TRB Special report 194, 1982, pp. 103-111.

19. Chin, Shih-Miao, Eiger, Amir. "Network Simulation Interactive Computer Graphics Program," TRB TRR 835, 1982, pp. 52-66.

20. Cilliers, M.P., May, A.D., Cooper, R. "Development and Application of a Freeway Priority-Lane Model," TRR 722, pp. 16-26, 1979.

21. Clelland, Alan. "A Systems Approach to Area Traffic Control," Traffic Engineering and Control, April 1984, pp. 171-175.

22. Clowes, David. "Traffic Control — The Way Forward," paper presented in seminar on "Traffic Operations and Management," PTRC Annual Meeting, 1985.

23. Cohen, Stephen L. and Liu, C.C. "The Bandwidth-Constrainted TRANSYT Signal-Optimization Program," TRB 1057, 1986.

24. Cohen, Stephen L. "Concurrent Use of MAXBAND and TRANSYT Signal Timing Programs for Arterial Signal Optimization," TRR 906, pp. 81-84, 1983.

25. Cohen, Stephen L. and Liu, Charles C. "The Bandwidth-Constrained TRANSYT Signal-Optimization Program," TRR 1057, PP.1-9, 1986.

26. Colyer, John. "SCOOT from Scratch — Experience in Worcester," paper presented in Seminar on "Traffic Operations and Management," PTRC Annual Meeting, 1985.

27. Courage, Kenneth G. and Wallace, Charles E. "Models for Design and Evaluation of Traffic Signal Timing," paper in "Application of Traffic Simulation Models," TRB special report 194, 1982, pp. 35-40.

28. Davila, M.C. and Lieberman, E.B. "Hybrid Macroscopic — Microscopic Traffic Simulation Model," TRB, TRR 772, 1980, pp. 15-18.

29. Easa, S.M. and Allen, B.L. "CORCON: Freeway CORridor Assignment and CONtrol Model," TRR 682, 1978, pp. 76-84.

30. Eiger, Amir and Chin, Shih-Miao. "Fisrt-Generation UTCS Simulation," TRR 906, pp. 57-60.

31. Fehon, K.J.; Moore, S.E. and Negus, B.J. "Validation of SCATSIM," IEE, papaer in The Second International Conferrence on Road Traffic Control, 1986, pp. 123-126.

32. Foraste, B. and Scemama, G. "Surveillance and Congested Traffic Control in Paris by Expert Systems," IEE, paper in Proceedings of the Second International Conference on Road Traffic Control, 1986, pp. 91-94.

33. Foulds, L.R. "TRANSYT Traffic Engineering Program Efficiency Improvement via Fibonacci Search," Transportation Research, Part A. Vol.20a, No.4, pp. 331-335, 1986.

34. Gartner, N.H. "Prescription for Demand-Responsive Urban Traffic Control," TRB 881, 1982.

35. Gartner, Nathan H. "Demand-Responsive Traffic Signal Control Research," Transportation Research, Part A, Vol. 19A, No. 5/6 pp. 369-373, 1985.

36. Gartner, Nathan H. "OPAC: A Demand-Responsive Strategy For Traffic Signal Control," TRR 906, 1983, pp. 75-81.

37. Gibson, David R.P. and Williams Llewellyn. "The Arterial Analysis Package," Public Roads, Vol.50, no.3, December 1986, pp. 91-96.

38. Hall, M.D., Vliet, D.V. and Willumsen. "SATURN — A Model for the Evaluation of Traffic Management Schemes," paper presented in Seminar on "Traffic and Environmental Management," PTRC Annual Meeting, 1979.

39. Hecht, James R. and May, Adolf D. "The Use of Vehicle Detectors for Freeway Traffic Management, Volume 2: The Calculation of Traffic Characteristics From Detector Data," University if California at Berkeley, Institute of Transportation Study, 1986.

40. Hsu, Y. S. and Munjal, P.K. "Freeway Digital Simulation Models," TRR 509, 1974, pp. 24-41.

41. Hunt, P.B. Robertson, D.I., Bretherton, R.D. and Winton, R.I. "The SCOOT on-line Traffic Signal Optimization Technique," Procs. Intern. Conf. on Road Traffic signalling, Institution of Electrical Engineers, London, UK 1982, pp. 59-62.

42. Hunt, P.B. Robertson, D.I., Bretherton, R.D. and Winton, R.I. "SCOOT — A Traffic Responsive Method of Coordinating Signals," Transport Road Research Laboratory, TRRL Laboratory Report 1014, 1981.

43. Jeffery, D.J. and Russam, K. "Information systems for Drivers," Journal of Transportation Planning and Technology, 1984, Vol.9, pp. 185-198.

44. JHK & Associates. "A Comparison of Methods for Evaluating Network Traffic Control Systems," Federal Highway Administration, Report No. FHWA/RD-80/040, 1980..

45. Kahng, S.J.; Jeng, C-Y, Campbell, J.F. and May, A.D. "Developing Segmentwide Traffic-Responsive Freeway Entry Control," TRR 957, 1984, pp. 5-13.

46. Kessmann, R.W. "The Overland Park Traffic Signal Control System — Blueprint for the Future?"; paper in Proceedings of International Conference on Road Traffic Signalling, 1982, pp. 75-81.

47. KLD Associates, Inc. "Integrated Simulation Models -- Phase II : TRAF User Guide," 1985.

48. KLD Associates, Inc. "Development and Testing of INTRAS, A Microscopic Freeway Simulation Model," Federal Highway Administration, 1980.

49. Kneebone, D.C. and Howie, D.J. "Micro-computer versus Traffic: Chipping Away at the Traffic Problem," Proceedings of the 12th Australian Road Research Board Conference, vol.12, Part 4, August 1984.

50. Lee, Clyde E. and Lee, Fong-Ping. "Simulation of Traffic Performance, Vehicle Emissions, and Fuel Consumption at Intersections: the TEXAS-II Model," TRR 971, pp. 133-140, 1984.

51. Lieberman, E.B.; Rathi, A.K.; King, G.F. and Schwartz, S.I. "Congestion-Based Control Scheme for Closely Spaced, High Traffic Density Networks," TRR 1057, pp. 49-57, 1986.

52. Lieberman, E.B. and Andrews, B.J. "TRAFLO: A New Tool to Evaluate Transportation System Management Strategies," TRR 772, 1980, pp. 9-15.

53. Lieberman, Edward B. "Review Comments on Traffic Control Systems Handbook," Federal Highway Administration, U.S. Department of Transportation, 1976, p. 16.

54. Lowrie, P.R. "The Sydney Co-ordinated Adaptive Traffic System — Principles, Methodology, Algorithms," Proc. Intl. Conf. Road Traffic Signalling, Institution of Electrical Engineers, London, U.K., 1982.

55. Luk, J.Y.K. "Two Traffic-Responsive Area Traffic Control Methods : SCAT and SCOOT," Traffic Engineering and Control, January 1984, pp. 14-22.

56. Luk, J.Y.K. "SCATS — application and field comparision with a transyt optimized fixed time system," Procedings of International Conference on Road Trffic Signalling, IEE, 1982, pp. 71-74.

57. Luk, J.Y.K., Lowrie, P.R., and Sims, A.G. "Using TRANSYT for Traffic Signal Optimization in Parramatta," Proceedings of the 11th Australian Road Research Board Conference, Vol. 11, Part 4, August 1982.

58. May, A.D. "Models for Freeway Corridor Analysis," Presented at Conference on the Application of Traffic Simulation Models, June 3-5, 1981, Williamsburg, Virginia.

59. McDonald, M. and Hounsell, N.B. "The Potential for Route Guidance in On-line Computer Controlled UTC Networks," paper presented in Seminar on "Traffic Operations and Management," PTRC Annual Meeting, 1985.

60. Messer, Carroll J., Fambro, Daniel B. and Richards, Stephen H. "Optimization of Pretimed Signal Diamond Interchanges," TRB, TRR 644, 1977, pp. 78-84.

61. Michalopoulos, Panos G. and Plum, Roger A. "Selection and Evaluation of Optimal Freeway Design by Computer Simulation," TRR 773, pp. 40-47, 1980.

62. Michalopoulos, Panos G. "Dynamic Freeway Simulation Program for Personal Computers," TRB, TRR 971, 1984, pp. 68-79.

63. Morales, Juan M. and Paniati, Jeffrey F. "Two-Lane Traffic Simulation -A Field Evaluation of ROADSIM," Public Roads, Vol.49, no.3, December 1985.

64. Negus, B.J. and Moore, S.E. "The Benefits of SCRAM: The Maroondah Highway Survey," Proceedings of the 12th Australian Road Research Board Conference, Vol.12, Part 4, August 1984.

65. Nemeth, Zoltan A. and Mekemson, James R. "Comparision of SOAP and NETSIM: Pretimed and Actuated Signal Controls," TRR 905, pp. 84-95, 1983.

66. Oaks, R.L. "Role of Computer Graphics in Realtime Traffic Control System," Public Roads, Vol.45, No.2, pp. 62-67, September 1981.

67. Payne, H.J., Brown, D. and Todd, J. "Demand Responsive Strategies for Interconnected Freeway Ramp control System, Volume 1. Metering Strategies," FHWA, 1985.

68. Payne, H.J., Brown, D. and Todd, J. "Demand Responsive Strategies for Interconnected Freeway Ramp control System, Volume 2. Program Documentation," FHWA, 1985.

69. Payne, H.J., Brown, D. and Todd, J. "Demand Responsive Strategies for Interconnected Freeway Ramp control System, Volume 3. User's Guide," FHWA, 1985.

70. Payne, Harold J. "FREFLO: A Macroscopic Simulation Model of Freeway Traffic," TRR 722, pp. 68-77, 1979.

71. Powell, R.J. "SCOOT in Southampton," paper presented in Seminar on "Traffic Operations and Management," PTRC Annual Meeting, 1985.

72. Rach, Leonard, Lam, Joseph K., Kaufman, David C., and Richardson, David B. "Evaluation of Off-line Traffic-signal Optimization Techniques," TRR 538, 1975, pp. 48-58.

73. Rioux, Thomas W. and Lee, Clyde E. "Microscopic Traffic Simulation Package for Isolated Intersections," TRR 644, 1977, pp. 45-51.

74. Robertson, Dennis I. "Research on the TRANSYT and SCOOT Methods of Signal Coordination," ITE Journal, January 1986.

75. Robertson, D.I. "Research on Strategies for Traffic Control and Management," paper in ARRB Proceedings, Vol. 11, Part 1, 1982.

76. Robertson, D.I. "Traffic Models and Optimum Strategies of Control — A Review," Proceedings of the International Symposium on Traffic Control System, 1979.

77. Robertson, G.D. "Dual Strategy U.T.C. — Vehicle Actuation Under Computer Control and SCOOT," paper in the Second International Conference on Road Traffic Control, Institution of Electrical Engineers, April 1986.

78. Rogness, Ramey O. "Possible PASSER II Enhancements," TRB 881, 1982.

79. Ross, Paul. "Review of Road Traffic Network Simulation Models," TRR 644, 1977, pp. 36-44.

80. Ross, Paul. "The Future of Traffic Simulation," Public Roads, Vol.45, No.2, pp. 75-79, September 1981.

81. Ross, Paul. "Possible Futures for Traffic Simulation," paper presented in "Application of Traffic Simulation Models," TRB Special Report 194, March 1982.

82. Santiago, Alberto J. "ITDS: A Data Base Driven Interface to Traffic Models Using a Microcomputer," Public Roads, Vol.49, No.4, March 1986, pp. 122-126.

83. Schneider, Jerry B., Combs, D.M. and Folsom, T.C. "NETGRAF: A Computer Graphics Aid to the Operation and Interpretation of NETSIM, A Traffic Simulation Model: Part I. Overview and Experimental Results," U.S. Department of Transportation, 1980.

84. Schneider, Jerry B., Jette, C.L. and Lewis, B. T. "A Graphic Freeway Simulation Model FREGRAF," Computer Graphics World, Vol.3, No.5, 1980.

85. Sibley, Scott W. "NETSIM for Microcomputer," Public Roads, Vol.49, No.2, September 1985.

86. Sims, A.G. and Finlay, A.B. "SCATS. Splits and Offsets Simplified (S.O.S.)," Proceedings of the 12th Australian Road Research Board Conference, Vol. 12, Part 4, August 1984.

87. Sims, A.G. and Dobinson, K.W. "SCAT, The Sydney Co-ordianted Adaptive Traffic System: Philosophy and Benefits," paper in Proceedings of the International Symposium on Traffic Control Systems, 1979, University of California, Berkeley.

88. Skabardonis, Alexander. "Microcomputer Applications in Traffic Engineering," Journal of Transportation Engineering, Vol.112, No.1, January 1986, ASCE, pp. 1-14.

89. Stockfisch, Charles R. "The UTCS Experience," Public Roads, Vol.48, No.1, June 1984.

90. Summer, R., et al. "Freeway Management Handbook," Department of Transportation, Federal Highway Administration, May 1983 (four-volume set).

91. Taylor, Michael A. and Hatjiandreou, Michael. "Traffic Modelling in the Design Office : A Microcomputer-Based Methodology," Paper presented in Seminar on "Traffic Operations and Management" at the PTRC 13th Summer Annual Meeting, 1985.

92. U.S. Department of Transportation. Traffic Control System Handbook, Federal Highway Administration, FHWA-IP-85-11, Revised, April 1985.

93. Vincent, R.A., Mitchell, A.I., and Robertson, D.I. "User Guide to TRANSYT Version 8," Transport and Road Research Laboratory, TRRL Laboratory Report 888, 1980.

94. Walmsley, J. "The practical implementation of SCOOT traffic Control Systems," paper in IEE International Conference on Road Traffic Signalling, 1982, UK, pp. 197-201.

95. Wicks, D.A. and Lieberman, E.B. "Development and Testing of INTRAS, A Microscopic Freeway Simulation Model: Volume 1, Progress Design, Parameter Calibration, and Freeway Dynamics Component Development," FHWA 1977.

96. Wicks, D.A. and Lieberman, E.B. "Development and Testing of INTRAS, A Microscopic Freeway Simulation Model: Volume 2, User.s Manual," FHWA 1977.

97. Wicks, D.A. and Lieberman, E.B. "Development and Testing of INTRAS, A Microscopic Freeway Simulation Model: Volume 3, Validation and Applications," FHWA 1977.

98. Wilbur, Toni. "A Survey of TRANSYT-7F Applications," Traffic Engineering and Control, October 1985, pp. 498-501.

99. Yagar, Sam. "CORQ — A Model for Predicting Flows and Queues in a Road Corridor," TRR 533, 1975, pp. 77-87.

100. Yagar, Sam. "Measures of the Sensitivity and Effectiveness of the CORQ Traffic Model," TRR 562, 1976, pp. 38-48.

101. Yagoda, H.N. Rubenstein, A., Wisepart, I.S. and Herstein, H.I.; "FORCAST On-Line: Making UTCS Easier to Use," IEE Proceddings of International Conference on Road Traffic Signalling, 1982, pp. 97-102.

102. Yedlin, M. and Lieberman, E.B. "Development of a TRANSYT-Based Traffic Simulation Model," TRR 772, pp. 6-15, 1980.