ARIZONA DEPARTMENT OF TRANSPORTATION

REPORT NUMBER: AZ-SP-9701

RHODES - ITMS

Prepared by:
Larry Head
Pitu Mirchandani
Systems and Industrial Engineering Department
The University of Arizona
Tucson, Arizona 85721

April 1997

Prepared for:
Arizona Department of Transportation
206 South 17th Avenue
Phoenix, Arizona 85007
in cooperation with
U.S. Department of Transportation
Federal Highway Administration

Technical Report Documentation Page

1. Report No. AZ-SP-9701	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle RHODES-ITMS		5. Report Date April 1997
		6. Performing Organization Code
7. Author Larry Head and Pitu Mirchandani		8. Performing Organization Report No.
9. Performing Organization Name and Address Systems and Industrial Engineering Department, The University of Arizona, Tucson, AZ 85721		10. Work Unit No.
		11. Contract or Grant No. STP-900-0(088) K93 1973 TRN
12. Sponsoring Agency Name and Address **ARIZONA DEPARTMENT OF TRANSPORTATION 206 S. 17TH AVENUE PHOENIX, ARIZONA 85007** Project Manager: Sarath Joshua		13. Type of Report & Period Covered Final Report 12/93 - 11/96
		14. Sponsoring Agency Code

15. Supplementary Notes
Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

This report documents the work performed and the results obtained on the RHODES-Integrated Traffic Management System (ITMS) Project. The project addressed the design and development of a real-time traffic adaptive control system for Freeway-Arterial Diamond Interchanges using the concepts underlying the RHODES traffic-adaptive signal control system. The traffic "controls" at a diamond interchange are the two sets of traffic signals located at the arterials, on both sides of the freeway, and the ramp meters at the on-ramps to the freeway. The RHODES-ITMS strategy considers, in real-time, the vehicle arrivals and the queues at the intersections and the on-ramps to optimally set these signals, also in real time, to decrease the overall delay of all the vehicles which use the arterials, the frontage roads parallel to the freeway (if they exist) and the ramps at the interchange. RHODES-ITMS prediction/optimization algorithms (1) predict the arrivals and queues of individual vehicles at the arterial approaches on both sides of the freeway, as well as arrivals from the off-ramps and the departures and queues at the on-ramps, and (2) determine the optimal phasing of the signals at the two intersections on either side of the freeway.

To test the algorithms and the RHODES-ITMS strategy, a CORSIM-based simulation model was used. The simulation tests showed that the average vehicle delay is significantly lower for RHODES-ITMS than for the current fully actuated controls; average delays decrease between 25% to 50%. Also, the variance of these delays is drastically reduced. In addition, the simulation tests appear to indicate that RHODES-ITMS probably delays the onset of queue spillback and oversaturated conditions when the traffic volume is very high.

17. Key Words		18. Distribution Statement Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161		23. Registrant's Seal
19. Security Classification Unclassified	20. Security Classification Unclassified	21. No. of Pages 60	22. Price	

METRIC (SI*) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	2.54	centimeters	cm
ft	feet	0.3048	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	6.452	centimeters squared	cm^2
ft^2	square feet	0.0929	meters squared	m^2
yd^2	square yards	0.836	meters squared	m^2
mi^2	square miles	2.59	kilometers squared	km^2
ac	acres	0.395	hectares	ha
		MASS (weight)		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		VOLUME		
fl oz	fluid ounces	29.57	millimeters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.0328	meters cubed	m^3
yd^3	cubic yards	0.765	meters cubed	m^3

Note: Volumes greater than 1000 L shall be shown in m^3.

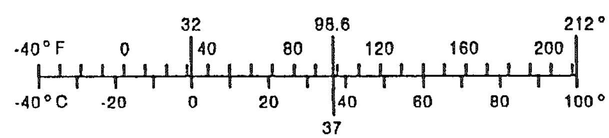
Symbol	When You Know	Multiply By	To Find	Symbol
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

These factors conform to the requirement of FHWA Order 5190.1A

*SI is the symbol for the International System of Measurements

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
yd	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	millimeters squared	0.0016	square inches	in^2
m^2	meters squared	10.764	square feet	ft^2
yd^2	kilometers squared	0.39	square miles	mi^2
ha	hectares (10,000 m^2)	2.53	acres	ac
		MASS (weight)		
g	grams	0.0353	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams (1000 kg)	1.103	short tons	T
		VOLUME		
mL	millimeters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	meters cubed	35.315	cubic feet	ft^3
m^3	meters cubed	1.308	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



PREFACE

This report documents the work performed on the *RHODES-ITMS* Project. This research effort was funded by the Arizona Department of Transportation (ADOT) and the Maricopa Association of Governments (MAG). Essentially, the scope of the Project was to develop a method to optimally control, in real time, the interchange traffic signal operations for the traffic that passes through the two intersections of a freeway-arterial diamond interchange (this excludes the freeway traffic going over, or under, the interchange). The method used was based on extensions of the concepts developed for the surface street network in the previous *RHODES* Project funded by ADOT and the Pima Association of Governments.

This report was written by the principal investigator, **Pitu B. Mirchandani**, and co-principal investigator, **Larry Head,** both of the Systems and Industrial Engineering (SIE) Department at the University of Arizona. It is based on the compilation of research efforts and results of various individuals who have been involved in the *RHODES-ITMS* Project. In particular, the efforts of the following individuals are acknowledged:

Steven Shelby	Graduate Assistant, SIE Department
Douglas Gettman	Graduate Student, SIE Department
Dennis Sheppard	Assistant Traffic Engineer, City of Tucson
Michael Whalen	Past Graduate Assistant, SIE Department (Currently with Gardner-Rowe Systems Inc.)

In addition, the principal investigators wish to acknowledge their appreciation to the Project's **Technical Advisory Committee** (TAC) whose continual active participation, technical input and support resulted in the *RHODES-ITMS* results being even more relevant to traffic engineering and control. The following individuals served on the TAC at various times:

Sarath Joshua	RHODE-ITMS Project Manager, Arizona Transportation Research Center (ATRC), ADOT
Jim Decker	Traffic Operations, City of Tempe
Dan Powell	ADOT District 1
Tom Parlante	ADOT Traffic Engineering
Larry A. Scofield	ATRC, ADOT
Tim Wolfe	ADOT Tech. Group
Alan Hansen	Federal Highway Administration
Cathy Arthur	Maricopa Association of Governments
Roger Herzog	Maricopa Association of Governments
Paul Ward	Maricopa Association of Governments
Tammy Flaitz	Maricopa Association of Governments
Thomas Buick	Maricopa County Transportation and Development Agency
Don Wiltshire	Maricopa County Transportation and Development Agency
Philip Lindsay	City of Phoenix
Doug Dykhouse	City of Phoenix
Monica Beeman	City of Phoenix

RHODES - ITMS: FINAL REPORT

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION AND BACKGROUND

1.1 Project Scope and Objectives

The RHODES-Integrated Traffic Management System (ITMS) Project addresses the design and development of a real-time traffic adaptive control system for Freeway-Arterial Diamond Interchanges using the concepts underlying the RHODES traffic-adaptive signal control system. The current approaches to control the traffic on the arterial of the interchange are (1) fixed time, perhaps based on time-of-day traffic conditions, and (2) actuated (or semi-actuated) where loop detectors detect traffic on specific lanes and/or movements and based on some designed logic provide pre-specified phases, phase skips, phase extensions, force-offs and gap-outs to allow for the movement of the detected traffic. The major deficiency for such types of strategies is that there is no way for the control system to respond to anticipated actual arrivals - by varying phase durations and/or using more appropriate cycle times and phase sequencing - even though detectors at the off-ramps and upstream intersections may have identified unusual traffic conditions (either unusually large volumes or very small volumes, due to, for example, events and incidents). Also, unusually large queues detected at the on-ramps are not considered in phase durations; vehicles may be directed on to the queued on-ramps which results in no apparent effect on their delays but instead induces queue spill-backs and possibly increases delays for other traffic.

The traffic "controls" at a Freeway-Arterial Diamond Interchange (FADI) are the two sets of traffic signals located at the arterials, on both sides of the freeway, and the ramp meters at the on-ramps to the freeway. To truly manage all the traffic at the FADI, we need to do both, set the phase durations of the intersection traffic signals and control the ramp-metering rates, taking into account local traffic objectives as well as network-wide objectives. However, the scope of the project was only to control the traffic signals, with ramp-metering rate given externally. There were two reasons for not adjusting the ramp-metering rates: (1) current practice is that the state traffic agency (e.g., ADOT in Arizona) sets these rates directly, with consideration of freeway flow management objectives, and (2) ideally, ramp-metering rates should take into account not only local flows at the FADI but also wide-area network flows, which would substantially increase the scope of the project. (A subsequent project titled *RHODES-ITMS Corridor Control*, currently underway, is addressing the control of ramp-metering rates from wide-area traffic considerations.) Nevertheless, as will be described in the report, it is necessary to have information about the queues at the on-ramps in order to decrease the overall delay of all the vehicles which use the arterials, the frontage roads parallel to the freeway (if they exist) and the ramps at the interchange. Thus, in summary, the scope was to develop a method to optimize only the interchange traffic signal operations, for only the traffic which passes through the two intersections of the interchange (and not the freeway traffic going over, or under, the interchange). Figure 1 gives a schematic diagram of the area from which signal and traffic information (from detectors) is collected, and the area (shaded) under real-time adaptive control using the RHODES-ITMS System.

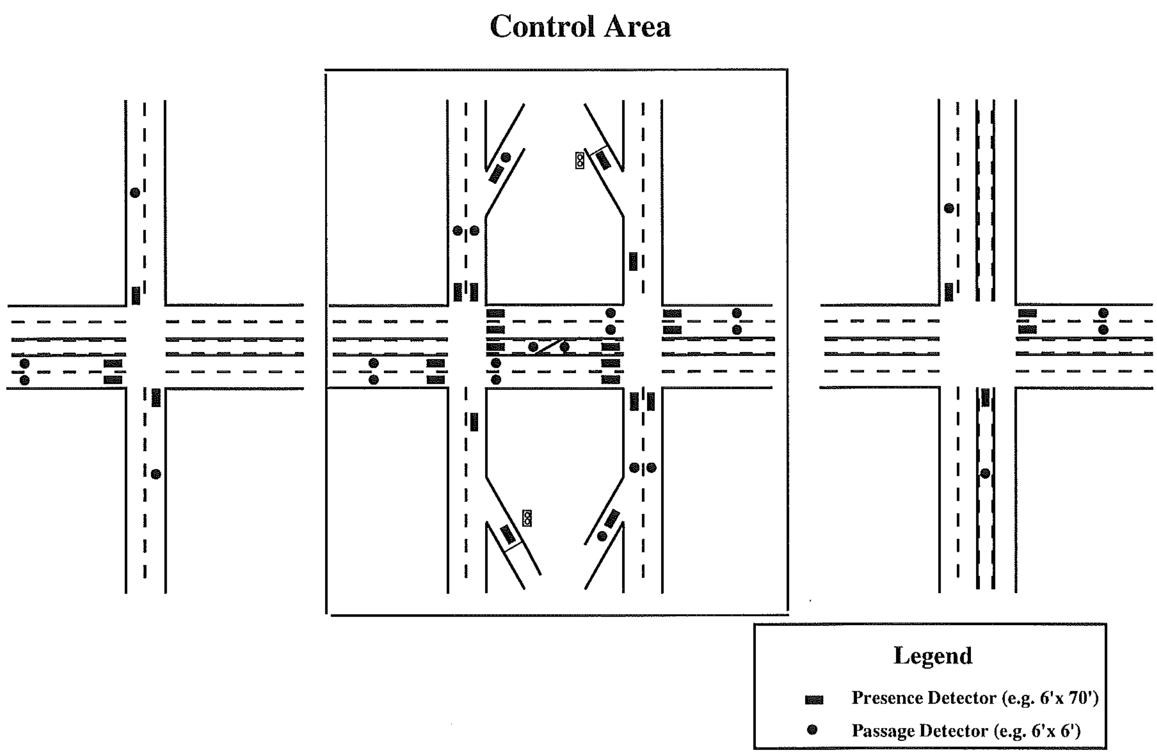


Figure 1 The interchange: Control area and detectors utilized.

To briefly review the RHODES concept, the RHODES architecture for surface streets is depicted in Figure 2. At the highest level of RHODES is the "dynamic *network loading*" model that captures the slow-varying characteristics of traffic. These characteristics pertain to the network geometry (available routes including road closures, construction, etc.) and the typical route selection of travelers. Based on the slow-varying characteristics of the network traffic loads, estimates of the load on each particular link, in terms of <u>vehicles per hour</u>, can be calculated. These load estimates then allow RHODES to allocate "green time" for each different demand pattern and each phase (North-South

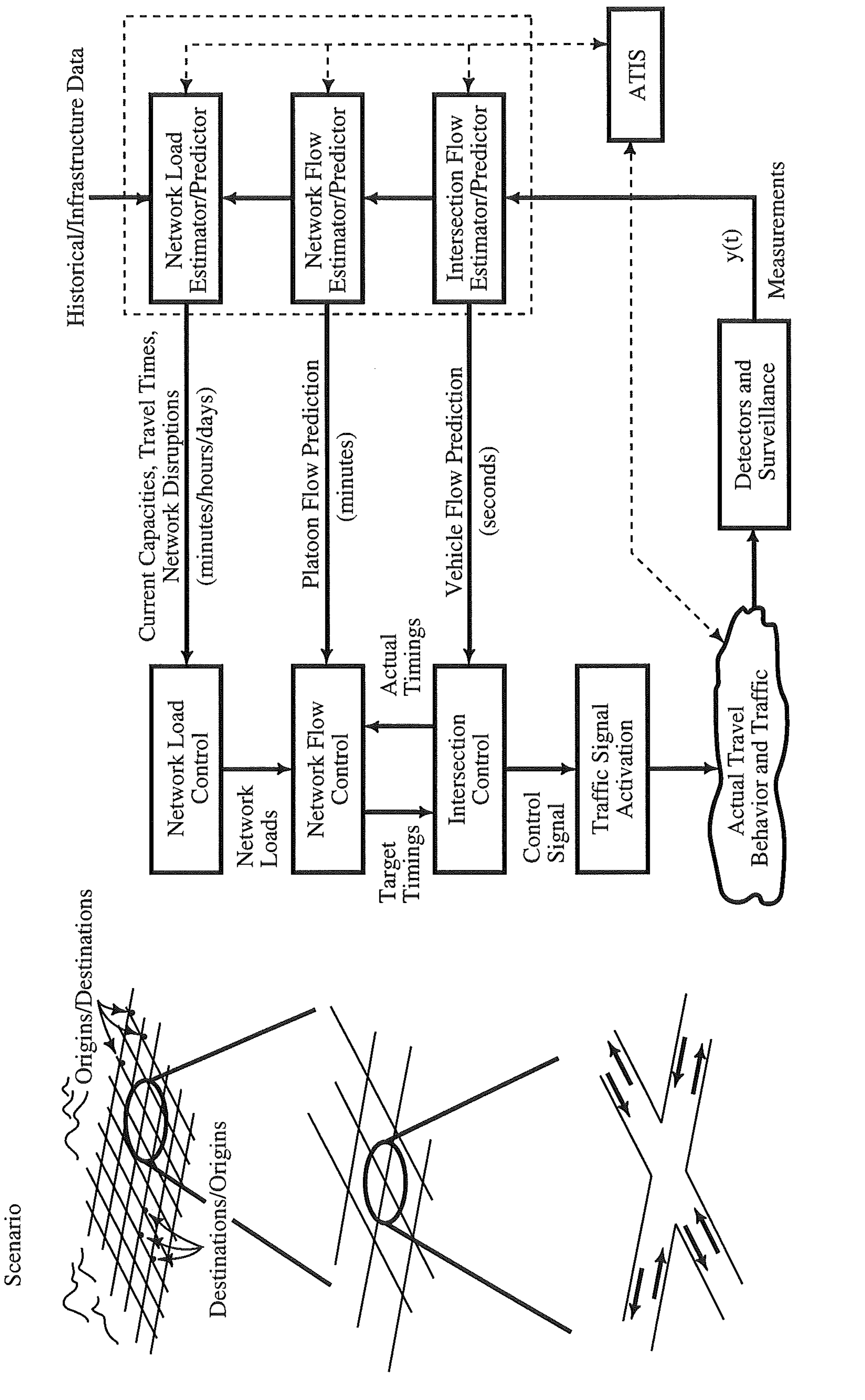


Figure 2 The *RHODES* Architecture

through movement, North-South left turn, East-West left turn, and so on). These decisions are made at the middle level of the hierarchy, referred to as *"network flow control"*. Traffic flow characteristics at this level are measured in terms of platoons of vehicles and their speeds. Given the approximate green times, the *"intersection control"* at the third level selects the appropriate phase change epochs based on observed and predicted arrivals of individual vehicles at each intersection. The RHODES architecture and its software implementation is modular; it allows the accommodation of new modeling methodologies and new technologies as they are developed.

A significant difference between RHODES and other "real-time" traffic control systems is that RHODES is being designed to accommodate real-time measurements of traffic and to become an integral component of *Intelligent Transportation System* (ITS). For example, integration of *Advanced Traveler Information Services* (ATIS) within ITS will result in (1) improved prediction and estimation of network loads, (2) allowing the traffic management system to provide drivers with real-time information about traffic conditions, and (3) advising the travelers of alternate routes. Priority and accommodation of public and private transit, emergency vehicles, and commercial vehicles, can be easily integrated into the decision-making structure of RHODES.

At the highest *network loading* level of the hierarchy we envision the decision time horizons to be 15 minutes to an hour. This model allows for integration of historical data (a priori information), observed traffic flows (posterior information) and potential ATIS information about ITS suggested routes and traffic conditions (congestion, accidents and other network events) to allow prediction of near future loads and hence exercise real-time proactive traffic control. The next level of the hierarchy utilizes the predicted and estimated network loads to control traffic on a network-wide basis. At this level the *network flow controller* integrates the network load information with observations of actual volumes/platoons and flow profiles to select major phases and their approximate durations. These timing decisions are passed to the *intersection controller* where decisions to shorten or extend the current phase, and the selection of target timings for the next, say, 30 - 60 seconds, are made based on actual observations of the current traffic arrival pattern and estimated queues at each intersection. The lowest level of the hierarchy, referred to as *traffic signal actuation*, is responsible for implementation of the intersection controller decision on the signal control hardware.

Although the RHODES-ITMS Project included some effort at the second level of the hierarchy - network flow control - to anticipate the logic for the eventual integration of freeway and arterial control (see Section 2.3), most of the Project's effort concentrated on the third level - here referred to as *intersection/interchange control* (recall that the diamond interchange includes intersections on both sides of the freeway). The basic goal of the project was to predict arrivals and queues of individual vehicles at the arterial approaches on both sides of the freeway, as well as arrivals from the off-ramps and the departures and queues at the on-ramps, and based on these predictions and a given criterion of performance to determine the optimal phasing of the signals at the two intersections on either side of the freeway. In Sections 3 and 4, the prediction and the optimization algorithms are developed for this purpose.

To test the algorithms, it was necessary to first do so in the "laboratory". In this regard the Project Team developed a TRAF-NETSIM-based* , and later a CORSIM-based* , simulation model which was used as a platform to test our RHODES-ITMS strategy at the intersection/interchange level. Issues related to the simulation modeling and the simulation-based testing of the algorithms, are discussed in Section 2.

1.2 History of the Project

In June 1991, the Arizona Department of Transportation, working closely with the City of Tucson and the Pima Association of Governments (PAG), supported the initial R&D efforts on the development of the RHODES surface street traffic control system within the Department of Systems and Industrial Engineering at the University of Arizona. Based on this work, the RHODES Research Team submitted a proposal entitled "*A Real-Time Traffic Adaptive Signal Control System*" to Federal Highways Administration (FHWA) on the design of a prototype real-time traffic-adaptive signal control system based on the RHODES architecture. The RHODES Team led a strong consortium, that included JHK & Associates, SRI International Inc. , The Analytical Sciences Corporation, Rensselaer Polytechnic Institute, and Hughes Aircraft Company, and submitted a consortium proposal to FHWA in January 1992. The proposal was not selected for funding; the contract was awarded to Farradyne Systems in June 1992.

---

* TRAF-NETSIM and CORSIM are software packages for modeling and simulting traffic on a network. Their development have been supported by FHWA.

However, in 1994, FHWA sent out another call for proposals for the development of alternative strategies for real-time traffic adaptive control, and the RHODES Team were one of five awarded a contract. The scope of that project was to develop a working prototype of the RHODES strategy, implementing only the last two levels of the hierarchy - intersection and network flow control, which was to be laboratory tested by a third party contractor. This FHWA contract was initiated in June 1994, with a completion date of October 1995. The contract will be completed sometime in 1996 (a few no-cost extensions were granted), and the prototype is currently being evaluated by Kaman Sciences (which won the third-party evaluation contract).

In parallel to the FHWA proposal development effort, the RHODES Team discussed with ADOT the possibility of developing an improved traffic control strategy at the interchange-arterial interface and eventually field testing some of the concepts at a location in Maricopa County. After several meetings with traffic engineers and technical managers within ADOT, Maricopa Association of Governments (MAG), Phoenix, and Tempe, the research team proposed the development of a real-time traffic adaptive control strategy for later field testing at an interchange along the I-17 Corridor. A draft proposal was developed and submitted to ADOT in April 1992. This was later revised to accommodate resource constraints, and resubmitted in September 1992 entitled "Real-Time Traffic Adaptive Control for Integrated Traffic Management of the I-17 Corridor" and referred to as the "RHODES-ITMS Project" for short. The proposal received favorable reviews from ADOT and MAG, and the RHODES-ITMS Project was approved for funding in December 1993. The funding was jointly provided by ADOT (from SP&R funds) and MAG, and the Project was administered by Arizona Transportation Research Center in ADOT. This report addresses the activities and the findings of that Project.

1.3 Project Tasks

RHODES-ITMS Project consisted of the following tasks:

- Task A: Conduct Literature Review
- Task B: Develop Simulation Models
- Task C: Develop Intersection/Interchange Control Models
- Task D: Develop Flow Control Models
- Task E: Integrate Hierarchical Control Models
- Task F: Laboratory Test
- Task G: ITS R&D Liaison

Although the principal researchers on the Project had a good knowledge and understanding of the literature on traffic models and traffic-adaptive control approaches for surface streets, and some know knowledge of freeway control, the Project Team felt that a thorough literature review was essential to understand the current state of the art and the state of the practice of both freeway traffic models and control algorithms, as well as current attempts to integrate surface street control with freeway control. Task A was for this purpose.

Task B consisted of evaluating current simulation models (for surface street traffic, for freeway traffic and for both freeway and surface street traffic) and specifying the requirements for a simulation model for demonstrating, testing and evaluating real-time control for diamond interchanges. It was decided that modification of the TRAF-NETSIM model would provide a suitable simulation environment. However, as we will discuss in the next section, development of a simulation platform posed major hurdles in the conduct of this Project.

As an exercise on the application of a freeway/surface street simulation model, and as a delivery requirement suggested by ADOT, a case study on the evaluation of strategies for high-occupancy-vehicle (HOV) freeway traffic management was conducted. The focus of the study was to evaluate HOV strategies such as "add a new HOV lane", "convert a general purpose lane to HOV", "add a general purpose lane", etc. A paper (Sheppard et al., 1996, included in Appendix A) on this case study was presented at the 1996 Transportation Research Board Meeting and recently published by TRB.

Task C focused on the investigation and development of algorithms for intersection/interchange traffic adaptive control. There were two challenges here: (1) to develop methods to predict arrivals and queues at the arterial approaches and the on-ramps, and (2) to develop an algorithm to determine the "optimal" phasing based on the above predictions. Section 3 describes our approach to the prediction problems and Section 4 to the intersection/interchange control problem.

Task D was a small effort to investigate network coordination from the perspective of interchange control. This was a complementary effort to the FHWA contract for the development of the RHODES prototype where a network flow control model was being developed [Dell'Olmo and Mirchandani, 1995]. Although no new network level algorithm was developed on this project, Task D efforts allowed the Team to provide the appropriate interface on the intersection/interchange algorithm to allow coordination between the freeway diamond signals and the network arterial signals.

Task E was to integrate the intersection/interchange control algorithms with the estimation/prediction algorithms, as well as integrate them with the simulation models for laboratory testing. Although the initial scope included the integration of a network flow component, this was not included in the final scope due to the reallocation of resources to address the unanticipated simulation and testing barriers (see Section 2).

In Task F, the RHODES-ITMS system was tested on the simulation platform developed in Task B and integrated in Task E. The results of the evaluation are given in Section 6.

In addition to these tasks, another task, Task G, was included in the RHODES-ITMS Project referred to as "ITS R&D Liaison". The principal researchers on the Project felt that it was important to interface with various transportation research forums (e.g. Transportation Research Board, ITS-America and ITS-Arizona meetings) to report on the Project's findings and develop proposals for further work on real-time traffic adaptive control systems. Hence, some effort was budgeted for such a task.

1.4 Project Oversight

Project oversight was provided by a Technical Advisory Committee (TAC) comprising of representatives from key agencies. The project was administered by the Arizona Transportation Research Center of ADOT. The members of the TAC were:

Name	Agency
Jim Decker	City of Tempe
Monica Beeman	City of Phoenix
Paul Ward	MAG
Alan Hansen	FHWA
Sarath Joshua (Project Manager)	ATRC/ADOT
Tom Parlante	ADOT Traffic Engr.
Dan Powell	ADOT District 1
Tim Wolfe	ADOT Tech. Group
Don Wiltshire	Maricopa County

At earlier stages of the project there were some other members on TAC, specifically Larry Scofield from ATRC/ADT, Tammy Flaitz, Cathy Arthur and Roger Herzog from MAG, Tom Buick from Maricopa County, and Philip Lindsay and Doug Dykhouse from the City of Phoenix, who had to be replaced due to their other commitments.

2. SIMULATION ISSUES, MODELING AND TESTING PLATFORM

In this section we discuss the simulation model used for studying and evaluating the effectiveness of traffic control algorithms. After a brief discussion of the issues related to simulation modeling and evaluation, we present our approach to develop a simulation model for testing real-time traffic-adaptive traffic control, which is based on the modification of the TRAF-NETSIM/CORSIM model. The modified model was validated by implementing external fixed-time (and external semi-actuated) signal control logic and comparing the performance of the external control logic with the corresponding logic that is internal to TRAF-NETSIM/CORSIM. Having validated the simulation model, the real-time traffic-adaptive intersection control algorithm, ICOP, as described in Section 4, was interfaced with the simulation model and evaluated.

2.1 Simulation Modeling Requirements and TRAF-NETSIM/CORSIM

It is clear that any type of traffic control algorithm needs to be tested in the "laboratory" before it is implemented and evaluated in the field. The most appropriate method to do this "laboratory" testing is to (1) have a realistic simulation model of traffic flow at an interchange, (2) emulate the (loop) detection of the traffic flow, and (3) observe the resulting changes that would come about if the algorithm was implemented in place of the current control system. Adopted from a previous report [Head and Mirchandani, 1994], the functional requirements for simulation models for development, testing and evaluation of real-time traffic-adaptive signal control logic in this setting include:

- the ability to realistically simulate the arriving and departing vehicular traffic at an interchange;
- the ability to generate dynamic traffic conditions, including recurrent and non-recurrent congestion such as incidents and special events;
- the ability to obtain surveillance/detector output at required frequencies;
- the ability to implement decisions (for, example from RHODES) to control traffic signals in real-time; and
- the ability to compute various measures of effectiveness based on traffic characteristics (including those that are not necessarily observable, such as queue lengths).

The ability to represent dynamic recurrent and non-recurrent congestion, as well as other non-congested traffic conditions, is needed for measuring the algorithm's capability to respond to real-time traffic conditions.

Simulation models used for testing must provide the same surveillance and detection information as that available in the field. The frequency of surveillance and detector system output and the frequency of the signal control input will dictate the minimal resolution, and hence the responsiveness, of the signal control logic. The simulation model must be able to represent rates that will be achievable when the control logic is implemented for field testing.

It may be desirable for the signal control algorithms to optimize different measures of effectiveness (MOE), based on traffic conditions or dictated by the operating jurisdictions. Therefore, it is essential that the simulation model provide a wide variety of MOEs to evaluate the real-time traffic adaptive signal control algorithms.

The simulation model requirements from a development and testing perspective differ from the requirements for performance evaluation. Clearly, the most important requirement of a simulation model is that it accurately represent the dynamics of traffic flow and its response to dynamic signal control. This requirement dictates that the simulation model chosen for development and testing not be based on a macroscopic flow model that assumes constant cycle length and deterministic traffic flow characteristics. Rather, the model should include microscopic flow characteristics, such as car-following, and include an ability to simulate real-time traffic controls (not necessarily constant cycle lengths) and attendant vehicle response to actual traffic signals.

During the development and testing phase it is essential to have access to both traffic and signal control variables so that detailed behavior can be studied. One may distinguish between traffic simulation information/data that is needed for validation and testing and that information/data which is available as traffic surveillance/detection data for the signal control algorithms. For example, for the purpose of testing a traffic model used in an optimization routine, it may be desirable to compare the traffic model's state-of-the-traffic measures, such as queue length, to the corresponding measures in the simulation model. This form of testing requires that the traffic simulation model provide accurate measurements of queue lengths despite the fact the existing traffic surveillance technology may not provide this information.

Another important consideration is the frequency at which required testing data is available. For example, the average queue length for a simulation period is insufficient for testing a routine that estimates real-time queue lengths. This information must be available as frequently as possible, at least as frequent as queue estimated are generated.

For the past four years we have been using the TRAF-NETSIM traffic simulation model for development and testing of traffic-adaptive signal control algorithms for surface streets. We have modified the TRAF-NETSIM model to (1) interface the simulated traffic surveillance equipment with the algorithm's database, and (2) interface the simulation model's traffic controllers with the algorithm's output. In addition, we have made information that is normally internal to the simulation model's database accessible to analyst; this allows the analyst to test and validate algorithms and software.

Also, during these four years FHWA has been actively improving the logic and implementation of TRAF-NETSIM. These improvements include a Windows-based implementation, integrating TRAF-NETSIM and TRAF-FRESIM into one module called CORSIM and developing an interface based on the Dynamic Link Library (DLL) in Windows. This makes the implementation of a real-time control strategy on CORSIM easier as well as improves the integrity of the simulation results.

The RHODES research team has worked closely with FHWA in implementing the RHODES algorithms on CORSIM and helping the CORSIM developers to make appropriate changes in their software to develop the DLL-interface. This cooperation has resulted an improved environment for simulation modeling and evaluation and a real-time control interface approach where both FHWA and the RHODES team to agree. Hence, at the later stages of this project the RHODES research team ported all its algorithmic software to CORSIM.

In testing RHODES' intersection control algorithm, the research team modified the simulation model to record the simulated signal states and compare them with the signal states downloaded by the algorithm. This ensured that the desired signal state was the signal state activated by the simulation model. The CORSIM surveillance logic was modified to allow the research team to record simulated traffic flow profiles on each link and compare them with the predicted traffic flow profiles that are generated by the traffic flow prediction models.

The interchange that we simulated in our test experiments consisted of the freeway, a cross arterial (which makes this a diamond interchange), four signalized intersections (two on each side of the freeway), two parallel frontage roads, a parallel arterial on either side of the freeway, and pairs of off-ramps and on-ramps (see Figure 1). To simulate actual operating characteristics of a diamond interchange, one such interchange was identified in the City of Phoenix and the associated traffic characteristics, volumes and signal configurations were obtained with the assistance of the City's traffic engineers. Thus our simulations and laboratory experiments were based on realistic data.

2.2 Interfaces with the Simulation Model

When the project team began the development of the simulation interface with TRAF-NETSIM, its objectives were (1) to minimize the modifications to the existing TRAF-NETSIM code and (2) to simulate the interface between the real-time traffic-adaptive signal control logic and the NEMA-type controllers that are currently used in Phoenix, Arizona. The actuated signal controller logic (software Q5 logic) was programmed, through the input data base, to have the desired set of phases with the desired minimum green intervals. Then detector information, contained in the TRAF-NETSIM internal data base, is first read, for the purposes of surveillance and, subsequently, it is either cleared or set to represent a CALL for the desired phase. When the signal state is updated, the CALL is processed thereby forcing the signal into the desired state. Figure 3 depicts the software implementation of this approach.

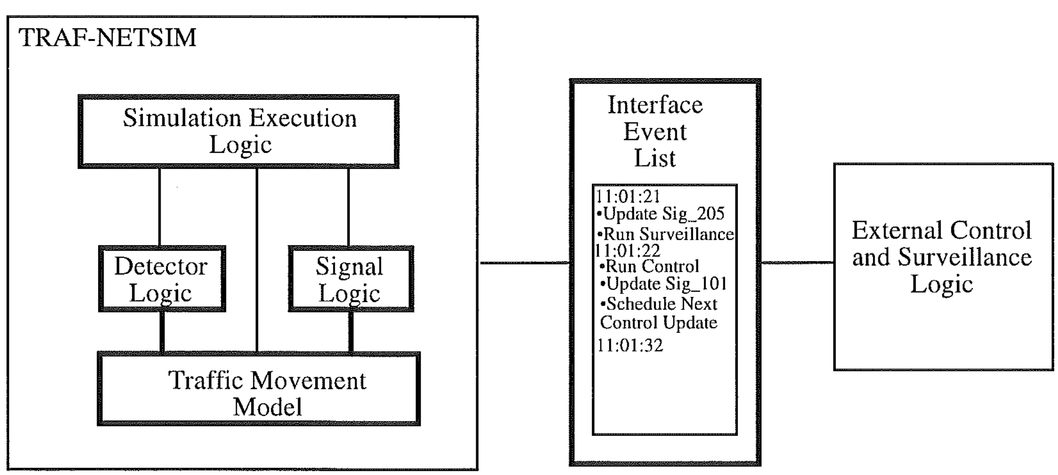


Figure 3 Software Interface for External Control and Surveillance Logic

However, with the availability of the DLL in CORSIM, it no longer becomes necessary to represent a CALL for a desired phase by forcing the actuated controller logic through detector counts. RHODES can directly set the DLL values which CORSIM then reads for desired phasing.

As it was done in the RHODES System (for surface streets), the interface implementation included the development of a library of interface functions that the external logic, either surveillance or control, can use to access database information, alter database information, place events on the event managers list, or remove events from the event manager list. This library facilitates the interface between the FORTRAN simulation model and the external logic that was developed in the C programming language.

An essential element in the development of an external signal control logic, especially real-time control, is the traffic surveillance system. In our effort, we have utilized the internal surveillance detector logic CORSIM for the placement and processing of detector events, but we utilize an external logic for processing the detector data. This approach has allowed us to estimate any traffic parameter that is desired, in addition to the standard count and occupancy values. The surveillance detector logic (SUBROUTINE DETECT) is used to generate events (pulses or signals) from each detector on a tenth-of-a-second basis. In particular, the surveillance detector logic determines the beginning and ending of detection events on tenth-of-a-second basis. These events are the externally translated into continuous binary signals. Each continuous signal represents either an occupied detector or an unoccupied detector. These signals are then processed into the proper parameter estimates and traffic measures for the associated signal control algorithms.

It is important to note that the model enhancements made were intended to support the research team's on-going activities in real-time traffic-adaptive signal control. The primary mission in these activities was not to develop a general platform for testing ITS components. The objective was to develop a platform for testing real-time traffic-adaptive signal control logic.

2.3 Complicating Factors in the Development of Simulation Models

Although the Project team felt that the simulation platform based on CORSIM was the most appropriate for its objective, it was not the ideal. This is because there were several complicating factors in adapting CORSIM for simulating a diamond interchange and for testing real-time control logic. These factors are discussed below.

Simulating Diamond Interchange Traffic

In the most recent off-the-shelf version of CORSIM, there are no "interchange controller" elements, that is where two intersection controllers are coupled to produce interchange phases. Therefore, to produce these interchange phases, the project team had to develop an external interchange controller (EIC) which downloaded concurrently the required phases for the corresponding two intersection signals for any desired interchange phase. In effect, this allowed us to take the appropriate detector signals, process them to provide predictions and "optimal" phasing (via ICOP), input these phase decisions to EIC, which in turn downloaded the corresponding phase signals to the intersection controllers. The graphical animation available with TRAF-NETSIM allowed us to view the traffic flow through the interchange thus defined, and indeed this appeared to validate the simulation model as far as the arterial flow was concerned.

However, there was still the problem of simulating the merge behavior at the on-ramps. Initially, a CORSIM "node" was defined for the ramp meters, with intersection controllers that only provided two phases "red" and "green". Unfortunately, the built in intersection logic for this node produce unrealistic queues on the freeway, since a priority is given to the on-ramp vehicles during the "green" phase. We then modified the on-ramp simulation element by including a dedicated link (lane) for the on-ramp traffic after the node (ramp meter) which, subsequently, experienced a "yield" behavior, by our inclusion of a "yield" function in the model, when it merged into freeway traffic some distance downstream. Although the animation of the freeway was still not ideal, the behavior of traffic through the interchange and the ramp meters was now very realistic; and this was all that ICOP attempted to optimize.

TRAF-NETSIM/CORSIM Support

As in the case of most software, especially those that are being continuously upgraded and for which new versions are being continuously developed, there exist software "bugs". In our case, during the process of simulation modeling and validation, we came across many such bugs. Initially, we found that the TRAF-NETSIM developers were quick to respond to our discoveries, sometimes showing us where our coding was in "error" and sometimes repairing the discovered bugs and providing us with a new version of TRAF-NETSIM. However, as our modeling got more complex, we were exercising parts of the codes that had hardly been tested and our responses to queries became much slower; sometimes the queries were not or could not be answered. Part of the reason was that we were conducting all our modeling and experiments on a Sun workstation and the contractors were not obliged to support the software on this platform. We had always found that the Sun environment was better for software development which could later be ported to another platform if a need arose.

Fortunately, our concurrent FHWA contract to develop a RHODES prototype for a surface street network required us to use a PC platform. Thus, in parallel to this RHODES-ITMS Project, we were developing simulation models and algorithms that were executable on a PC. Thus, in the middle of this Project we ported the simulation and the algorithmic development to a PC platform (Pentium 120 Mz, Windows NT 3.51). Note that our major reason to do this was that we were able to get better TRAF-NETSIM support for the PC platform. However, this change required the adaptation of the DLL-based CORSIM interface, which required considerable re-engineering effort.

Graphics /Animation Visualization

We had realized earlier in our RHODES' efforts that an essential preliminary validation of our simulation models was to get a visual animation of the traffic flow on the surface streets and the interchange. Early versions of TRAF-NETSIM could only provide animation for few seconds due to memory requirements. This is because the whole simulation run needs to be first executed, and the data required for the animation has to be stored. Subsequently, by an appropriate execution call, a graphical animation of the simulation run is presented. In this way, one could see if the traffic flow "looked right" for the scenario (and signals) just simulated. The Sun workstation allowed us to get animations of longer simulation runs, using a package referred to as GTRAF. However, in the middle of the Project period, this package was no longer supported and we were not able to fix the discovered "bugs" on the Sun. We proceeded with a somewhat "blind" development of the simulation model and the ICOP, only to later realize that there were huge queues being built up in the simulation runs, that were neither being serviced by the signal control decisions nor were they being considered in the MOE computations. This was partly due to the space limitation specified by TRAF-NETSIM, and partly due to the manner in which RHODES processed detector data for over-saturated queues.

In general, RHODES is not effective in over saturated conditions. However, it should not have a problem when occasionally a lane becomes over-saturated and it takes more that one cycle to clear the queue. Thus, the prediction algorithm must realize that the queue did not clear during a green phase and the delays of the remaining cars should be taken into account in the subsequent optimization of phase durations. However, when the remaining queue is past the detector then no arrivals are recorded and the logic makes the erroneous assumption that there were no arrivals in that lane. Then ICOP does not give the corresponding movement a green phase resulting in the queues becoming even longer. We did not see this because of the unavailability of the animation function. On the other hand, when queues become so large that generated vehicles cannot come into the study area, TRAF-NETSIM just keeps them on hold outside the study area but **does not include their delays in the computation of the performance measure.** Hence, in the algorithm for optimization of phase durations neither were the over-saturated queues considered, nor were their delays measured in the TRAF-NETSIM simulation. The reported measures showed better performance than fixed time control or the semi-actuated control (the current strategy at the simulated interchange); it was only after detailed (and tedious) examination of the simulation output did we realize the build-up of undetected queues.

Fortunately, we were able to get better visualizations on the PC platform and we were able to modify our queue prediction algorithms to make ICOP run appropriately. Further, ICOP performed much better than the current strategies (see Section 5). But unfortunately, this was done much later in the Project, delaying its deliverables and decreasing the scope of the Project.

Setting Signals with ICOP

In the first RHODES project, to set a desired phase at a given time, appropriate detectors calls were inputted to actuated signal controllers so that the desired signals for that phase were set. In other words, we disabled the link from the detector calls to the controllers, but instead linked the calls to the external RHODES logic, which after optimization inputted to appropriate detector calls (not the true calls) to the controllers to get the desired phase (see section 2.2; for more details see Head and Mirchandani, 1994).

However, in the RHODES-ITMS Project, this scheme did not work for one of the interchange phases (recall in our implementation an interchange phase was a set of coupled movements for two intersections). Due to lack of TRAF-NETSIM support and unavailability of graphic animation, we were not able to discover why one of the interchange phases could not be implemented in the simulation model. However, this did not become an issue when we ported the development to the PC-based CORSIM platform.

3. MODELS FOR PREDICTING VEHICLES ARRIVALS AND QUEUES

In the earlier RHODES Project Report [Head and Mirchandani, 1994] the importance of prediction of vehicle arrivals, turning probabilities and queues at an intersection, in order to compute phase timing that optimize a given measure of effectiveness (e.g. average delay), was discussed. To emphasize this importance, consider the intersection shown in Figure 4. This intersection has four approaches. (An interchange consists of two coupled intersections; later we will discuss how the predictions are done for an interchange.) Associated with each approach are several possible traffic movements: left turn, right turn and a through movement. For the purpose of signal timing, the right turn and through movements are generally considered as a single movement. Any non-conflicting combination of movements that can share the intersection at any one time can be assigned a signal phase that allows those movements protected use of the intersection. The traffic demand for a phase is determined by the approach volume (measured using a group of loop detectors on the approach to each intersection and in the left-turn pockets) and the turning probabilities associated with vehicle routes.

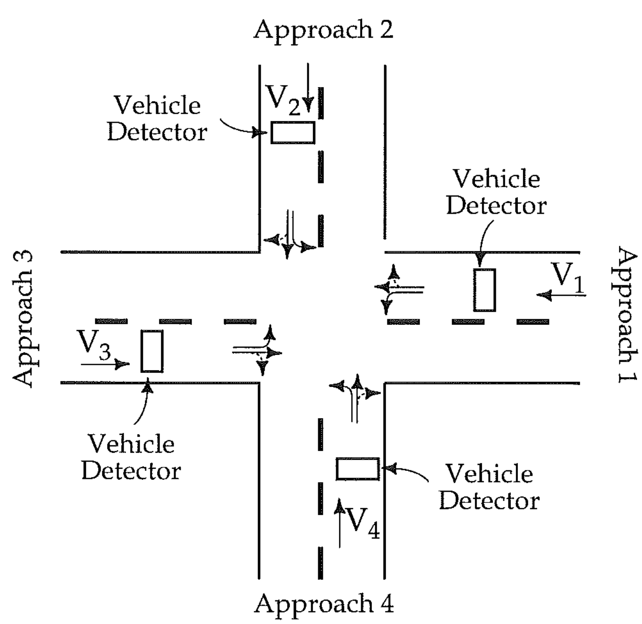


Figure 4. Basic traffic intersection showing approaches, approach volumes, movements and vehicle detectors.

Now consider the signal timing problem given two possible perfect predictions of arrivals during the planning horizon as depicted in Figure 5. Each arrival pattern represents the number of vehicles to arrive at an intersection in fixed time intervals[1]. Both arrival patterns are identical until time t_0 when the signal control has to decide whether to serve this approach or to serve another approach. In the top case, the demand occurs immediately following t_0, where as in the bottom case there is little demand immediately following t_0 and greater demand in the future. In each case the total number of vehicle arrivals are equal. However, the optimal signal timings could be significantly different. It is of fundamental importance to know the temporal arrival distribution to build a truly real-time traffic-adaptive signal control logic.

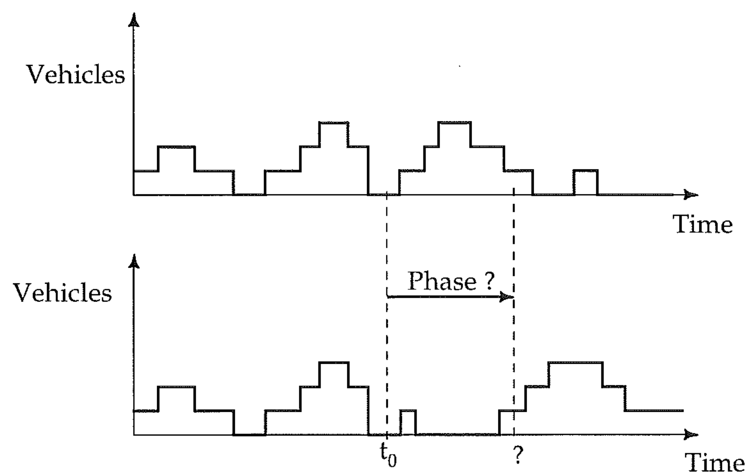


Figure 5. Graphical depiction of the effect of future arrivals on scheduling phase sequences and durations.

Three issues are important to predicting traffic flow: (1) the length of the time horizon, (2) the number of prediction points per time horizon (called the prediction frequency) and (3) the number and location of information sources. The prediction time horizon provides the real-time traffic-adaptive signal timing control logic with the ability to plan future signal timing decisions. If the prediction horizon is short, perhaps several seconds, then the signal timing decisions are restricted. For example, if the predictions are made over a 10-second horizon, the signal timing logic can only make timing decision which extend or shorten the current phase. On the other hand, if the predictions are made over a longer horizon, the signal timing decisions can include decisions on phase termination times and phase sequencing. For example, if the prediction horizon is 30-40 seconds, then the signal

[1]The use of "the number of vehicles" during fixed-time intervals is primarily to display the data. The arrival of vehicles can best be thought of as a point-process characterized by an instantaneous arrival rate with the additional characteristics of position, velocity and acceleration that represent the vehicle as a dynamic entity.

timing logic might schedule the next two phases and their durations based on the predicted demand instead of following a fixed phase sequence.

The prediction frequency provides information about the distribution of vehicle arrivals over time. If the predictions are made at a frequency of only one prediction for the decision time horizon, then the signal timing logic must assume that the vehicles are distributed uniformly over that time. If the predictions are made more frequently, say 10 to 30 times over the prediction horizon, then the signal timing logic will have a more accurate representation of the distribution of vehicle arrivals over time. Figure 6 depicts the information content of predictions at a frequency of once per time horizon and 10 times per horizon.

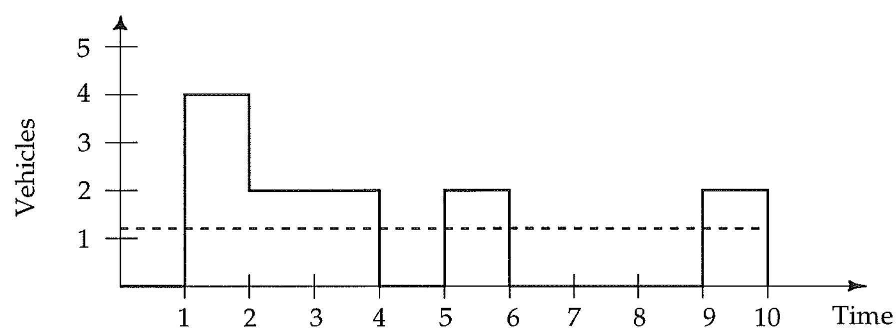


Figure 6. Illustration of the relationship between the prediction horizon (T=10) and the prediction frequency.
(Note: The dashed line shows an average of 1.2 vehicles/time unit. The solid line shows the number of vehicles predicted per each time unit).

Traffic flow is, in general, a time-space phenomenon. The number and location of information sources determine the ability of any prediction algorithm to predict future conditions based on current conditions at other spatial points. For example, if a detector is located, say, 10 seconds upstream of the desired prediction point, then prediction will be easier but only for a 10-second horizon. The further away the location of other information sources, the longer the potential prediction horizon. But, the temporal information may become more distorted (e.g. platoon dispersion) and thus less valuable for prediction. In addition, the further away the information sources are located, the greater are the effects of exogenous factors, such as traffic signals and traffic sources/sinks, on prediction. Clearly, a system with many well placed detectors will provide the best information for prediction.

3.1 The PREDICT Approach

The PREDICT algorithm [Head, 1995] of RHODES uses the output of the detectors on the approach of each upstream intersection, together with information on the traffic state and planned phase timings for the upstream signals, to predict future arrivals at the intersection/interchange under RHODES control. This approach allows a longer prediction time horizon since the travel distance to the intersection is longer and the delays at the upstream signal are considered. A benefit of this approach is that it includes the effects of the upstream traffic signals in the intersection/interchange control optimization problem.

This prediction approach is data driven. That is, the prediction of each arrival at an intersection/interchange depends on the event of a vehicle crossing some detector on the approach to an upstream detector and not (directly) on the traditional time-averaged detection parameters of count and occupancy.

To understand how this approach works consider the scenario shown in Figure 7. It is desired to predict the flow approaching intersection A at detector d_A. Making the prediction for the point d_A is important because it is a point on link AB where the actual flow can be measured, hence the quality of the prediction can be assessed in real-time.

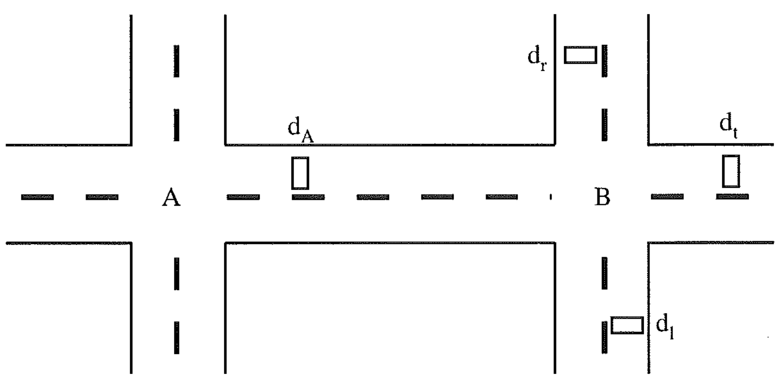


Figure 7. Prediction scenario based on detectors on the approaches to the upstream intersection (B).

Traffic contributing to the flow at d_A originates from the approaches to intersection B and can be measured at detectors d_l, d_t and d_r representing the flows that will turn left, pass through and turn right, respectively, onto link AB. Other traffic that originates at sources between intersections A and B are possible, but will be considered as unmeasurable "noise". Also, it is possible that vehicles passing over d_l, d_t, and d_r will terminate their trip before arriving at d_A. This will also be considered as "noise" in the prediction.

When a vehicle passes a detection point, say d_i where $i \in \{l,t,r\}$, several factors affect when it will arrive at d_A including (1) the travel time from d_i to the stop bar at intersection B, (2) the delay due to an existing queue at B, (3) the delay due to the traffic signal at B, and (4) the travel time between B and d_A.

Figure 8 (a)-(d) depict the delay associated with each of these factors. In Figure 8(a) the vehicle arrives at detector d_i and passes freely to detector d_A. The arrival time, denoted t_a at d_A can be estimated as

$$t_a = t_{d_i} + T_{d_i,s_B} + T_{s_B,d_A}$$

where T_{d_i,s_B} is the travel time from d_i to the stop bar at intersection B and T_{s_B,d_A} is the travel time from the stop bar at intersection B to the detector at d_A. Each of these travel times can be estimated based on the approach speed and the link flow speed, respectively.

In Figure 8(b) the vehicle arrives at detector d_i and is delayed by the signal at intersection B. Hence the travel time from d_i to d_A must account for the travel time from d_i to the stop bar, the delay due to the signal and the travel time from the stop bar to d_A. The arrival time at d_A can then be estimated as

$$t_a = t_{d_i} + T_{d_i,s_B} + T_{u_B} + T_{s_B,d_A}$$

where T_{u_B} is the delay until the signal timing plan advances to a phase that will serve the desired movement.

In Figure 8(c) the arrival at d_i encounters delay for the signal as well as a standing queue, and has to travel from d_i to the stop bar at B, and from the stop bar to d_A. The signal delay can be computed based on the signal timing plan as described above. The delay due to the standing queue can be estimated using a relationship of the form

$$T_{u_B} = a_o + a_1 N_{q_i}$$

where a_0 and a_1 are parameters that can be selected based on the particular intersection and N_q is the number of vehicles in the queue. (The above equation has the form of Greensheild's equation used to estimate the amount of time required to clear a queue. It is based on the dynamics of traffic queue dispersion and has been determined empirically.)

Figure 8(d) depicts the case when the arrival at d_i occurs after the signal has begun serving the desired phase, but a standing queue is present. This case is similar to the above, except that the delay due to the standing queue must be adjusted based on the amount of time that has elapsed between the onset of the signal and the arrival of the vehicle at d_i and the travel time to the back of the queue.

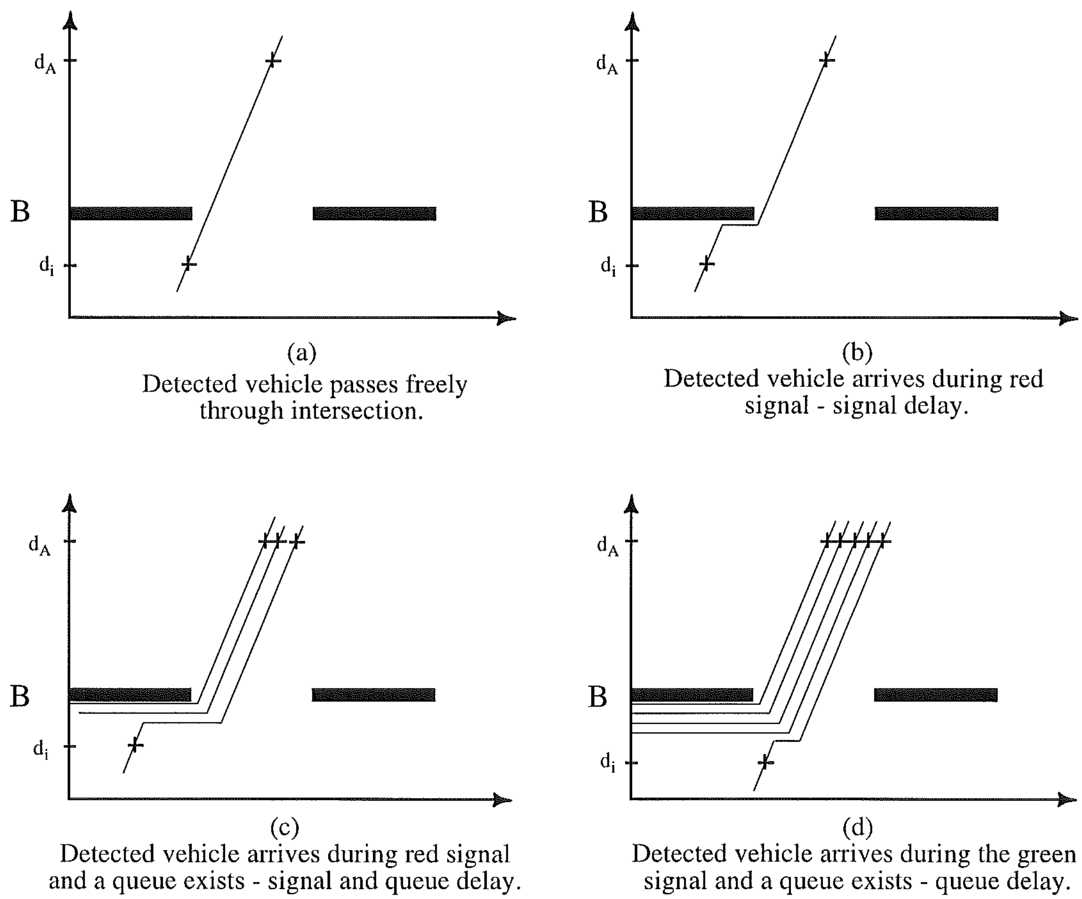


Figure 8. Delays associated with the prediction of arrivals at the detector d_A.

Once the arrival time at d_A is predicted, this model adds the probability of the arrival to the current estimate of the expected number of arrivals at that time. For example, if 15% of the vehicles that pass over d_i continue on to d_A, then 0.15 will be added to the current estimate of the expected number of arrivals at the predicted arrival time t_a.

Note that to use the PREDICT model, several parameters (given in bold) need to be provided: (1) travel times on links (detector to detector) which depends on the **link free-flow speed** and current traffic volumes, (2) **queue discharge rates** which also depends on volumes (as well as opposing- and cross-traffic volumes), and (3) **turning probabilities**. In addition to these parameters, to estimate arrivals and demand for various phases we also need to have **estimates of queues** at the intersections and the ramps.

Link free-flow speed can be estimated from historical data and capacity analysis. Link free-flow speeds are even needed in traditional off-line models to optimize fixed phase timings (cycle times, offsets, splits) so to obtain these should pose no major problem.

Through-traffic queue discharge rates are effected by preceding through-traffic volumes, which can be easily measured. Likewise, left-turn queue discharge rates depend on opposing traffic volumes, and right-turn queue discharge rates depend on cross-traffic in that direction. These three discharge rates can initially be given from calculated default functions - functions of traffic volumes, but can be adjusted based on how well they predict remaining queues at the stop-bar presence detectors. For example, if the left-turn queue estimate tends to be larger than actually measured queue (see next subsection on the queue estimate algorithm, QUEUE) then the left-turn discharge rate is adjusted upwards.

Besides the estimation of queues one also needs to estimate turning probabilities. The TURN algorithm given in a following subsection updates turning probabilities every few minutes based on recent turning percentages.

It is important to note that the PREDICT model is based on processing arrival data as it becomes available. At any point in time the predicted arrival flow pattern at d_A accounts for vehicles that have already passed the detectors d_r, d_t and d_l. The benefit of this vehicle-additive process of the predictor is that it constantly provides, for a given prediction horizon, (1) nearly complete information of anticipated vehicle arrivals in the very near future (of those vehicles that have already passed the upstream intersections) and (2) partial information of anticipated vehicles in remaining part of the prediction time horizon (of those vehicles that have not passed the upstream intersections, since some new vehicles may still arrive that will effect the delays in the prediction time horizon).

Also, this algorithm is distributed in that it can be applied for every approach of every intersection/interchange in a large urban traffic signal control network. Results of an evaluation study of the PREDICT algorithm for arrivals at an intersection have been reported by Head [1995].

QUEUE Algorithm

There have been a few algorithms that have been reported in the literature that address the problem of estimating queues at an intersection using detector information, most notably that of Baras et al. [1979]. However, these are not applicable because they require excessive computational effort and time that is not available in our real-time prediction scenario. Instead, for our purpose, we developed a simple estimation procedure of accounting arrivals and departures at the stop-bar detectors. Suppose at the beginning of a green phase, say at time t_0, our initial queue estimate at some stop-bar is $q(t_0)$. At the end of the green phase, say at time t_1, the remaining queue $q(t_1)$ is given by

$$q(t_1) = q(t_0) + a(t_1, t_0) - d(t_1, t_0)$$

where $a(t_1, t_0)$ is the number of predicted arrivals between t_1 and t_0, and $d(t_1, t_0)$ the predicted number of departures (using a given queue discharge rate). What allows us to keep biases from creeping into the estimates is that at some epochs we are certain of the remaining queue lengths: when there are no queues at the stop bars - as confirmed by the stop bar presence detectors. If the estimated queue is positive while the stop-bar presence detectors indicate no queues then we effectively decrease our estimate and make it zero. If the estimated queue was zero while the stop-bar indicates a positive queue, the estimated queue is increased to one.

When the queue discharge rate at the stop bar is estimated well, it would be expected that, on the average, half the time the estimated queues will be greater than the actual queues and half the time less than actual queues. If the estimated queues more often than not tend to be higher than the actual queues (i.e., when there were no vehicles while the queue estimate was nonzero) then we adjust the queue discharge rate upwards by a small amount, if it tends to be less (i.e., when there were vehicles while the estimates were zero) then we adjust it downwards by a small amount. We note that the adjustment of queue discharge rate is only possible when there are no queue spillback into the intersection; in this case queue discharge rate is zero because of blockage and does not depend on traffic volumes. This consideration is important in our interchange scenario since there are times when the on-ramp may result in momentary queue spillbacks due to some surges of high traffic volumes.

TURN Algorithm

Both the ICOP and the PREDICT models require that turn probabilities are specified. An assumption for RHODES (as well as current off-line methods to set signal timings) is that some estimates for turn probabilities at the intersection are given. Even the CORSIM model and existing off-line signal optimization algorithms need this information. However, from the real-time traffic control perspective, these probabilities are not deterministic; they change stochastically over time. For example, suppose P_{WN} is the probability that a vehicle arriving from the West to some intersection will turn left (North), then it is clear that P_{WN} will depend on the time of the day, the volume of traffic, and the particular mix of the origins/destinations in the group of arrivals being modeled. In other words, P_{WN} is described by a random process.

Our assumptions for P_{WN} are (1) a prior estimate is available whose uncertainty is modeled with a Normal distribution with known mean and variance; (2) at any given time, we have measured the percentage of vehicles that have turned left in the last, say, five minutes, as well as percentages that turned right and driven straight through the intersection (these three percentages should add up to 100% since U-turns are considered to be negligible); and (3) we know the error distributions for these measurements. We had a choice of three turning probabilities models: (1) Information Minimization/Entropy Maximization [Mekky, 1979; van Zuylen, 1979; Hauer, Pagitsas and Shin, 1981]; (2) Bayesian [Maher, 1984]; and (3) Maximum Likelihood [Maher, 1984; Nihan and Davis, 1989]. The Bayesian model was picked for implementation since the other two models involved a nondeterministic number of iterations based on an error tolerance whereas Bayesian method consisted of exactly 7 iterations. In the Bayesian method, prior variances for the turning volume errors can be used along with the prior means, unlike for the other two methods. The covariances of the turning volume errors are assumed to be zero since the traffic detectors are assumed to operate independent from one another.

Researchers have discussed different methods for determining priors, using historical data, intersection categorization or simply taking educated guesses [e.g., van Zuylen, 1979; Hauer, Pagitsas and Shin, 1981; Schaeffer, 1988; Furth, 1990]. Currently, we plan to use historical data when available, or expert judgments of local traffic engineers, or the default values of 25% for left turns, 50% for straight through and 25% for right turns.

The TURN algorithm uses a Bayesian approach to update current prior estimates given the current detector measurements [Maher, 1984].

To describe the TURN algorithm, let

T_{ij} - The number of cars going from i to j.
P_{ij} - The turning probability for going from i to j.
X_{ij} - The prior turning probability for going from the i to j.
O_i - The number of cars entering the intersection from the i direction.
D_j - The number of cars exiting the intersection to the j direction.
β - The variance in the turning volumes.

At an intersection there are four measured input streams (from origins, O_N, O_E, O_S, O_W), four measured output streams (from destinations, D_N, D_E, D_S, D_W). These inputs and outputs result in 12 movements, each corresponding to a turning probability. We can relate these measurements and turning probabilities using eight linear equations, one of which is linearly dependent. Thus we can iteratively solve for the turning probabilities in seven iterations using the algorithm described below. First, to define notation we let

$$\text{Let } \mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{g} = \begin{bmatrix} O_N \\ O_E \\ O_S \\ O_W \\ D_N \\ D_E \\ D_S \end{bmatrix}.$$

where $H = \{h_{ij}\}$ with

$$h_{ij} = \begin{cases} 1 \text{ if input/output stream } i \text{ contributes to movement } j \\ 0\text{, otherwise.} \end{cases}$$

TURN Algorithm

Step 1: $y = \dfrac{\sum_j D_j - \sum_i O_i}{\sum_j D_j + \sum_i O_i}$; $\hat{O}_i = O_i(1+y)$ for all i ; $\hat{D}_j = D_j(1-y)$ for all j.

Step 2: k=1; compute

$$\mu_1^0 = X_{\mathrm{NE}} O_{\mathrm{N}} , \ \mu_2^0 = X_{\mathrm{NS}} O_{\mathrm{N}} , \ \mu_3^0 = X_{\mathrm{NW}} O_{\mathrm{N}} ,$$
$$\mu_4^0 = X_{\mathrm{ES}} O_{\mathrm{E}} , \ \mu_5^0 = X_{\mathrm{EW}} O_{\mathrm{E}} , \ \mu_6^0 = X_{\mathrm{EN}} O_{\mathrm{E}} ,$$
$$\mu_7^0 = X_{\mathrm{SW}} O_{\mathrm{S}} , \ \mu_8^0 = X_{\mathrm{SN}} O_{\mathrm{S}} , \ \mu_9^0 = X_{\mathrm{SE}} O_{\mathrm{S}} ,$$
$$\mu_{10}^0 = X_{\mathrm{WN}} O_{\mathrm{W}} , \ \mu_{11}^0 = X_{\mathrm{WE}} O_{\mathrm{W}} , \ \mu_{12}^0 = X_{\mathrm{WS}} O_{\mathrm{W}} ;$$
$$v_{ii}^0 = \beta \mu_i^0 \text{ for all } i , \ v_{ij}^0 = 0 \text{ for all } i \neq j .$$

Step 3: $\delta_i = \sum_{j=1}^{12} h_{kj} \mu_i^{k-1}$, for all i; $S_i = \sum_{j=1}^{12} h_{kj} v_{ij}^{k-1}$ for all i; $W = \sum_{i=1}^{12} h_{ki} S_i$.

Step 4: $\mu_i^k = \mu_i^{k-1} + \dfrac{S_i}{W}(g_k - \delta)$, $v_{ij}^k = v_{ij}^{k-1} - \dfrac{S_i S_j}{W}$ for all i and j.

Step 5: If k=7 then Stop, else let k=k+1 and go to Step 3.

Then the estimated turning volumes are:

$$T_{\mathrm{NE}} = \mu_1^7 , \ T_{\mathrm{NS}} = \mu_2^7 , \ T_{\mathrm{NW}} = \mu_3^7 , T_{\mathrm{ES}} = \mu_4^7 , \ T_{\mathrm{EW}} = \mu_5^7 , \ T_{\mathrm{EN}} = \mu_6^7 ,$$
$$T_{\mathrm{SW}} = \mu_7^7 , T_{\mathrm{SN}} = \mu_8^7 , \ T_{\mathrm{SE}} = \mu_9^7 , T_{\mathrm{WN}} = \mu_{10}^7 , \ T_{\mathrm{WE}} = \mu_{11}^7 , \ T_{\mathrm{WS}} = \mu_{12}^7$$

and the estimated turning probabilities are

$$P_{ij} = \frac{T_{ij}}{\sum_k T_{kj}} \text{ for all } i \text{ and } j.$$

In Step 1, we eliminate inconsistent traffic volumes, when the number of cars entering is not equal to the number of cars exiting. Because of counting errors, time delays and other errors, those sums might not be equal. The procedure implemented is the one due to van Zuylen [1979], where the volumes D_{W} are changed if these volumes are not balanced.

If a negative turning probability results in some computation, then this probability is set to a low value of 0.01 and the other two probabilities for that movement are adjusted proportionally [Hall, van Vliet and Willumsen, 1980; Mountain, Maher and Maher, 1986].

3.2 Interchange Predictions

Unlike the case of predictions of vehicle arrivals and queues at an intersection, in the case of the interchange we need to get detector data from additional sources such as off-ramps and on-ramps detectors, and predict arrivals/queues at two intersections and two on-ramps.

Referring to Figure 9, in the interchange scenario the PREDICT/QUEUE/TURN methods takes (1) input from passage detectors (one for each lane) at upstream arterial intersection locations {A, B, C, D, E, F}, from passage detectors at interchange locations {G, H, H", I, J, K, K", L}, and from passage detectors at off-ramps M and N, for predicting arrivals, (2) input from presence detectors (one for each lane) at locations {A', B', C', D', E', F'}, from presence detectors at interchange locations {G', H', I', J,' K', L'}, from passage detectors at off-ramps M' and N', and from presence detectors at on-ramps {O, O', P, P'} for predicting queues, and (3) outputs prediction of arrivals/queues at interchange locations R and S and at on-ramps U and V. It is important to note that predictions requires the state of the signals at all times and the ramp metering rates. As we indicted earlier, the scope of this project does not allow our RHODES-ITMS algorithm to set ramp metering rates; the algorithm assumes that these rates are set externally to the interchange control. These rates need to be provided to PREDICT for estimating on-ramps queues in real-time.

The prediction equations for the interchange are similar to those for PREDICT for the intersection but there are more cases to consider. For example arrivals at location K' depends on the arrival streams from locations D', E', F', L', and J' and the corresponding phase durations at intersections S and T. Queues at on-ramp V depends on arrival streams from locations J', H' and L', the signal timings at intersection S, and the ramp-metering rate at on-ramp V.

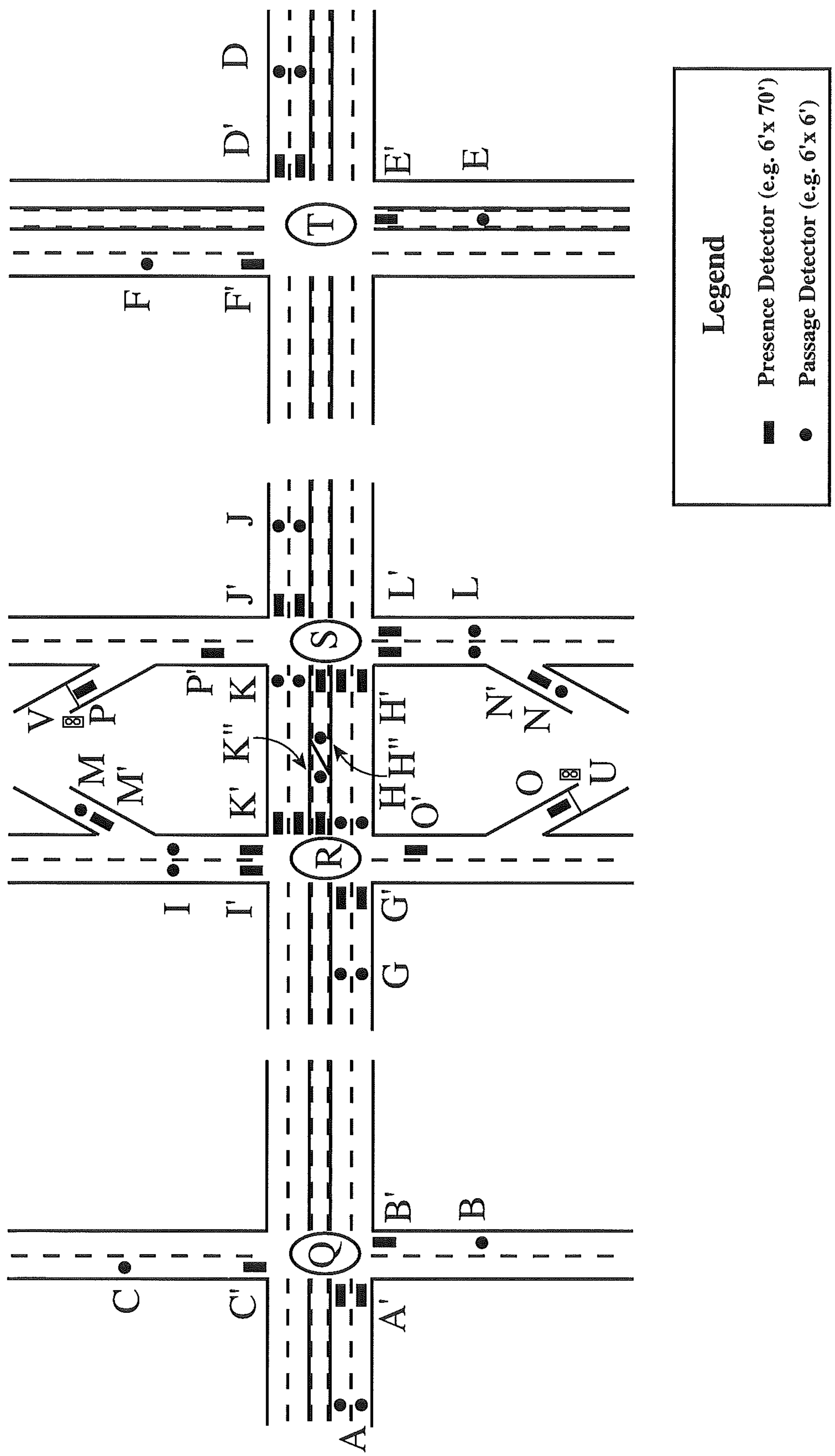


Figure 9 Detector locations for intersection predictions

4. INTERCHANGE TRAFFIC CONTROL

In the earlier RHODES Project report we reviewed the state of the art on real-time traffic-adaptive signal control and established that with the current communication and computer technology a dynamic programming (DP) based algorithm is practical and would produce better traffic-based performance measures then existing methods such as Time-of-day Fixed Time Control, Semi-actuated Control, and Actuated Control, and other traffic-adaptive approaches currently operating.

Existing control strategies are based on a signal timing plan defined in terms of operating parameters for traditional signal control, namely *cycle time, splits*, and *offsets*. These parameters are generally developed based on traffic studies and standard procedures, such as the Highway Capacity Manual, or signal timing software such as TRANSYT and PASSER. The traffic studies result in estimates of traffic conditions, link volumes and turning percentages, for specified time periods. Signal timing <u>parameters</u> are developed for each of these time periods and, typically, implemented on a time-of-day basis with no consideration of current actual traffic conditions. In many cases, even the use of standard procedures for the development of signal timing plans is abandoned and traffic engineers operate in a judgment-based fashion with moderate levels of success. None of these approaches is truly traffic-adaptive or even attempt to actually minimize some measure of traffic performance such as average vehicle-delay.

Currently available traffic responsive systems attempt to address the problem of responding to actual traffic conditions by switching these <u>parametric</u> signal timing plans based on current wide-area traffic conditions rather than time of day. This requires that signal timing parameters be developed for a variety of possible traffic conditions. Nevertheless, implicit in the usage of <u>parametric</u> timing plans, is the assumption that for the next several minutes, or hours, the traffic is the network can be well characterized by the measured <u>average</u> flows and parameters. No account is taken of the fact that the second-by-second and minute-by-minute variabilities of traffic are significant and plans based on averages produce unnecessary delays for some traffic movements when the traffic on conflicting movements is absent, or very small, during some periods.

The RHODES approach is to predict both the short-term and the medium term fluctuations of the traffic (in terms of individual vehicle arrivals and platoon movements respectively), and explicitly set phases that maximize a given traffic performance measure. Note that we do not set timing plans in terms of cycle times, splits and offsets, but rather in terms of phase durations for any given phase sequence. (A pre-specified phase sequence is not necessarily required by RHODES, but since many traffic engineers prefer a pre-specified sequence RHODES has been developed to allow the traffic engineer to specify a desired sequence.)

There are other signal timing schemes which have been experimented that do not provide parametric timing plans but instead provide phase durations, notably OPAC [Gartner, 1983; Gartner et al., 1991] and PRODYN [Khoudour et al., 1991] and UTOPIA [Mauro and DiTaranto, 1990]. In some ways these too use dynamic programming or related optimization schemes, but, in their current implementations, the underlying models are more approximate and the methods less efficient [Sen and Head, 1997].

In the RHODES traffic-adaptive control strategy, the emphasis shifts from changing timing parameters in <u>reacting</u> to traffic conditions just observed to <u>proactively</u> setting phase durations for predicted traffic conditions. The RHODES system is a hierarchical decomposition of the system-wide traffic control problem into three levels of subproblems: network loading, network flow control, and intersection/interchange control. In order to be proactive, predicted information is used at each level in the hierarchy for the optimum allocation of available traffic carrying capacity as effected by the control decisions at each of the hierarchical levels [Head, Mirchandani and Sheppard, 1992].

At this lowest level of the control hierarchy for a surface street network, that is, at the intersection control level, we have developed the optimization algorithm COP, reported earlier [Head and Mirchandani, 1994]. For this project it was generalized to the freeway-arterial diamond interchange. The tightly-coupled operation of the two intersections comprising the interchange has been traditionally addressed by using a specialized phasing strategy. Here we develop a real-time interchange control scheme where the total and average vehicle delays at the interchange are greatly reduced by explicitly considering these delays in selecting phase durations.

The extension of adaptive optimization algorithms to arterial-freeway interchanges, which are frequent components of urban traffic networks, has been only marginally addressed in the literature. However, many studies have addressed the trade-offs associated with the use of various types of phasing and timing strategies [Messer et al., 1977; Messer and Chang, 1987; Herrick, 1992], geometric configurations [Messer and Malakapalli, 1992], and progression schemes for interchanges [Messer et al., 1974]. The general conclusions of such investigations including field studies and performance comparisons [Munjal, 1971] indicate that specific geometric configuration, phasing, and timing strategies should be selected for each interchange based on local traffic volumes, safety considerations, future volume predictions, and other local environmental variables upon initial installation. Regular updates of the timing plan are recommended, as with any intersection, to account for changes in the various parameters, as changes in traffic conditions evolve over time. With these concerns in mind, an adaptive control strategy becomes more desirable if such a scheme can emulate various types of phasing strategies and timing choices, eliminating the need to choose a regimented timing and phasing strategy "up front" for a given geometric configuration.

In many jurisdictions, the need for an adaptive control strategy at the arterial-freeway interchange has been identified and addressed by the use of a fully-actuated controller. Such a strategy may balance the needs of freeway on/off-ramp traffic with the needs of arterial cross-traffic to some degree, but usually disrupts any progression along the arterial and does not consider potential ramp spillbacks. There have been successful efforts to integrate interchanges in a progression scheme by using a semi-actuated control system at the interchange, but such systems have been shown to perform poorly as the traffic volume approaches the capacity of the interchange.

In light of the fact that many such interchanges exist in the interior of a metropolitan area, an adaptive control strategy, that uses information, namely predicted vehicle arrivals, from the surrounding intersections and detector locations, can provide simultaneously the benefits of arterial progression and better freeway service. In this project, ICOP, a real-time traffic-adaptive control strategy for a freeway-surface street diamond interchange was developed and evaluated with a TRAF-NETSIM simulation model.

4.1 Controlled Optimization of Phases (COP) for Intersections

COP is a real-time traffic-adaptive control algorithm based on rolling-horizon dynamic programming [Sen and Head, 1997]. Multivariable dynamic programming (DP) is notorious for its "curse of dimensionality" casting initial doubt on the possibility to obtain an optimal, or even a sufficient solution, in real-time to a complex control problem. Unlike OPAC, which uses a restricted search heuristic to reduce the solution space for its DP optimization [Gartner, 1983; Gartner et al., 1991], the COP algorithm was formulated with a single state variable (s_j) to solve the general intersection control problem efficiently and to optimality. By doing so, COP provides, in real-time, the phase timings and switching decisions after each stage of optimization to minimize some cumulative performance measure for the intersection, such as the total vehicle delay. Additionally, the generality of the DP formulation allows the use of a variety of other performance measures such as stops and/or queue lengths (although queue lengths are generally not directly measurable).

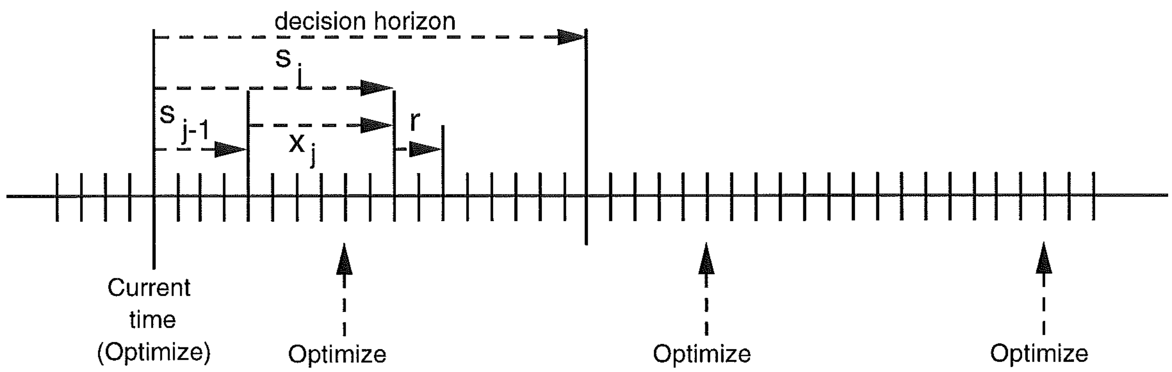


Figure 10. Implementation of Single-Variable Rolling Horizon Approach

Since phasing sequence and timing decisions are applicable only for a short time-horizon, the problem is formulated such that the optimization horizon T is long enough to account for sufficient propagation of traffic flow to get good predicted performance measures, but "short enough" to allow for computations in real time. Furthermore, COP implements timing decisions only for the first few seconds of the optimization horizon T, as depicted in Figure 10. As predicted vehicle arrival information is updated with additional detector actuations, the algorithm runs again to re-optimizes the phase timings and sequence for the un-implemented portion of the previous optimization horizon, and propagates the optimization an additional t time units into the future. In this way, COP produces a rolling-horizon optimization that can extend the timing of phases previously scheduled given new updated vehicle arrival information.

4.2 Controlled Optimization of Phases (ICOP) for Diamond Interchanges

Given the last predictions provided by the PREDICT model, ICOP proceeds as follows at any iteration: (1) collect current detector actuations at the interchange and adjacent intersections and update the detector database, (2) use the PREDICT model to propagate those detector actuations into the future and update flow predictions, (3) assess various phase timing/switching decisions with the ICOP rolling-horizon dynamic programming algorithm. The algorithmic basis of ICOP is exactly like that of COP, but the computation of performance function is different. For the sake of completeness, we will reintroduce a slightly modified notation of COP [Sen and Head, 1997] and restate the algorithm before we discuss the computation of the performance function.

<u>Notation</u>

T = Total number of discrete time steps. Each time step is of length Δ, and is indexed by t in $[1,T]$.

$P = \{ \phi_1, \phi_2, \ldots \phi_N\}$, set of possible phases.

x_j = Control variable denoting amount of green time allocated to stage j.

s_j = State variable denoting the total number of time steps allocated after stage j is completed.

$g(\phi_j)$ = Minimum green time (integer number of time steps) for phase ϕ_j.

$r(\phi_i|\phi_k)$ = Clearance ("all-red" phase + start-up loss time) interval, if required (integer number of time steps) between phases ϕ_i and ϕ_k.

$p(j)$ = phase decision in stage j; $p(0)$ = initial phase in P

$X_j(x_j)$ = Set of feasible control decisions, given state s_j.

$f_j(s_j,x_j)$ = Performance function at stage j and control x_j

$V_j(s_j)$ = Value function (cumulative value of prior performance) given stage s_j is completed.

The major difference in this formulation as compared to those that have appeared in the literature (e.g., OPAC), is that, say, k^{th} stage in the DP denotes the allocation of green time to the k^{th} phase in any given sequence. For example if the phase sequence ABCABCABC was been considered, first stage makes the decision whether phase A should be allocated x_1 amount of green time, the second stage whether phase B should be allocated x_2 amount of green time after phase A is finished, and so on. Effectively, from the fact that some phases may be allocated zero green time then any sequence of green phases are possible. For example x_1=2, x_2=0, x_3=4, x_4=0, and x_5= 3, means that we start with 2 time steps for phase A, then 4 time steps for phase C and finally 3 time steps for phase B, effectively resulting in the sequence ACB. That is, any phase sequence can result from the unrestricted ICOP algorithm. However, it is important to note that a sequence of phases must be given (in this case ABCABCABC) and considered in order for the DP to work and to identify phases with stages. We shall refer to this phase sequence as *phase order* for the algorithm.

The ICOP's DP framework allows us to take into account any operating and jurisdictional constraint imposed by the traffic engineer. For example, if it is required that a phase sequence is specified and phases cannot be skipped, then the phase order will correspond to the given phase sequence and a non-zero minimum green time threshold will be imposed on all the phases in the phase order. Or, if it is allowed that non-conflicting movements of two consecutive phases "overlap" instead of requiring explicit clearance phase, then this is also possible in the DP framework.

Suppose any phase sequence is allowed, but a minimum green time of $g(\phi_j)$ is required if phase ϕ_j is used. Then, given the value for the state variable s_j, the control variable x_j (for the j^{th} stage and the corresponding phase in the given phase order) can assume values from the following feasible discrete set.

$$X_j(s_j) = \begin{cases} 0 & \text{if} \quad s - r < g, \\ 0, g, g+1, \ldots, s_j - r & \quad \text{otherwise.} \end{cases}$$

If a clearance interval of length r is required between phases, then we use the following relationship between the stages of the DP (see Figure 10):

$$s_{j-1} = s_j - h_j(x_j)$$

where

$$h_j(x_j) = \begin{cases} 0 & \text{if} \quad x_j = 0, \\ x_j + r(\Phi_k, \Phi_j) & \quad \text{otherwise.} \end{cases}$$

On the other hand, if some phases are allowed to overlap, then depending on the current non-zero duration phase decision $p(j)$ and the last non-zero duration phase decision $p(j\text{-}k)$, clearance interval $r(p(j)/p(j\text{-}k))$ may be set to zero.

An important assumption for dynamic programming is that the objective function of the optimization should be additive, namely

$$v_j(s_j) = f_1(s_1, x_1) + f_2(s_2, x_2) + ... + f_j(s_j, x_j)$$

which is true for many types of performance functions. For the standard traffic performance measures, total delays and stops, this performance measure is additive. In this project we used delays (historical delays and expected delays) as our performance criteria that we tried to minimize with our traffic controls.

The algorithm is now presented briefly in the standard two stages of dynamic programming. First, a forward recursion is performed for the possible phase sequences and timings and the associated performance measures are calculated at each recursive step. To initialize, $v_0(s_0)$ is set to zero. The DP starts with j=1 and proceeds recursively for j=2,3,... until $v_j(T)$ does not change for a whole cycle in the phase order. In the implementation of the algorithm, for computational efficiency, we simply terminated the DP after a fixed number (M) of phases were determined. The resulting errors from the optimal solution were small. At each stage in the forward recursion, the method computes the optimal value of $v_j(s_j)$, and stores the optimal $x_j(s_j)$ and the corresponding phase ϕ_j, for all possible values of s_j:

Forward Recursion

Step 0: Initialize $v_0 = 0, j = 1$

Step 2: For $s_j = r, ..., T$

$v_j = Min_{x_j} \{f_j(s_j, x_j) + v_{j-1}(s_{j-1}) \mid x_j \text{ in } X_j(s_j)\}$

Record $x^*_j(s_j)$, an optimal soultion to above

Step 3: If $j < M, j \leftarrow j+1$ then repeat Step 1.

Otherwise, when $j = M$, STOP.

The optimal phase sequence and timings are retrieved in a backward pass, beginning at the end of the time horizon and progressing towards the beginning. Since M denotes the last stage for which the value function $v_j(s_j)$ has been computed, we simply perform the following backward recursion starting with $j = M-1$, and ending with $j = 1$.

Retrieval of Optimal Phase Decisions

Step 0: $s^*_M = T$

Step 1: For $j = M, M-1, ..., 1$

Read $x^*_j(s_j)$ from forward recursion record; $s^*_{j-1} = s^*_j - h_j(x^*_j(s_j))$

Computation of Performance Function

If one does not consider the queues at the ramps, then the essential differences between COP + PREDICT and ICOP + PREDICT are:

(1) the phases and the phase order are different,
(2) additional constraints for clearing the queued vehicles on the arterial segment between the two intersections (of the interchange) may be imposed, and
(3) predictions of arrivals and queues for more locations are needed.

Here the performance functions based on delays are essentially the same. We refer to the Sen-Head [1996] paper for details on the computation of these performance functions. In the computation of the delay-based performance measure, every second a vehicle is delayed at the intersections being controlled contributes an additional one vehicle-second to the overall objective function value. Equivalently, one-second delay of a vehicle C_T going straight through on the arterial is equal to one-second delay of a vehicle C_L waiting to make a left turn on to the on-ramp.

However, when one considers the queues at the on-ramps which may result in an additional delay for C_L, then the last statement is not necessarily correct. For example, if one knows that it will take 10 seconds for vehicle C_L to reach the ramp-meter but the current queue q(t) at the ramp meter is large so that the vehicle will have to wait an additional 3 seconds at the on-ramp anyway, then there is no decrease in the total delay of vehicle C_L through the interchange if it is held back another 3 seconds at the intersection traffic signal. In the meantime, other vehicles like C_T may be able to get through the interchange without incurring traffic signal delay.

On the other hand, if there is no queue, or a small queue, at the ramp meter then one-second delay of C_L does in fact contribute 1 vehicle-second to the overall objective function value. Thus, depending on the size of queue, if it exists, at the on-ramp, as well as on the turning probability for getting on to the ramp, a one-second delay of vehicle C_L may have to be "discounted" using a multiplicative factor ρ, $0 \leq \rho \leq 1$.

Referring to Figure 11, we let P_{LR} and P_{LS} be the respective probabilities that vehicle C_L will go to the on-ramp for the freeway, or straight through on the frontage road;

$P_{LR} + P_{LS} = 1$. Let H be the travel time from the intersection to the ramp meter. Let $q(t)$ the queue at the ramp at time t, and $\lambda(t)$ the ramp-metering rate in vehicles per second. Then the time to clear the queue is $q(t)/\lambda(t)$. Finally, let Q_R be the queue capacity (buffer) from the intersection to the ramp meter, which means $q(t) \leq Q_R$.

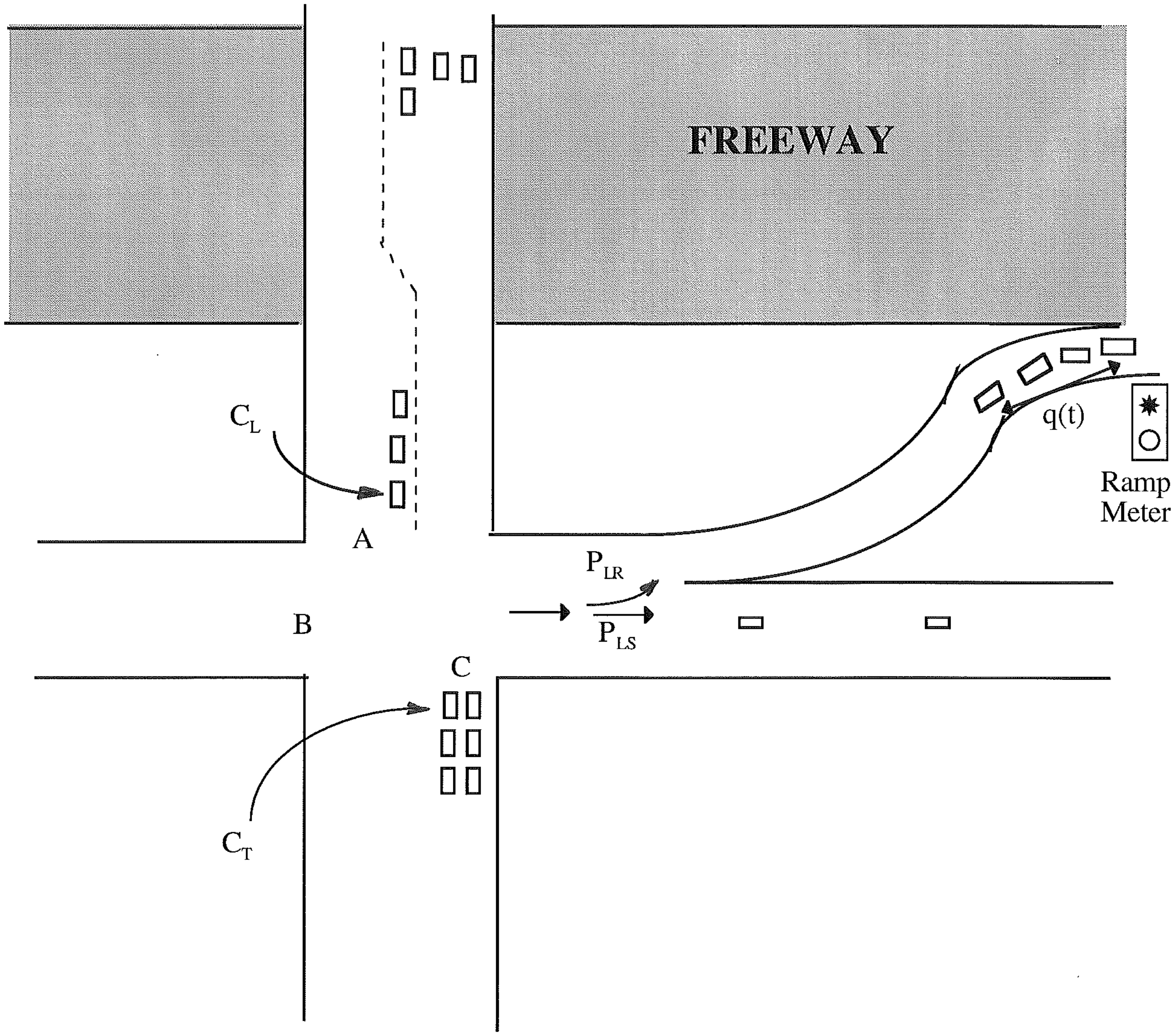


Figure 11 Turn probabilities and queues at ramps

When $q(t)/\lambda(t) \leq H$, a delay of one second for vehicle C_L contributes 1 vehicle-second to the objective function value because by the time the vehicle gets to the ramp meter there will be no queue there. Then, ICOP uses exactly the same delay functions as COP.

When $H\,\lambda(t) \leq q(t) < Q_R$, then a one-second delay of vehicle C_L will either result in the contribution of 1 vehicle-second to the objective function value if the vehicle is planning to go down the frontage road -- with probability P_{LS} -- or no contribution if the vehicle is planning to use the on-ramp where it will stay in a first-come-first served queue -- with probability P_{LR}. Hence the delay in this situation is "discounted" with a multiplicative factor $\rho = P_{LS} \bullet 1 + P_{LR} \bullet 0 = P_{LS}$.

When $Q_R = q(t)$ then there is a momentary spillback and vehicle C_L cannot go anywhere, even if it receives a green signal. In that case we just as well provide the green phase to one of the conflicting movements and discount the delay completely (ρ =0).

In summary,

$$\rho = \begin{cases} 1 & \text{when} \quad q(t) \le H\,\lambda(t) \\ P_{LS} & \text{when} \quad H\,\lambda(t) \le q(t) \le Q_R \\ 0 & \text{when} \quad q(t) = Q_R. \end{cases} \tag{1}$$

The same applies for <u>all</u> the vehicle streams going towards the on-ramps at the interchange, including the straight-through vehicles from point B and right-turn vehicles from point C.

There is one other case that needs to be considered in this scenario. If there is a left-turn bay at point A, as it is generally so, then holding the vehicles at A may result in a queue spillback into the through lane adjoining A. In that case it may be better to "dump" some of the vehicles on to the ramp buffer as long as $q(t) < Q_R$. In this case, (eqn. 1) is modified as follows

$$\rho = \begin{cases} 1 & \text{when} \quad q(t) \le H\,\lambda(t) \\ 1 & \text{when} \quad H\,\lambda(t) \le q(t) \le Q_R \text{ and } q_L(t) \ge Q_L \\ P_{LS} & \text{when} \quad H\,\lambda(t) \le q(t) \le Q_R \text{ and } q_L(t) \le Q_L \\ 0 & \text{when} \quad Q_R = q(t). \end{cases}$$

where $q_L(t)$ is the queue of left-turn vehicles at point A at time t, and Q_L is the capacity of the left-turn bay. We remark that if ρ remains 0 for a long time then this implies that the interchange is over-saturated and our RHODES strategy becomes applicable no longer.

5. LABORATORY TESTING AND RESULTS

5.1 Integration of CORSIM, PREDICT and ICOP

As described earlier, CORSIM was used for laboratory testing the real-time traffic control algorithms for the freeway-arterial diamond interchange. An actual interchange was selected for the simulation model. The particular, the interchange that we selected for laboratory testing the algorithms was the diamond interchange between freeway I-17 in Phoenix and the arterial Indian School Road. Indian School Road is four-lane highway, two lanes each direction, with left-turn bays at the interchange. Vehicle volumes and turning movements were estimated from observing the traffic at the interchange where possible, otherwise educated guesses were made. Current system timing strategies, and associated timing parameters were also obtained from ADOT. Physical characteristics of the interchange were observed and required measurements such as distances from the intersections to the ramp-meters, the lengths of left-turn bays, distances between the detectors and stop bar, etc., were obtained from a overhead photograph of the interchange. Although the simulation model was not validated - to see if it exactly represented the I-17/Indian School interchange, the project team felt that it was sufficiently realistic for laboratory testing of the algorithms. The simulation model was also demonstrated to the Project's Technical Advisory Committee to ascertain its realism.

Once the Project Team had the simulation model running, it integrated the PREDICT and the ICOP algorithms as was done in the previous project (Head and Mirchandani, 1994). Briefly, the simulation procedure for the integrated set-up worked as follows:

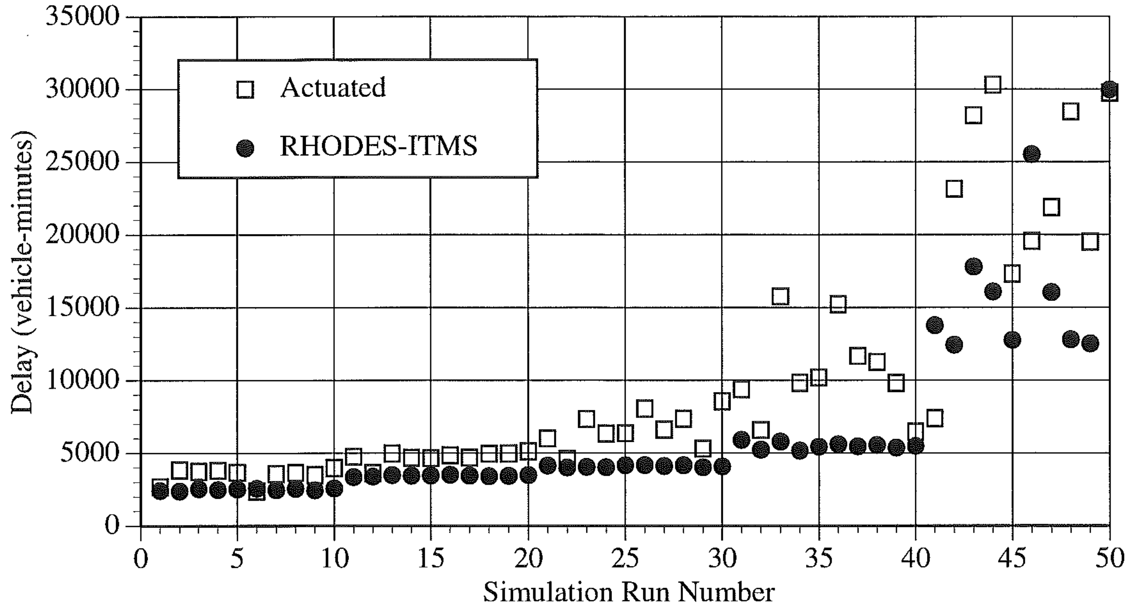
1. Check if specified simulation end time has been reached. If yes, STOP.
2. Otherwise, simulate the traffic dynamics for one second (by execution of CORSIM)
3. Collect specified measures of effectiveness
4. Collect detector outputs and the signal states in the last second
5. Update predicted arrivals and queues (using PREDICT)
6. Check if ICOP needs to be executed for the given predictions. If no, then go to Step 1.
7. Otherwise, run ICOP for the specified rolling horizon. Schedule time for next ICOP execution
8. Implement recommended phase decisions, and go to Step 1.

5.2 Simulation Results

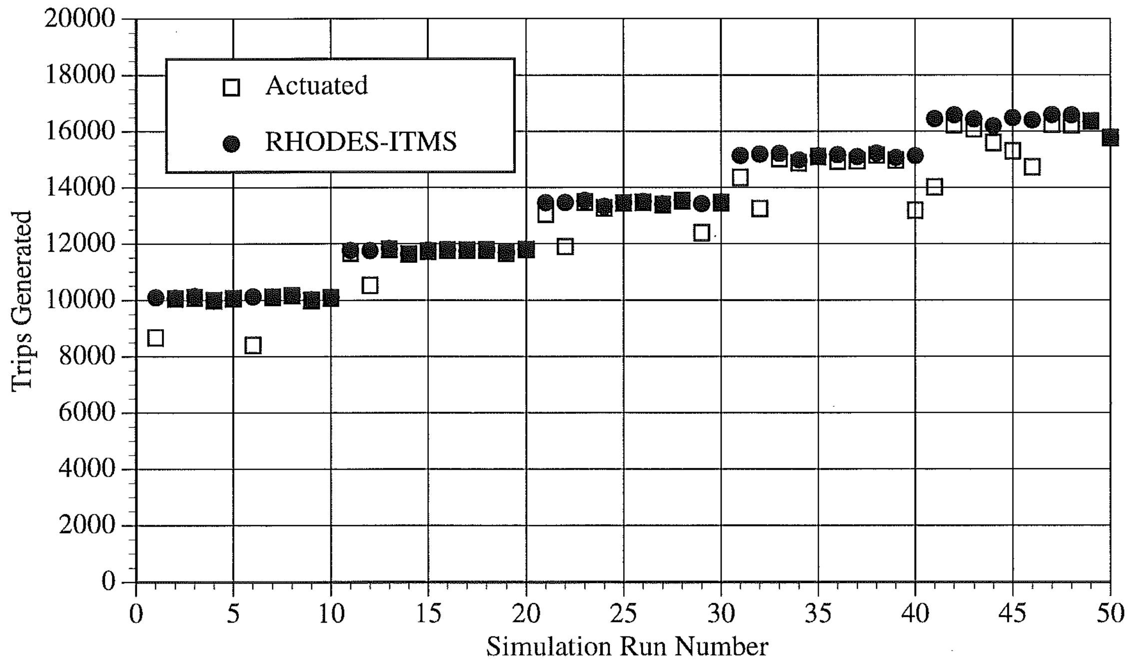
The simulation results available to date are from runs of the simulation model (1) with fully actuated control (which is how the interchange is currently controlled), and (2) with parts of the RHODES-ITMS system (the parts currently implemented). Specifically with regard to the runs with RHODES-ITMS, the estimation and consideration of queues at the ramp meters have not yet been included in the simulation results reported below. It is anticipated that RHODES-ITMS system will show even better performance if the ramp queues and delays are considered.

Three scenarios were considered in the simulation testing. The first scenario consisted of a well-timed fully actuated control using signal phasing currently adopted by ADOT. This was for comparison purposes. The second scenario used an implementation of the RHODES algorithms that included some consideration of past delays of queued vehicles - with the perspective of not allowing some vehicles to have very large queue delays. The third scenario did not include past delays, but instead constrained the green time to be below a given maximum to assure that no vehicles were queued for a very long time in highly congested situations. Preliminary tests showed that the third version of RHODES with maximum green time specified (we used the maximum green time used by ADOT in their fully actuated control logic) performed most consistently at the diamond interchange. The results reported below for "RHODES-ITMS" scenario are based on that version.

Five loading factors were considered. For each of the loads, ten runs were conducted for the "actuated" scenario and ten for the "RHODES-ITMS" scenario. Each run was for 120 minutes (in simulation time). In the low load case approximately 5000 vehicle-trips per hour were sent through the interchange. In the low-medium load case, approximately 5900 vehicle trips per hour were loaded. In the medium load case, about 6800 vehicle trips per hour were loaded. In the high-medium load case, about 7500 vehicle trips per hour were loaded. Finally, in the high load case, about 8300 vehicle trips per hour were loaded. Figure 12 shows the results of these experiments. Figure 12(a) shows total delay (in vehicle minutes) of the vehicles through the interchange; and 12(b) the total number of vehicles trips on the links of the interchange. (In Figure 12(b) the many of the squares indicating "actuated" results cannot be seen because they are covered by the solid circles indicating "RHODES-ITMS" results.)



(a)



(b)

Figure 12 (a) Total vehicle delay and (b) vehicle trips generated using actuated control and RHODES-ITMS control strategies

Observe that except for the high-load case, the total vehicle delay for RHODES-ITMS was <u>always</u> lower than for the fully actuated scenario. RHODES-ITMS reduced average delays by 28% for the low and low-medium loads, 38% for the medium load, and 48%

for the high-medium load. Also, very importantly, the standard deviations of these delays were reduced by more than a factor of ten. This indicates that RHODES-ITMS is able to reduce significantly the variance of the delays of the vehicles moving through the interchange. Observe also that, in these cases, the number of trips through the interchange were mostly the same for actuated and RHODES-ITMS scenarios, which should be expected since in steady state the number entering must equal to the number exiting for otherwise the queues would grow indefinitely and oversaturation would take place.

In the high load case, not all vehicles make it through the interchange for some of the runs, and few large queues remain at the end of the simulation period. This is specially true for some of the "actuated" runs where the number of trips through the interchange are significantly lower than the loaded trips (see runs 31, 35 and 36 in Figure 12b). Still, RHODES-ITMS reduced the average delay by 25% for these trips, while the standard deviations wereof the same order of magnitude. Also, RHODES-ITMS was able to get about 5% more trips through the interchange during the simulation period.

To account for the delay based on throughput, one can plot delay per vehicle (in seconds) versus trips per hour, by dividing the total delay (in vehicle seconds) by the number of vehicles trips completed during the simulation period. Figure 13 gives such a plot.

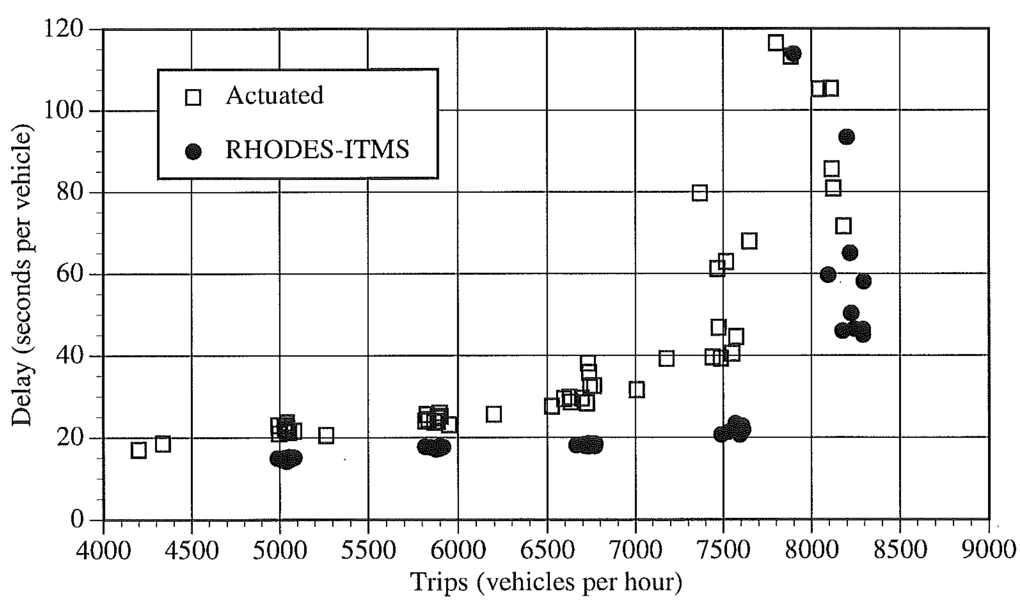


Figure 13 Average vehicle delay versus throughput (vehicle trips per hour) using actuated control and RHODES-ITMS control strategies

Using the data of Figure 13 we see that, in the low load case, RHODES-ITMS results in vehicle delays of 14.8 seconds versus the 21.3 seconds for the actuated case - a 30% improvement. For the low-medium case the respective numbers are 17.5 seconds for RHODES-ITMS and 24.3 seconds for the actuated case (a 25% improvement); for the medium case they are 18.2 seconds for RHODES-ITMS and 30.2 seconds for the actuated case (a 40% improvement); and for the high-medium they are 21.8 seconds for RHODES-ITMS and 43.2 seconds for the actuated case (a 50% improvement). For the high load case most vehicle delays are between 45 - 65 seconds for RHODES-ITMS while they are between 65-85 for the actuated case. (As we mentioned above, because the large queues at the end of the simulation period, the large delays of over 100 seconds are unrealistic, and are probably due to the artifacts of the experiments and the limitations of the simulation model). Nevertheless, the average vehicle delay is lower for RHODES-ITMS than for the fully actuated case; 62.4 versus 85.7 seconds (a 27% improvement)., again the standard deviation of the of the vehicle delays deceased about ten-fold for the low to high-medium cases, and decreased by 70% for the high load case.

Another observation and conjecture can be made from the high-load results of the simulations. In near oversaturated conditions, large queues result in significant queue spillback and, hence, sometimes there are excessive delays through the interchange. It is perhaps because of this that the variances in average delays are high for the high-load case. However, because of the lower variance for RHODES-ITMS and fewer incidents of excessive delays, one may conjecture that RHODES-ITMS probably delays the onset of queue spillback and oversaturated conditions. To test this, it would be useful to perform transient analysis for studing performance measures such as “time to queue spillback”, and “time to settle back to steady state” during large surges of traffic; the research team plans to do that in the future.

6. CONCLUSIONS AND DIRECTIONS OF FUTURE RESEARCH

The RHODES-ITMS Project addressed the design and development of a real-time traffic adaptive control system for Freeway-Arterial Diamond Interchanges using the concepts underlying the RHODES traffic-adaptive signal control system. The traffic "controls" at a Freeway-Arterial Diamond Interchange are the two sets of traffic signals located at the arterials, on both sides of the freeway, and the ramp meters at the on-ramps to the freeway. To truly manage all the traffic at the interchange, we need to do both, set the phase durations of the intersection traffic signals and control the ramp-metering rates, taking into account network-wide freeway objectives as well as local traffic objectives. However, the scope of the project was only to control the traffic signals, with ramp-metering rate given externally. Nevertheless, the current design of the RHODES-ITMS system, as described in the report, uses information about the queues at the on-ramps in order to decrease the overall delay of all the vehicles which use the arterials, the frontage roads parallel to the freeway (if they exist) and the ramps at the interchange. The basic goal of the project was to predict arrivals and queues of individual vehicles at the arterial approaches on both sides of the freeway, as well as arrivals from the off-ramps and the departures and queues at the on-ramps, and based on these predictions and a given criterion of performance (e.g., minimize total vehicle-delay) to determine the optimal phasing of the signals at the two intersections on either side of the freeway.

To test the algorithms, it was necessary to first do so in the "laboratory". In this regard the Project Team developed a CORSIM-based simulation model which was used as a platform to test our RHODES-ITMS strategy at the intersection/interchange level. This project was able to demonstrate the applicability and the significant potential of the RHODES concept towards traffic responsive control of the signals at an interchange. Of course, it is clear that the project team has just conducted some preliminary research in this important traffic management area. Much more needs to be done to further implement the RHODES-ITMS system discussed in the report and to field test the system. In particular, the research team needs to (1) implement the estimation of queues at the ramp-meters, (2) implement the ICOP with consideration of these queues, (3) laboratory test the resulting RHODES-ITMS system, (4) implement the field hardware, and (5) test the RHODES-ITMS System in the field. The currently funded RHODES-ITMS Phase II Project is performing these tasks which are scheduled to be completed in December 1997. In a way, this report may be considered an Interim Report for the overall ITMS Project.

REFERENCES

J. S. Baras, W. S. Levine and T. L. Lin (1979), "Discrete-time Point Processes in Urban Traffic Queue Estimation, " *IEEE Transactions on Automatic Control, AC-24*, pp. 12-27.

P. Dell'Olmo and P. B. Mirchandani (1995), "REALBAND: An Approach for Real-Time Coordination of Traffic Flows on Networks" *Transportation Research Record 1494*, pp. 106- 116.

P. G. Furth (1990), "Model of Turning Movement Propensity," *Transportation Research Record 1287*, pp. 195-204.

N. H. Gartner (1983), "OPAC: A Demand-Responsive Strategy for Traffic Signal Control." *Transportation Research Record 906*, pp. 75-81.

N. H. Gartner, P. J. Tarnoff and C. M. Andrews (1991), "Evaluation of the Optimized Policies for Adaptive Control (OPAC) Strategy," *Transportation Research Record 1324*, pp. 105 -114.

M. D. Hall, D. van Vliet and L. G. Willumsen (1980), "SATURN - A Simulation-Assignment Model for the Evaluation of Traffic Management Schemes," *Traffic Engineering and Control 21*, pp. 168-176.

E. Hauer, E. Pagitsas and B. T. Shin (1981), "Estimation of Turning Flows from Automatic Counts," *Transportation Research Record 795*, pp. 1-7

K. L. Head (1995), "An Event-Based Short-Term Traffic Flow Prediction Model," *Transportation Research Record 1510*, pp. 45-52.

K. L. Head and P. B. Mirchandani (1994), "RHODES PROJECT Phase II (A)" Final Report, FHWA-AZ94-383, Arizona Department of Transportation, Phoenix, Arizona.

K. L. Head, P. B. Mirchandani and D. Sheppard (1992), "Hierarchical Framework for Real-Time Traffic Control," *Transportation Research Record 1360*, pp. 82-88.

G. C. Herrick and A. M. Carroll (1992), "Strategies for Improving Traffic Operations at Oversaturated Diamond Interchanges," Technical Report (Volume 4). 1148-4F, Texas Transportation Institute, Texas.

L. Khoudour, J-B Lesort and J-L Farges (1991), "PRODYN - Three Years of Trials in the ZELT Experimental Zone", *Recherche - Transports - Securite*, English Issue, Special Traffic Management, pp. 89-98.

Michael J. Maher (1984), "Estimating the Turning Flows at a Junction: A Comparison of Three Models," *Traffic Engineering and Control 25*, pp. 19-22.

V. Mauro and D. Di Taranto (1990), "UTOPIA" *Proceedings of the 6th IFAC/IFIP/IFORS Symposium on Control and Communication in Transportation*, Paris, France.

A. Mekky (1979), "On Estimating Turning Flows at Road Junctions," *Traffic Engineering and Control 20*, pp. 486-487

C, J. Messer, R. H. Winston and J. D. Carvell (1974), "A Real-Time Frontage Road Progression Analysis and Control Strategy," *Transportation Research Record 503,* pp. 1 -12.

C. J. Messer, D. B. Fambro and S. H. Richards (1977) "Optimization of Pretimed Signalized Diamond Interchanges," *Transportation Research Record 644*, pp. 78-84

C. J. Messer and M. S. Chang (1987) "Traffic Operations of Basic Traffic-Actuated Control Systems at Diamond Interchanges," *Transportation Research Record 1114*, pp. 54-62

C. J. Messer and M. P. Malakapalli (1992),"An Applications Manual for Evaluating Two and Three Level Diamond Interchange Operations using TRANSYT-7F". Technical Report 1148-3, Texas Transportation Institute, Texas.

L. J. Mountain, M. J. Maher and S. Maher (1986), "The Estimation of Turning Flows from Traffic Counts: 1. At Four-Arm Intersections," *Traffic Engineering and Control 27*, pp. 501-507

P. K. Munjal (1971), "An Analysis of Diamond Interchange Signalization " *Highway Research Record 349*, pp. 47-64

N. L. Nihan and G. A. Davis (1989), "Application of Prediction-Error Minimization and Maximum Likelihood to Estimate Intersection O-D Matrices from Traffic Counts," *Transportation Science 23*, pp. 77-90.

D. E. Sheppard, K. L. Head, S. Joshua and P. B. Mirchandani (1996), "Simulation-Based Methodology for Evaluation of High-Occupancy-Vehicle Facilities", *Transportation Research Record 1554*, pp. 90-98.

M. C. Schaeffer (1988), "Estimation of Intersection Turning Movements from Approach Counts," *Institute of Traffic Engineering Journal 58*, pp. 41-46.

S. Sen and K. L. Head (1997), " Controlled Optimization of Phases at an Intersection" *Transportation Science 31*, pp. 5-17.

H. J. van Zuylen, (1979 "The Estimation of Turning Flows on a Junction," *Traffic Engineering and Control 20*, pp. 539-541.

APPENDIX

Simulation-Based Methodology for Evaluation of High-Occupancy-Vehicle Facilities

DENNIS E. SHEPPARD, K. LARRY HEAD, SARATH JOSHUA, AND PITU B. MIRCHANDANI

There has been an increasing interest in improving the use of transportation facilities as environmental and social concerns have grown and as financial resources for infrastructure expansion have become increasingly scarce. Numerous programs for increasing carpooling, vanpooling, and transit usage have been undertaken to decrease reliance on single-occupant vehicles and increase the use of multioccupant vehicles. One program has been to develop facilities that give preferential treatment to high-occupancy vehicles (HOVs). Although HOV facilities have been implemented, they often have been found to be unsuccessful in attaining their stated or implied goals. Because interest in the use of HOV facilities is growing, there is a need to improve the ability to evaluate and compare design alternatives in the context of realistic (stochastic) environments. Simulation modeling has long been recognized as a powerful tool for such purposes. A structured simulation-based methodology for the evaluation of HOV design alternatives is presented. An example case study for a corridor in the Phoenix, Arizona, metropolitan area is used to demonstrate the methodology.

Over the past few decades there has been a growing interest in increasing the use of transportation resources in response to environmental and social concerns. One issue of interest is the traveler's predominant dependence on the single-occupant vehicle (SOV). Many facilities such as freeways are often under used because of dependence on the SOV. One method to entice drivers, particularly daily commuters, to multioccupant vehicles or high-occupancy vehicles (HOVs) has been through preferential treatment and incentive programs. These programs can range from reduced toll charges and parking fees for HOVs to the establishment of HOV lanes and HOV bypass ramps on freeways.

Although the political drive is often present for HOV implementation, transportation planners and engineers have found it difficult to rationalize diversion of the scarce financial resources for highway construction and operation to implement HOV facilities. Two primary obstacles to HOV construction are (a) the lack of clearly defined, attainable goals and objectives and (b) the lack of an effective planning and evaluation process.

The lack of clearly defined objectives was addressed in a study of HOV operations conducted by Turnbull et al. (*1*). That study reviewed the evaluation procedures and outcomes of a large number of HOV facilities throughout the country. It noted that the failure to have clearly defined goals, objectives, and measures of effectiveness, although not exclusive to HOVs, made the performance of HOV operations difficult to assess. The study concluded that HOV analyses suffer from a variety of issues including insufficient data on short- and long-term performance, a lack of consistent benchmarks, and uncertainty of the effects of HOV implementation on the overall facility performance.

When conducted, evaluations of HOV performance have been through the use of before-and-after studies to provide the necessary measures of effectiveness. In addition, the measures considered often reflect a limited view of the effects of HOV operations on the entire corridor. An appropriate evaluation should include the entire travel corridor consisting of the freeway, frontage roads, and parallel and cross arterials. This effort will require a substantial data collection effort to properly and fully evaluate the effects on the corridor. Furthermore, the reliance on before-and-after studies for establishing the effectiveness of a facility may overlook the serious negative impact of HOV operations on the overall performance of a corridor without a reasonable expectation of attaining established objectives. For example, unattainable HOV penetration levels may be required to yield a successful outcome.

Some studies have been conducted to assess the feasibility of HOV operations before implementation. In a comprehensive review of HOV operations, Batz (*2*) noted that HOV feasibility studies have had serious limitations because of the expense of the evaluation process, often being restricted to evaluating a single design alternative rather then a wide range of options. Batz found that HOV treatments frequently failed to meet stated objectives or were deemed ineffective, often for reasons that remained undefined during the preimplementation studies but that were revealed after implementation. The inability to perform comparative analyses of various design options can lead to the selection of an inadequate design. Early comprehensive analyses could reveal potential shortcomings of the design of a particular option and therefore allow for either design modifications before implementation or the development of other alternatives with more acceptable, expected performance.

Given the extensive costs required to implement an HOV facility, it would be desirable to establish a methodology to evaluate the potential performances of various design options and to determine the performance sensitivity to variations in underlying conditions. In this paper a simulation-based methodology for evaluating design options is discussed. An example case study for a corridor in Phoenix, Arizona, will be used to demonstrate this methodology.

D. E. Sheppard, K. L. Head, and P. B. Mirchandani, Department of Systems and Industrial Engineering, University of Arizona, Tucson, Ariz. 85721. S. Joshua, Arizona Transportation Research Center, Arizona Department of Transportation, Tempe, Ariz. 85284.

SIMULATION AS EVALUATION TOOL

Simulation modeling has long been recognized among both practitioners and researchers as a powerful tool for investigating and evaluating design concepts. With increasing accessibility to more powerful computers, simulation is frequently being used as a design tool. Simulation models can readily provide insight into the attributes and shortcomings of a design and can yield realistic expectations of the design's performance. Thus, a simulation-based methodology can provide the ability to assess a broad set of what-if scenarios and to garner a wide range of performance measures for comparison among these scenarios, ultimately allowing engineers and decision makers to select an option that provides the highest probability of success.

Within the transportation community simulation models have long been used in a wide range of applications. Of interest in this paper is the application of simulation to the operations and interactions of arterial and freeway traffic systems. Simulation models for the freeway environment (FREFLO and FRESIM) and the urban street network (NETFLO and NETSIM) have been developed by FHWA under the TRAF family of simulation models. Privately developed simulation programs, such as INTEGRATION, add to the options available for evaluating transportation designs. These simulation packages offer a wide range of capabilities for efficiently and effectively characterizing, modeling, and evaluating transportation alternatives.

The use of traffic simulation models as a transportation design tool was effectively demonstrated by Cohen and Clark (*3*) in their evaluation of alternative designs for freeway bridges entering the Washington, D.C., area. Their study showed that a simulation-based methodology could be used to compare a wide range of design concepts with sparse data, whereas traditional evaluation methods (e.g., the *Highway Capacity Manual*) were vulnerable to the lack of data and failed to provide insight into the dynamics of traffic flow. Besides the study by Cohen and Clark the literature on freeway design and, specifically, HOV evaluations indicates that simulation modeling is only now beginning to receive attention as a design tool. This may be because of the complexity of the required input, a lack of understanding of the simulation process, and the identification of appropriate HOV performance measures. At first glance simulation modeling methodologies may appear to be overly complicated and therefore too costly to incorporate into the design process. A structured modeling approach can reduce this complexity and can provide the designer with the ability to access a potentially powerful evaluation tool. It is also crucial that the designer identify the variables that measure the performance of design alternatives and that the simulation model be able to compute these variables.

EVALUATION METHODOLOGY

A structured methodology requires that a framework be established within which studies are conducted, options are assessed, and decisions are evaluated. This framework should include the establishment of goals, objectives, and evaluation criteria as well as the identification of the relevant data that are required to perform the simulations. The limitations of the simulation models must be clearly understood since the models may not be able to evaluate all aspects of a particular design. Considerations such as safety, for example, generally do not fall within the realm of currently available traffic simulation models. A framework for the evaluation of HOV design options is provided in Figure 1.

The first step of this methodology plays a significant role in the development of a successful study. As the evaluation process continues, particularly in the testing of the simulation model (Step 5) and performing experiments (Step 6), simulation results can lead to insights into the operation of the real system that may be further explored by repeating parts of the evaluation process. This iterative process can deepen the understanding of the underlying dynamics of the system, thereby improving the ability to effectively assess and compare alternative designs.

EXAMPLE CASE STUDY

Throughout this paper a case study will be used to illustrate the methodology. The corridor used in the example consists of a 13-km segment of I-17 in Phoenix, bounded by Thomas Road on the south and Thunderbird Road on the north (Figure 2). The corridor consists of three mainline freeway lanes and two-lane frontage roads in each direction, with access points spaced at approximately 1.6-km (1-mi) intervals where they intersect with major urban arterials. The corridor has no HOV facilities, but they are being considered for implementation.

In addition to the freeway, three streets that run parallel to the freeway are included in the study area: 19th Street, 27th Street, and 35th Street. These street segments consist of five- and six-lane cross sections with primarily four-phase signals at the intersections. To simplify the model, minor arterials that do not provide direct access to the freeway were not considered since these would have a minimal impact on traffic flow within the corridor.

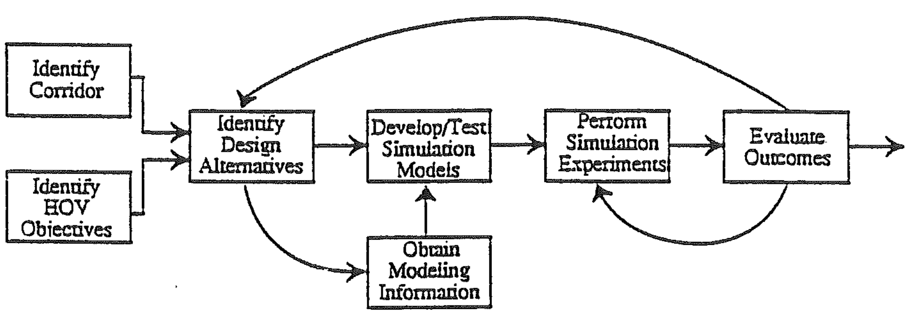


FIGURE 1 Simulation-based methodology for design of HOV facilities.

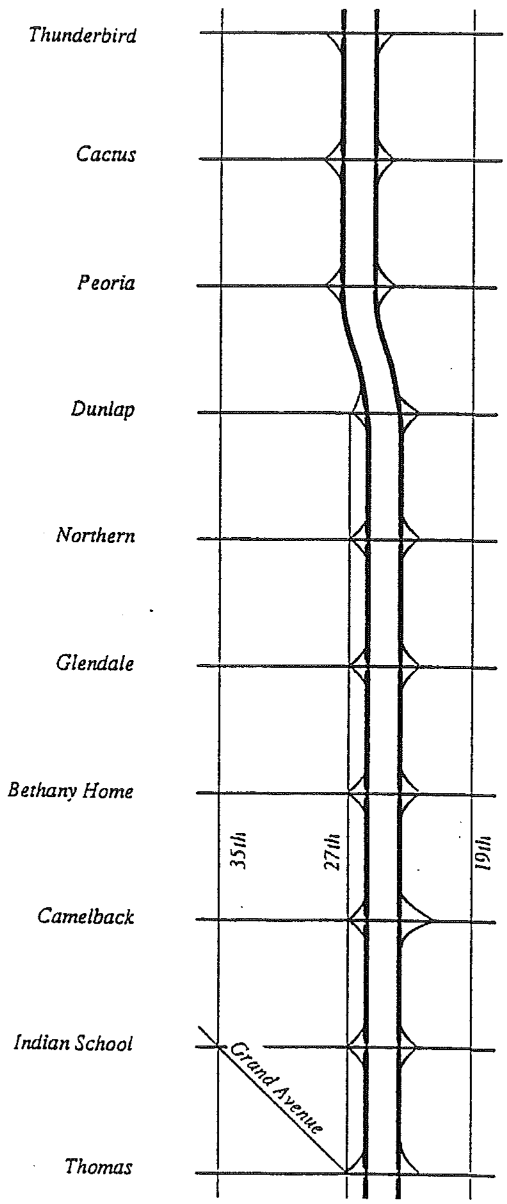


FIGURE 2. I-17 HOV study corridor.

DISCUSSION OF STEPS IN METHODOLOGY

Identify Corridor Area of Interest

The identification of the corridor boundaries defines the scope of the study and models. It should include, minimally, the freeways, frontage roads, and major arterials that could be affected by the implementation of an HOV facility. Realistically, the impact of an HOV facility diminishes with distance from the facility at a rate dependent on its extent as well as on the likelihood of HOV trips, their lengths, and their destinations. To evaluate the impact on parallel streets, they can be modeled, where appropriate, as a single facility instead of numerous smaller facilities. Streets running to the HOV facility should extend beyond any adjacent parallel street segments to mimic the network operations and turning movements appropriately.

Identify HOV Objectives

The identification of the HOV objectives was noted by both Turnbull et al. (*1*) and Batz (*2*) as a element that is critical to the successful implementation of HOV facilities. Batz found that HOV facilities were often deemed to have failed because of the different objectives of those charged with the evaluation process and of the decision makers; the objectives of the latter often involved political considerations rather than solely traffic performance measures. For facilities to be deemed effective, a better understanding of the objectives of all involved parties, including transportation planners, engineers, decision makers, and users, needs to be incorporated into the design and evaluation as early and as fully as possible. Care needs to be taken to ensure that attendant measures of effectiveness reflect measurable and realistic expectations.

The study by Turnbull et al. (*1*) identified a wide range of objectives that included slowing the growth in vehicle miles traveled (VMT), reducing congestion, increasing HOV usage in the corridor, and encouraging public transit. Imprecise goals, such as reducing congestion or slowing the growth in VMT, may be neither realistic nor attainable since these goals depend on the traffic demand within the corridor remaining unchanged. Latent demand, currently unserved by existing facilities, could fill newly gained capacity, maintaining VMT for the existing general-use lanes at current levels while increasing the number of person trips or person miles traveled. Generally, those objectives that are framed in terms of passengers may be suitable goals that can lead to realistic measures of effectiveness for HOV operations.

Reliance on a single measure of effectiveness as the test of success or failure of a facility should be avoided. For example, if increasing the number of person trips is viewed as the exclusive objective, it would be easy to determine the point at which HOV operations are functionally successful, yet the solution might ignore other factors that are negatively affected or unnoticed if only this objective is used. Therefore, care should be taken at early stages to ensure that realistic objectives are established, with an understanding that these may need to be modified as the evaluation proceeds.

For the purposes of the case study, HOV penetration and vehicle occupancy rates were developed from a study conducted by Lee Engineering (*4*) that assessed occupancies and vehicle classifications within the Phoenix metropolitan area. Three levels were used to establish evaluation points for HOV penetration on the freeway: a lower bound of 5 percent, a nominal value of 10 percent, and an upper bound of 15 percent of the entering traffic volumes. Average occupancy rates of 1.3 persons per vehicle for general-use lanes and 2.5 persons per vehicle for HOV lanes were used. This implies that at the upper bound, 25 percent of all person trips would be carried by the HOV lane(s), which is well in excess of the current level of less than 17 percent person trips during peak hours and 11 percent during uncongested periods but that is in conformance with the national lower-bound use rate cited by Emerson and Strickland (*5*).

Identify Appropriate Design Alternatives

The selection of appropriate design alternatives should take into consideration the economic, social, physical, and operational factors that may make HOVs effective. For the freeway environment, applicable alternatives might include (a) conversion of one or more existing general-use lanes to HOV operations, (b) the addition of one or more HOV lanes, (c) the addition of general-use lanes, and (d) preferential ramp treatments.

Simulation models may not be appropriate for assessing the effectiveness of some strategies; however, elements of these strategies may be incorporated into the models. Incentives such as HOV park-and-ride lots that are operated to increase use and that may affect portions of the corridor under study may be incorporated into the options model. In developing alternative representations of incentives, a series of submodels may be used to determine desirable combinations of incentives. The same methodology applied for the evaluation of HOV alternatives can be used for detailed evaluation of the HOV incentives.

Obtain Modeling Information

The collection of information to calibrate and validate a simulation model is often the most difficult and costly portion of the evaluation of HOV alternatives. In general, the greater the detail and accuracy of the data collected or estimated, the more accurately the simulation emulates the existing and future system conditions. Yet, from the study by Cohen and Clark (*3*), it is evident that although data quality affects simulation results, acceptable results can be achieved with sparse data sets and by appropriately estimating unknown parameter values. In those cases in which estimates are used, consideration should be given to treating such parameters as variables for the purposes of testing the model's sensitivity to the estimates.

Information requirements for simulation programs such as CORFLO or CORSIM can be categorized into four types: physical network description, traffic flow data and characteristics, origin-destination data, and signal operations data.

Physical Network Description

Physical network description and signal operations information are data types that are readily accessible. At a minimum physical data should include street, freeway, frontage road, and ramp-lane characteristics, signalized intersection geometrics, link or segment lengths, vehicle detection locations, and any lane controls in effect. Other physical data that substantially affect flow should be included where appropriate.

Traffic Flow Data and Characteristics

In contrast to the ease and relatively low cost of collecting information on the physical network and signal operations, collecting data on traffic flows can be time-consuming and costly. Besides existing and projected traffic volumes and turning movement data, information such as (a) operating speeds, (b) HOV percentages, and (c) the vehicle mix (trucks, buses, etc.) for each roadway is needed and should be estimated if insufficient data exist. However, because of the uncertainty of traffic forecasts, the operating speeds, HOV percentages, and vehicle mix may be better treated as parameters considered in the experimental design. Other information on, for example, vehicle headways, signal lost times, and HOV occupancy can be established by using default settings from traffic engineering manuals (e.g., the *Highway Capacity Manual*) or estimated from available information. Often, realistic estimates of many parameters can be made by local operators and planners.

Origin-Destination Information

Origin-destination information represents the most difficult data type needed in the simulation model. If this type of information is available, traffic assignment models can be used to assign vehicles to routes across an entire network. For static conditions these assignment models can produce turning percentages. The locations of sources and sinks (shopping malls, employers, etc.) that generate (or absorb) significant volumes should be identified and vehicle generation rates estimated for the period(s) under consideration. Special attention should be paid to HOV-oriented facilities, such as park-and-ride lots, when high turnover can be anticipated during short time periods.

Modeling Information for Phoenix Care Study

For the Phoenix case study details on the network and freeway flows were available from a variety of sources and provided much of the background information needed to develop the model. Turning movement and traffic volume information was obtained from an I-10–I-17 corridor study (*6*). Volumes were increased on a few segments to emulate current conditions. Signal operations, with the exception of the interchange signals, were set to provide coordinated movement between signalized intersections. Interchange controller parameters were obtained from the Arizona Department of Transportation. Furthermore, the intersection and interchange controllers were coordinated. Coordination of the interchange and network signals was a departure from existing conditions of no coordination, but this reflects planned operational changes. Information on vehicle occupancy, HOV percentages, and vehicle mix was obtained from the study by Lee Engineering (*4*). No origin-destination information was available for the study area.

Develop, Test, and Validate Simulation Models

The first step in the process is to develop a model of existing conditions. The importance of this step is twofold. It forms the basis for model validation, and it establishes a baseline condition for comparison with the design alternatives. Beyond the preparation of the model, a number of steps are integral to model development and are required to run the experiments needed to assess the alternative designs. These can be categorized as (a) validation of the simulation model, (b) selection of time frames for simulation, (c) selection of variable ranges, and (d) establishment of the number of replications.

Validation of Model

Validating the model for the existing conditions is a vital step in the simulation process and can be achieved in numerous ways. Kelton and Law (*7*) identify several methods for validating models, including empirical testing of model assumptions and statistical comparisons of output results. Comparison of simulation results with traditional traffic measures (travel times, stopped time delay, and numbers of stops) collected from the field would allow for tuning of the model to actual conditions and would ensure the highest confidence in the model. However, limited field data in conjunction with expert opinion can be used to tune and validate a model. Transportation professionals familiar with traffic networks and system operations, in conjunction with available field information, often have the knowledge and experience required to assist in detecting irregularities and fine-tuning the simulation model.

Selection of Simulation Run Times

The simulation time period should be of sufficient length to ensure that the collected statistics have stabilized. Generally, simulation times may be obtained through trial and error although a time period greater than twice the mean travel time through the corridor should be sufficient (the experimenter should check this assumption before accepting it).

Most simulation software, such as the TRAF family, begin collecting statistics after the model has attained some equilibrium condition. In the TRAF family this condition is defined as the point in time when the number of vehicles exiting the network is equal to the number entering the network. Although this definition represents one form of equilibrium, it does not guarantee statistical steady-state conditions, and an initial bias can result. The initial bias can be avoided by defining two or more simulation time periods. By experimenting with different durations, the statistical output for the two periods can be compared to see if steady state has been achieved.

The experimental results for the example network are presented in Figure 3. The results demonstrate how one particular statistic, person delay, collected in the second time period varies as a function of the first or initial period. With a freeway entry volume of 5,000 vehicles per hour (vph), the results indicate that after an initial period of 14 min network delay reached a steady-state condition. In this case the first time period could end and the second time period could begin at the 15-min mark (an even number of signal cycles for each simulation run).

The effort to eliminate initial bias is to ensure (a) consistency among replications and (b) that the variance observed is due to the stochastic nature of the traffic activity and not the simulation process itself. It should be noted that the network may not reach equilibrium as discussed by Rathi and Venigalla (*8*), in which case a closer examination of the results may be in order to assess and correct (where appropriate) the causes for this.

Selection of Ranges for Parameter Values

Ranges selected for parameter values should reflect base and projected values or the range of uncertainty for each type of parameter, or both. Typically, a lower and an upper bound should be selected for each parameter type. In conjunction with these bounds, midpoints may provide insight into the effects of parameter changes over the given range. As the simulation study proceeds, additional parameter values may be needed to gain greater insight.

An experimental design technique such as Yates's algorithm (*9*) may prove helpful as a means of selecting parameter values for each set of experiments. Such techniques allow for investigation of the individual effects of each parameter as well as estimation of the effects between parameters while ensuring that all experiments necessary to evaluate a set of parameters are performed. For small numbers of parameters at a limited number of levels (two or three), all combinations may be tried. However, as the numbers of parameters and levels increase, the numbers of combinations increase dramatically. By using fractional factorial design, the number of experiments needed can be significantly reduced with only a minimal loss of information on parameter effects.

Establishment of Number of Simulation Replications and Variance Reduction

When used as a tool for comparing alternative designs, simulation-based methodologies rely on the development of statistical tests of significance differences among various measures of effectiveness (MOEs). In tests for significance, a confidence interval on the difference between the means must be estimated. Generally, outcomes with large confidence intervals can result in an inability, in a statistical sense, to distinguish between two designs. The ability to establish significance is rooted in the variance of the replications and is by nature proportional to the inverse of the number of runs.

The variances required for developing confidence intervals are estimated by drawing a number of point estimates of the mean by using randomly generated sequences of events. Each point estimate is a random sample from an unknown probability distribution of the MOE of interest and is used in the development of the associated confidence interval. As such, it is desirable to balance the number of point estimates or replications that are needed against the resources required to perform the simulations. If no limit were placed on the available resources and time, very large numbers of runs could be performed to obtain the smallest confidence interval possible. However, the number of replications required to improve a confidence interval becomes unrealistic beyond some point. To this end a vari-

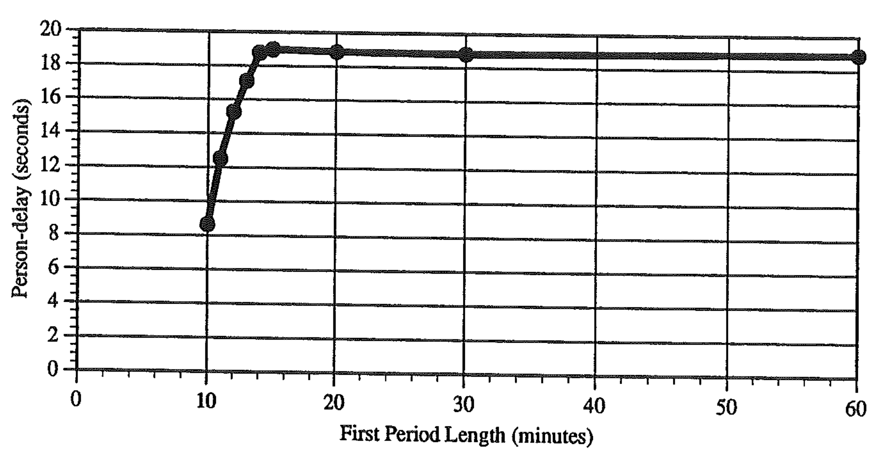


FIGURE 3 Second period person delay as function of first period duration.

ety of schemes have been devised to determine the required number of replications that balance the decrease in the confidence interval with the time required to complete the runs.

Techniques known as variance reduction methods (*7,8*) can be applied to improve the ability to detect statistical differences with smaller numbers of replications. Two methods are typically used: antithetic variates and common random numbers. Antithetic variates rely on using random number seeds that result in a negative correlation between successive observations for each design under consideration (i.e., if the first observation is above the true mean, the next observation will have a high probability of being below the true mean) thereby reducing the MOE variance for each design alternative. The common random numbers scheme reduces the variance in MOE differences by using identical random number seeds for the corresponding experiment for each design alternative. For evaluating transportation designs by using the TRAF family, Rathi and Venigalla (*8*) noted that the use of common random numbers generally led to a reduced variance and therefore an improved ability to detect statistical differences among alternative designs.

With regard to the number of replications required, a sequence of test runs should be performed to determine when the confidence intervals for one or more MOEs do not improve significantly. Typically, a minimum of 10 to a maximum of 30 replications may be required for each design alternative, depending on the model size and complexity. Figure 4 provides the confidence interval versus the numbers of runs for the case study. Ten replications were found to be sufficient.

Perform Simulation Experiments

By using the results from Steps 5c and 5d (selection of ranges for parameter values and establishment of the number of replications), a batch procedure can be established to perform and record the necessary output from each of the experimental runs that was performed. The batch procedure can be viewed as a series of embedded "do loops" to control and replace parameter values in the simulation input file, which rewrites the file and substitutes the replacement values when appropriate. A procedure should be used to record the order of the experiments so that the appropriate input-output data can be easily extracted for statistical comparison.

For a reasonably sized corridor, the output file generated by the TRAF family can be very large. As such it may be desirable to remove portions of the output file before archiving it to reduce the space required for its storage. For 10 replications of three alternative designs plus the current system, with three parameters at three levels, 360 output files need to be stored, requiring 180 or more megabytes of disk space. At a minimum the input information mirrored in the output files should be removed, at a substantial savings of disk space but without a loss of detail.

Evaluate Outcomes

The final step is to evaluate the results of the various experiments and to compare various alternatives. In reality this step may pose a new set of questions that require further investigation. Ultimately, a single design or set of designs may be recommended for further studies or design detailing.

Within the scope of the experiments and the outcomes analysis a wide variety of tests that can be used to evaluate virtually all aspects of the scenarios being modeled are available. Experimental design techniques, such as factorial analysis (*9*), can be applied to user-identified parameters to establish their contribution to the experimental outcomes. Statistical tests such as the paired *t*-test can be used to distinguish between competing designs and to determine whether a significant difference exists between the outcome MOEs.

The level of detail can vary greatly in an evaluation, from area-wide statistics for each replication to segment-by-segment details. Often, the latter is used early in the methodology to investigate behavioral abnormalities or to experiment with parameter values to improve model performance.

As an illustration the outcomes for the case study will be evaluated. Four design alternatives were initially selected to be modeled: (a) do nothing (baseline case), (b) convert an existing general-use lane to an HOV lane, (c) add an HOV lane, and (d) add a general-use lane. Figure 5 indicates the travel speeds experienced on the freeway for the four options at several hourly freeway entry volumes

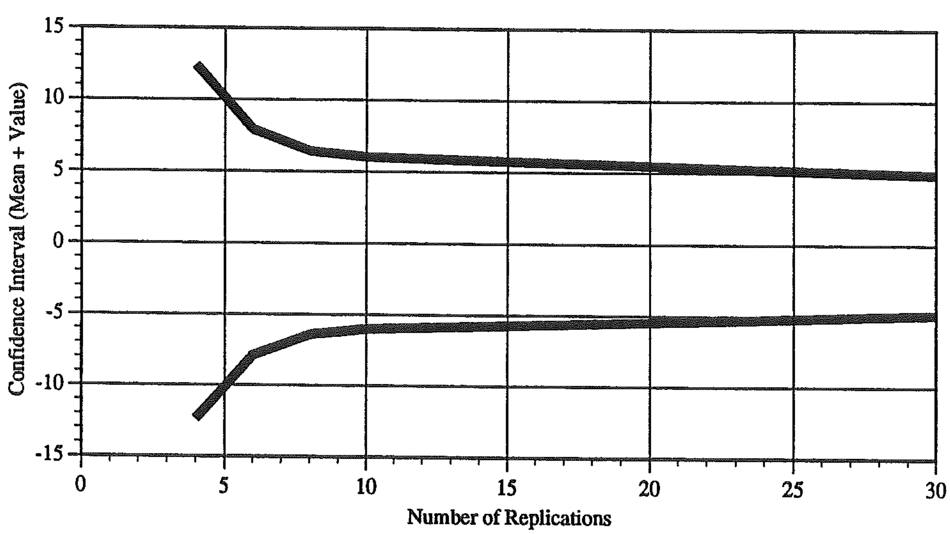


FIGURE 4 Variation of the confidence interval about the mean as a function of the number of replications.

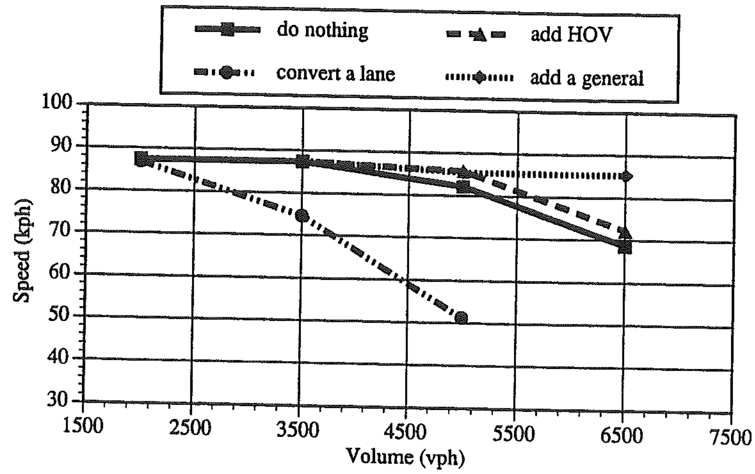


FIGURE 5 Facility speeds for the four alternative designs.

(2,000, 3,500, 5,000, and 6,500 vph). Although each design indicated a drop in the freeway speed as the volume increased, the option of converting a general-use lane to an HOV lane indicated a far greater degradation than the other three design options. Thus, a decision was made to eliminate the convert-a-lane option from further consideration.

HOV lane use was one of the variables investigated in the case study. Figures 6(*a*) and 6(*b*) indicate the variations in the effect of person trips for 10 and 15 percent HOV penetration levels. An increase in HOV penetration had only a subtle effect on the overall freeway speed. The most of the HOV effect was due to (a) the higher number of HOV vehicles traveling at or close to the free flow speed

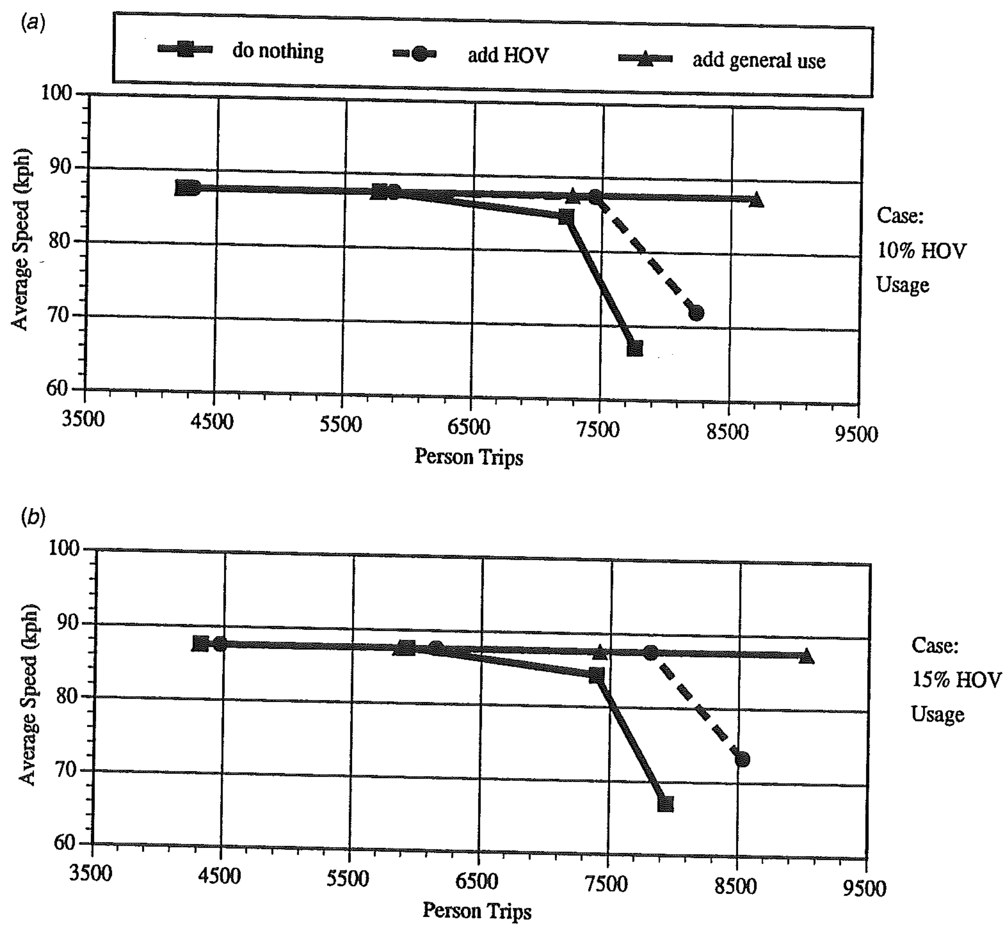


FIGURE 6 (a) Average freeway speed as a function of person trips at 10 percent HOV penetration. (b) Average freeway speed as a function of person trips at 15 percent HOV penetration.

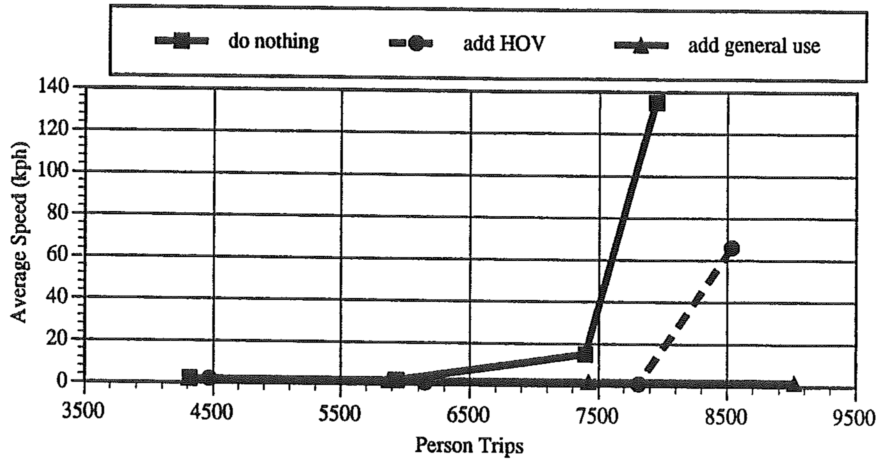


FIGURE 7 Person delay for increasing person-trips and for 10 percent HOV.

and (b) increased speeds in the general-use lanes. However, the speed increases in the general-use lanes were not statistically significant.

A second measure used to evaluate design performance was person delay (Figure 7). As would be expected, there are significant increases in person delay as the traffic on the roadway reaches and then exceeds capacity. For the existing conditions person delay rapidly increases at about 7,500 person trips. When a new general-use lane is added, no significant increase in the delay occurs for the traffic volumes studied. Some increases in delay were expected when the volumes reached about 8,000 person trips to match the increase for the do-nothing option, which occurred at 6,000 person trips, but no such change was observed. On examination of the individual freeway segments, it was observed that much of the delay occurs on downstream segments with high entering volumes, a situation that does not occur for the add-a-general-use lane option.

The comparison of the add-an-HOV-lane option with the add-a-general-use-lane option indicates both options appear to handle increases in numbers of person trips well, with the HOV option having a slightly lower delay until some of the freeway segments again approach oversaturation conditions. At between 8,000 and 8,500 person trips, the same segments noted earlier started to experience saturation, with an associated increase in delay for the add-an-HOV-lane option; this was not the case for the add-a-general-use-lane alternative. The slightly lower delay experienced under the add-an-HOV-lane option, when person trips are fewer than 8,000, can again be attributed to the lower volume and the associated higher speeds in the HOV lane. However, as before, this difference was not statistically significant.

CONCLUSION

Simulation can be a powerful tool for assessing design alternatives. Increasingly complex transportation systems require innovative approaches to assist planners and design engineers in making informed design decisions. Simulation models provide decision makers the opportunity to investigate transportation alternatives in realistic stochastic environments; deterministic models emulate only the average environment.

In this paper a simulation-based evaluation of HOV design alternatives has been proposed. Using simulation models, analysts and designers can assess and explore aspects of each design under different scenarios, much as if they were conducting field studies. However, simulation models allow them to explore and evaluate many what-if scenarios that would normally be infeasible with field studies.

Through the use of simulations within an experimental design context, powerful statistical tools can be used to compare alternative designs. The repeatability of simulated traffic streams leads to the application of variance reduction techniques that can be used to improve the ability to statistically differentiate among design alternatives with fewer simulation runs.

Undoubtedly, the use of simulation modeling for the evaluation of transportation alternatives is starting to be realistic and cost-effective. With increasing costs and concerns about environmental and social issues related to transportation, it can be expected that there will be an increased reliance on simulation models to evaluate transportation decisions.

ACKNOWLEDGMENTS

This research was supported by the Arizona Department of Transportation and the Maricopa Association of Governments through the direction of the Arizona Transportation Research Center. The authors acknowledge the active involvement and discussion of issues with the members of the Technical Advisory Committee of this research project.

REFERENCES

1. Turnbull, K. F., R. W. Stokes, and D. H. Henk. Current Practices in Evaluating Freeway HOV Facilities. In *Transportation Research Record 1299*, TRB, National Research Council, Washington, D.C., 1991, pp. 63–73.
2. Batz, T. M. High-Occupancy-Vehicle Treatments, Impacts, and Parameters: Procedures and Conclusions. In *Transportation Research Record 1181*, TRB, National Research Council, Washington, D.C., 1988, pp. 25–37.
3. Cohen, S. L., and J. Clark. Analysis of Freeway Reconstruction Alternatives Using Traffic Simulation. In *Transportation Research Record 1132*, TRB, National Research Council, Washington, D.C., 1987, pp. 8–13.

4. Lee Engineering. *1992 Study of Occupancy and Vehicle Classification in the Phoenix Metropolitan Area.* Maricopa Association of Governments, 1993.
5. Emerson, J. W., and S. G. Strickland. HOV Facilities: Status and National Perspective. *60th Annual Meeting, Institute of Transportation Engineers: Compendium of Technical Papers*, ITE, 1994, pp. 70–73.
6. JHK & Associates. *I-17/I-10 Corridor Study.* Arizona Department of Transportation Planning Division, 1986.
7. Kelton, W. D., and A. M. Law. *Simulation Modeling and Analysis*, 2nd ed. McGraw-Hill Book Co., New York, 1991.
8. Rathi, A. K., and M. M. Venigalla. Variance Reduction Applied to Urban Network Traffic Simulation. In *Transportation Research Record 1365*, TRB, National Research Council, Washington, D.C., 1992, pp. 133–146.
9. Montgomery, D. C. *Design and Analysis of Experiments*, 3rd ed. John Wiley and Sons, Inc., New York, 1991.

Publication of this paper sponsored by Committee on High-Occupancy-Vehicle Systems.