1991

Freeway Control Via Ramp Metering: Development of a Basic Building Block for an On-Ramp, Discrete, Stochastic, Mesoscopic, Simulation Model within a Contextual Systems Approach

Mohamed Yaser Alkadri
Portland State University

FREEWAY CONTROL VIA RAMP METERING: DEVELOPMENT OF A BASIC BUILDING BLOCK FOR AN ON-RAMP, DISCRETE, STOCHASTIC, MESOSCOPIC, SIMULATION MODEL WITHIN A CONTEXTUAL SYSTEMS APPROACH

by

MOHAMED YASER ALKADRI

A Dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
SYSTEMS SCIENCE

Portland State University

TO THE OFFICE OF GRADUATE STUDIES:

The members of the committee approve the dissertation of Mohamed Yaser Alkadri presented June 28, 1991.

Roy W. Koch, Chair

Wayne W. Wakeland

Kenneth J. Dueker

Warren Harrison

Eugene A. Enneking

APPROVED:

Roy W. Koch, Director, Systems Science Ph.D. Program

C. William Savery, Vice Provost for Graduate Studies and Research

AN ABSTRACT OF THE DISSERTATION OF Mohamed Yaser AlKadri for the Doctor of Philosophy in Systems Science presented on June 28, 1991.

Title: Freeway Control Via Ramp Metering: Development of a Basic Building Block for an On-Ramp, Discrete, Stochastic, Mesoscopic, Simulation Model Within a Contextual Systems Approach.

APPROVED BY THE MEMBERS OF THE DISSERTATION COMMITTEE:

[redacted]
Roy W. Koch, Chair

[redacted]
Wayne W. Wakeland

[redacted]
Kenneth J. Dueker

[redacted]
Warren Harrison

[redacted]
Eugene A. Enneking

One of the most effective measures of congestion control on freeways has been *ramp metering*, where vehicle entry to the freeway is regulated by traffic signals (meters). Meters are run with calibrated influx rates to prevent highway saturation.

However, recent observations of some metering sites in San Diego, CA indicate that metering, during peak hour demand, is helping freeway flow while sometimes creating considerable traffic back-ups on local streets, transferring congestion problems from the freeway to intersections. Metering problems stem largely from the difficulty of designing an integrated, dynamic metering scheme that responds not only to changing freeway conditions but also to fluctuating demand throughout the ramp network; a scheme whose objective is to maintain adequate freeway throughput *as well as* minimize disproportionate ramp delays and queue overspills onto surface streets.

Simulation modeling is a versatile, convenient, relatively inexpensive and safe systems analysis tool for evaluating alternative strategies to achieve the above objective. The objective of this research was to *establish a basic building block for a discrete system simulation model, ONRAMP, based on a stochastic, mesoscopic, queueing approach.* ONRAMP is for modeling entrance ramp geometry, vehicular generation, platooning and arrivals, queueing activities, meters and metering rates. The architecture of ONRAMP's molecular unit is designed in a fashion so that it can be, with some model calibration, duplicated for a number of ramps and, if necessary, integrated into some other larger freeway network models. SLAM.II simulation language is used for computer implementation. ONRAMP has been developed and partly validated using data from eight ramps at Interstate-8 in San Diego.

From a systems perspective, simulation will be short-sided and problem analysis is incomplete unless the other non-technical metering problems are explored and considered. These problems include the impacts

of signalizing entrance ramps on the vitality of adjacent intersections, landuse and development, "fair" geographic distribution of meters and metering rates throughout the freeway corridor, public acceptance and enforcement, and the role and influence of organizations in charge of decision making in this regard. Therefore, an outline of a contextual *systems approach* for problem analysis is suggested. Benefits and problems of freeway control via ramp metering, both operational short-term and strategic long-term, are discussed in two dimensions: *global* (freeway) and *local* (intersection). The results of a pilot study which includes interviews with field experts and law enforcement officials and a small motorist survey are presented.

To my grandfather
for his guiding spirit and exemplary scholarship!

To my parents
for their love and support throughout my years in college!

&
To my wife
for her love and comradeship during my doctoral studies!

Mohamed Yaser

Portland, Oregon U.S.A.
June 28, 1991

ACKNOWLEDGEMENTS

This research was supported by District 11 of California Department of Transportation (Caltrans) in San Diego. Thanks to Stewart Harvey, the deputy director for Traffic Systems/Operations who approved the idea; Joel Haven, the chief of Traffic Systems Branch at District 11 who acted as project administrator (at Caltrans); Arian Abrishami, Don Day and Susie Martin who provided valuable help in computer and electronic data.

I would like to acknowledge the support of Dr. Nancy Nihan, the director of Transportation Northwest (TransNow) and Dr. Kenneth Dueker, the director of the Center for Transportation Studies at Portland State University in providing a fellowship to match Caltrans contribution.

Special thanks to Dr. Adolf May from UC Berkeley for his comments and feedback; Dr. Kent Lall from the Department of Civil Engineering at Portland State for his help and support; and Wes Brenner from the Computer Consulting Services at Portland State for his help with SLAM.II

Finally, I am especially grateful to my dissertation advisers for their tremendous help, encouragement and the invaluable advice they gave to me; Dr. Roy Koch, the Committee Chairman and the director of the Systems Science Ph.D. Program; Dr. Wayne Wakeland from the Systems Science Ph.D. Program; Dr. Kenneth Dueker from the School of Urban Studies; Dr. Warren Harrison from the Department of Computer Science, and Dr. Eugene Enneking from the Department of Mathematics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

FREEWAY CONTROL VIA RAMP METERING: INTRODUCTION, EXECUTIVE SUMMARY, & FRAMEWORK OF RESEARCH

Freeway control via ramp metering (*) has become a widely accepted technique in the area of freeway management in the United States. Traffic experts consider it *one of the most effective means of control* of freeway traffic and congestion. The philosophy of ramp control is that entrance ramps are channels to a large canal -the freeway. And such channels must be controlled in order to prevent random disruption to the freeway flow, achieve balanced input-output in traffic mass (especially in the event of freeway incidents), and maximize freeway throughput during peak-hour demand.

The technique was first introduced in Chicago and later adopted in New York and Los Angeles, and eventually became a widely-accepted traffic engineering technique for freeway control nationwide. During the past two decades, highway departments across the country added ramp metering as an essential part of their traffic management systems. However, ramp metering is not free of problems. Field implementations have met varying degrees of success and some encountered operational

---

(*) To the casual observer, the word "freeway" means "freedom", and hence "freedom" and "control" seem to be mutually exclusive. To the traffic engineer, they are not. *Control of the operation* is necessary to preserve *a "relative" freedom for the flow.*

difficulties, public discontent, and differing organizational views and support. This research is a comprehensive systems study on freeway control via ramp metering with the following objectives:

1. Examine the theory, the state-of-the-art, the benefits, and problems of ramp metering by reviewing the literature, making some site observations at entrance ramps in San Diego, California, and conducting a pilot study and a survey in Portland, Oregon.
2. Collect and analyze traffic data at eight metered entrance ramps in San Diego to support the first objective and to establish a basic building block for a non-deterministic, discrete simulation model "ONRAMP" for modeling entrance ramps. ONRAMP, as an architectural molecular unit, could be used as a stand-alone unit or in future modeling projects as an entrance ramp block in an integrated freeway/ramp simulation model. The development of ONRAMP will help explain the ramp metering system better, and allow the traffic engineers to simulate various ramp layouts and traffic conditions in order to devise better operational plans and run the system with less problems.

REAL-WORLD PROBLEM

While ramp metering has been proven to be an effective method to improve flow on freeways, repeated field observations at some sites in San Diego, CA have indicated lately that ramp metering, during peak-hour demand, tends to help the freeway flow while creating traffic back-ups on local streets, thus transferring many congestion problems from the freeway to the nearby intersections. The state highway authority, Caltrans in this case, considers that unacceptable and is determined to

correct the problem. Support to this project has been provided by Caltrans within this context.

From a systems engineering point of view, the analytical techniques of the *Highway Capacity Manual* and other widely-used references are insufficient to investigate this problem for two reasons:

a. The manual's analytical methods are all deterministic. Not only does the manual offer very little in this regard, but also there are no probabilistic methods or stochastic techniques presented to deal with the uncertainty, randomness, and dynamic behavior of traffic, particularly the on-ramp vehicle arrivals and queueing activities.

b. The manual discusses traffic engineering points only. It does not address larger strategic issues that influence congestion or system planning problems which are created by signalizing entrance ramps. Furthermore, it does not provide any guidelines to study the socioeconomic costs/impacts that may result from the implementation of ramp metering.

To create a comprehensive approach & compensate for the second point of "systemic" deficiency, a non-reductionist *systems approach* for problem solving is suggested in this project. This research will begin in chapter two by reviewing the history and theory of ramp metering, analyzing its operational advantages from a traffic engineering point of view, and its strategic benefits from a transportation planning point of view. Possible problems that result from ramp metering, both operational (short-term) and strategic (long-term), will also be discussed. A detailed, multiple-perspective systems approach for understanding and examining the physical, economic, and societal impacts of freeway

control via ramp metering on the global (freeway) and local (intersection) system activities and interests is proposed. The results of a limited pilot study conducted by the author using this approach are also presented in chapter two. The pilot study includes personal interviews with transportation, traffic, planning and law enforcement key officials. It also includes a small motorist survey.

To compensate for the first point of "technical" deficiency, the use of *simulation modeling* is proposed. By simulating entrance ramps (ramp meters, light cycles, and traffic operations), alternative strategies to control and improve the freeway flow while preserving the integrity of the surface street system can inexpensively be simulated, tested, and evaluated as often as desired. Arguments for and against computer simulation modeling and building a new model are discussed in chapter two.

SYSTEMS MODELING PROBLEM

A question might be raised about the true need for a new modeling approach or a simulation model. The fact is that numerous computer models have been developed over the years in response to traffic flow modeling needs, including freeways with metered ramps. But those which do simulate on-ramp (like FREQ10) are macroscopic-deterministic. As will be discussed in chapter three, a good number of the rest suffer from severe limitations, inconsistency problems, and implementation difficulties. Therefore, there exists an opportunity for improvement.

Generally, models are either *macroscopic* or *microscopic*. Each type serves a particular need and is useful for certain applications, but

deficient (sometimes grossly) for others.

The macro approach views traffic as a mass of compressible laminar fluid. This view has been the basis of the *Highway Capacity Manual*. The individual microscopic entities (vehicles) and the random behavior of traffic are ignored. The macroscopic view is adequate for deterministic analysis of traffic supply and demand and corridor route assignment. But it is inadequate for investigating traffic operations that are vehicle-oriented or involve extensive real-time random or discrete traffic behavior.

The microscopic approach views traffic as a collection of micro-entities that move in accordance with physics principles of the linear motion. This approach has developed less rapidly than the macro approach and only little emphasis has been placed on the use of microscopic models by the industry because of their complex input requirements, complicated programming structure, long execution time, and the lack of robustness and comprehensiveness (despite their hugeness). For example, it was reported [41] that in some microscopic models, 5 seconds of computer time are required for each car for each one second of real time. Some more recent microscopic models still require overnight execution for each simulation run.

Several microscopic models have also reported validation and inconsistency problems. Only one microscopic model for freeway corridors (INTRAS family) is available for use today [51]. Numerous limitations, however, have been associated with INTRAS and CARSIM, which is another microscopic model based on INTRAS.

Many model limitations (both microscopic and macroscopic) are the

result of focusing on a one-sided or short-sided approach. Namely, they are either completely microscopic or completely macroscopic and they are developed and deployed without due consideration of implementation contexts. Model implementation problems also arise from their complexity, inflexibility, and detachment from the actuality of the real-world. Awareness of the real-world implementation environment and the unique characteristics of the studied system is absolutely necessary to overcome implementation problems.

In this research, a "systems approach" was deemed necessary. An intermediate "mesoscopic" modeling approach is that which is based on *a combination of microscopic system decomposition/analysis, mixed microscopic-macroscopic model building and verification, and macroscopic model validation and system representation.* For example, ramp arrivals and metering rates were analyzed and modeled in microscopic and stochastic details. Queue lengths and queue delays were viewed from a dynamic but semi-macroscopic perspective. Internal queue relationships between vehicles were considered to be less important microscopic details and thus were ignored. Average queue lengths and time delays were also evaluated from an overall macroscopic perspective. A complete discussion of mesoscopic modeling is given in chapter three.

THE REFERENCE BEHAVIOR MODE

This research has focused on data collected from a specific site in San Diego, California. In order to study the problem properly, a reference behavior mode (RBM) of the studied system (San Diego site) is established. RBM work encompassed field observations and video recording

of eight entrance ramps at a 5-mile segment along I-8 for the westbound morning traffic. Complete description of the site and data collection and analysis is given in chapter four. Specification for the model conceptualization, input, output, validation, and the validation criteria are also established and discussed in chapter four.

THE PROPOSED MODEL

The objective of the modeling work in chapter five is to *establish a basic building block of a discrete system simulation model, ONRAMP, which is based on a stochastic, mesoscopic, queueing approach.* The architecture of ONRAMP is defined as a *basic* building block because ONRAMP can be duplicated and used as is to model any *standard* entrance ramp. A standard entrance ramp means a loop or tangential on-ramp with 1-lane/1-server meter and no HOV lane. However, the basic block of ONRAMP can be easily modified to incorporate special ramp configurations or characteristics such as HOV lanes, etc. ONRAMP is built, tested, and validated based on the modeling approach and the reference behavior mode established in chapters three, four, and five.

ONRAMP is envisioned as the entrance ramp component of a larger ramp-mainlane freeway system model VQUE, *Virtual Queue*, which may be developed in the future. ONRAMP is developed and tested with VQUE in mind. Occasionally in this report, ONRAMP is referred to as phase I of VQUE. But *ONRAMP is not intended to be a final working simulator.*

The immediate application of ONRAMP is to model each ramp of a network of eight entrance ramps along Interstate-8 in San Diego, California. ONRAMP can model the geometric layout vehicular generation,

platooning, arrivals, queueing activities, and ramp meters and their rates (service times). ONRAMP block is capable of representing the following by adequately specifying its input parameters:

1. Platooned (clustered) and non-platooned vehicular arrivals.
2. Single or multiple ramp meters with static, dynamic, or random metering rates.
3. 1-car per green or 2-cars per green.
4. Single or multiple primary queues (queues on ramp lanes).
5. Primary queue blockage and single or multiple secondary queues (queue overspills on surface streets).
6. M number of time slices, each is "SILICSZ" minutes long.
7. K number of top-of-the-minute "TOTM" system status checks.

Eventually, ONRAMP could be further developed, expanded and integrated into the envisioned VQUE and used as a research/design instrument for designing operational strategies for ramp management and freeway control at any ramp-freeway system with or without meters.

SLAM.II (Simulation Language for Alternative Modeling), a widely-used, advanced simulation language, is used to implement ONRAMP on the computer.

RESEARCH TASKS & DISSEMINATION OF RESEARCH RESULTS

The research work was divided into four primary tasks. The objectives of these tasks were to:

Task 1: Review the theory of ramp metering and establish an overall systems approach for the research and field-test it in a pilot study (chapter II).

Task 2: Discuss the viability of computer modeling for investigating and solving ramp metering problems. Review existing traffic models & establish a modeling process to be implemented (chapter III).

Task 3: Establish the reference behavior mode of the studied system by collecting data and interpreting observation (chapter IV).

Task 4: Develop the basic building block of ONRAMP and test its validity (chapter V).

Chapter six is a short chapter which contains the final conclusions and some recommendations.

Besides the final dissertation submitted to the System Science Ph.D. program at Portland State University, research reports are submitted to the project sponsors CalTrans (California Department of Transportation), and TransNow (Transportation Northwest) and to the external project reviewer, professor James Banks from San Diego State University.

Further development of ONRAMP basic building block into a fully-operational VQUE simulator could be pursued as an extension to the project, if ONRAMP is adopted and the sponsors so desire, after the completion of the dissertation.

The prospective beneficiaries of the outcome of this project are the transportation industry, traffic engineering professionals, system analysts, and modeling and simulation practitioners.

---*---*---

CHAPTER II

REVIEW OF THE HISTORY, THEORY, AND OPERATIONAL/STRATEGIC BENEFITS AND IMPACTS OF FREEWAY CONTROL VIA RAMP METERING

CHAPTER PREVIEW

According to the U.S. Department of Transportation, the Federal Highway Administration, and the General Accounting Office [8 & 30], freeway congestion is increasingly a serious problem around the U.S. Numerous solutions have been proposed and many were tried. One of the most effective has been *ramp metering*. Federal Highway Administration experts concur that ramp metering is a proven way to maximize the capacity of freeway systems. Perhaps no other improvement can contribute to the safe and efficient movement of traffic as a carefully-designed and well-operated ramp metering system [12, 1989].

Despite the benefits of ramp metering, many problems come along with, or arise subsequent to the implementation of this technique. Not enough research has been done to study these problems. Some problems are operational, such as the difficulty to design an integrated, dynamic metering scheme that can respond to changing freeway conditions *as well as* shifting traffic demand throughout the ramps, to maintain adequate freeway throughput *and* prevent unnecessary or disproportionate ramp queueing and ramp delays. Other problems are less technical and have to do with: the impacts of metering on the vitality of adjacent intersections, land use implications, geographic distribution of meters

and "just" rationing of metering rates throughout a corridor (fairness issues), public acceptance, enforcement difficulties, and the objectives and political power of the organizations in charge of decision making.

The objective of this paper is to review of the history and theory of ramp metering, analyze the operational advantages of this technique from a traffic engineering point of view, and the strategic benefits from a transportation planning point of view. Problems that result from ramp metering, both operational short-term and strategic long-term, will also be discussed. A multiple-perspective, systems approach for understanding and examining the physical, economic, and societal impacts of freeway control via ramp metering on the global (freeway) and local (intersection) system activities and interests is proposed. The results of a limited pilot study conducted by the author using this approach are presented. Finally, questions for future research are suggested.

DEFINITION OF RAMP METERING

Ramp metering is a process whereby access to the freeway is controlled by a traffic light (a meter) at freeway entrances (on-ramps). Entry to the freeway is controlled and paced by the meters so that demand on the freeway is regulated and it does not exceed the freeway capacity. Metering rates normally allow one or two cars per green light in order to prevent platooned vehicular discharge. Meters are normally installed at a mid-point on the entrance ramp called "ramp metering limit line" or RMLL (figure 1). The ramp portion upstream of the RMLL is used as a storage for approaching vehicles. The ramp portion downstream of the RMLL is used as an acceleration lane for departing vehicles.

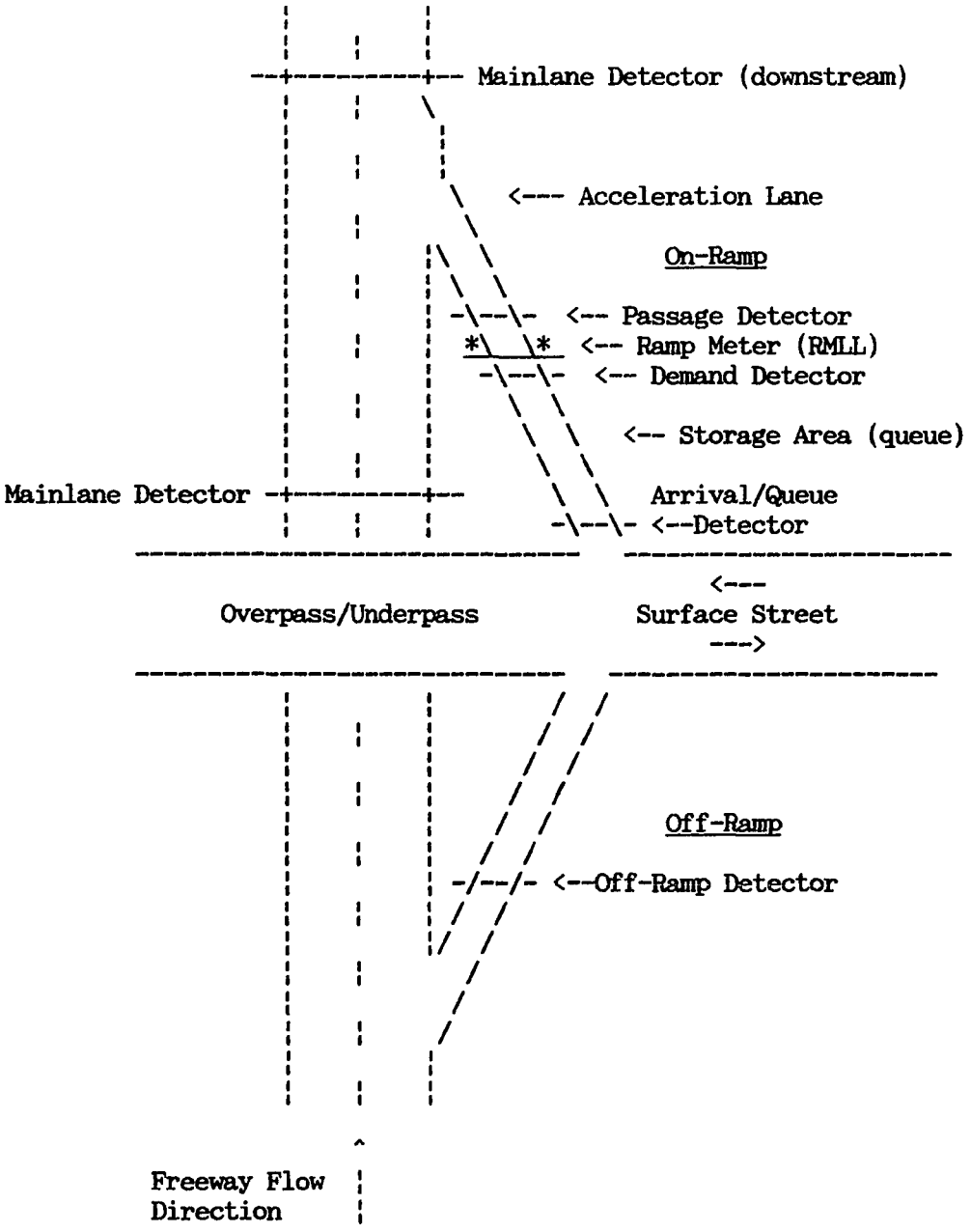


Figure 1. Typical On-Ramp, Mainlane, & Loop Detector Layout.

There are two types of ramp meter operation: fixed-rate and freeway traffic-responsive. In fixed-rate metering, ramp traffic is discharged according to a preset rate regardless of freeway conditions. In freeway traffic-responsive metering, rates are determined instantaneously according to freeway conditions. Mainlane traffic loop detectors (installed in the freeway pavement in the vicinity of the on-ramp as shown in figure 1) control the discharge rate via a central computer to prevent demand beyond the freeway capacity. Note that figure one is an outline of a standard "tangential" ramp. "Loop" ramps have the same metering structure.

RAMP METERING OVERVIEW: ORIGINATION & HISTORY OF RAMP CONTROL

Ramp metering started more as a temporary type of measure on many freeways until some other major freeway capacity improvement was made. Until the mid-seventies, new freeways were built before old ones reached points of saturation. Highway funds [state and federal] were plentiful and freeway space was available. Whenever the level of service in a freeway declined below "C", the freeway was either widened or a new one was built. Less attention was given to the subject of *efficiency*. But efficiency can no longer be ignored. Funding available for highway improvements is insufficient to keep pace with the demand for new freeways. Space is also less available, and the anti-freeway sentiment (environmental and otherwise) is growing. In many urban areas, capacity improvements are becoming impossible and, gradually, ramp metering is there to stay as a method of increasing the capacity by improving the efficiency of operations rather than enlarging the size of the freeway.

The ramp metering idea originated in Chicago in 1962 [29, 1989]. It began by the placement of a policeman at an entrance ramp of a heavily congested intersection at the Eisenhower freeway. The policeman would watch the freeway operation and allow the ramp traffic to enter the freeway at rates of whatever the freeway can accommodate by continuously observing available gaps on the freeway. The idea worked very well. Two years later, the policeman was replaced by a permanent fixed-cycle traffic light. However, the light was less effective than the policeman who was able to interact with the freeway.

The method of ramp metering was tried in other areas of the country. It has been a permanent fixture in the operating systems of many highway. For example, ramp metering in Los Angeles began in 1968 [28, 1989]. That system has been growing where there are over 900 meters currently operating in the L.A. metropolitan area -the largest system in the country. By 1989, ramp metering systems were in operation in over twenty metropolitan areas in north America (figure 2). These systems vary from a fixed-rate operation at a single ramp to complex, traffic-responsive, inter-connected, centrally-controlled, multi-meter systems along many miles of a freeway. The ramp metering system along I-8 in San Diego, CA is one such complex system. One evidence of the effectiveness of metering is the fact that, after years of testing and operation, every existing system has been or is proposed to be expanded [28].

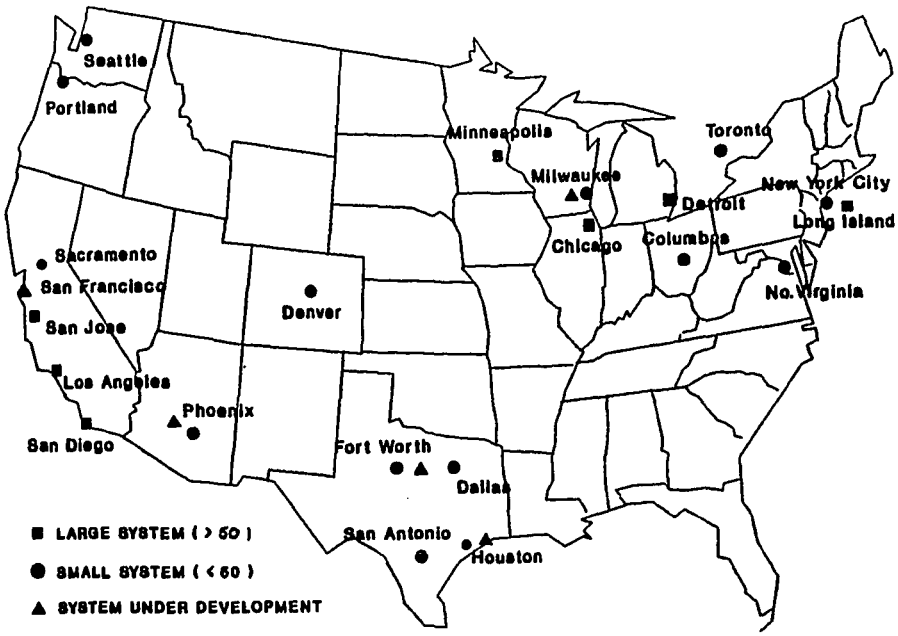


Figure 2. Ramp Metering Systems in the United States. Source: 27

THEORY OF RAMP CONTROL

The basic operational argument of the theory of ramp control contends that in order to minimize local congestions at the points where on-ramp lanes merge into the freeway mainlanes; to maintain smooth and continuous flow throughout the mainlanes; and to help keep the mainlane traffic running uniformly at high level of service (without bottlenecks, stop-and-go flow, recurrent delays during peak hours etc.), some measure of control over the influx from the entrance ramps is required. This kind of control is necessary not only to determine the vehicle *discharge rate* into the freeway, but also to regulate vehicle *discharge patterns*. Ramp meters are the ideal control tool to create these patterns and set these rates. The most effective ramp metering systems thus far are those

which are freeway traffic-responsive.

The theory of ramp control has evolved recently to incorporate strategic systems planning arguments and socioeconomic considerations besides the operational aspects. Traffic and transportation experts agree that there are many other systemic, societal and environmental benefits that might be realized by this technique. The more developed philosophy has two arguments:

1) Freeway management -a highway engineering operational argument.

2) Systems planning -a transportation engineering strategic argument.

1. FREEWAY MANAGEMENT: The Highway Engineering Operational Argument

Arguments for using ramp metering for managing freeway traffic are classified and discussed in the following section:

a. Control Freeway Saturation. There is no justification for allowing more cars into the freeway than what the freeway can accommodate. Assuming that the 1:4 delay:recovery rate is correct (i.e., one minute disruption on the freeway requires four minutes recovery time), and knowing that this rate is considerably more balanced at ramp queues [4, 1979], it is more efficient to hold vehicles at the entrance ramps when freeway capacity is all used up than to let these vehicles on freeway to queue there. Ramp meters can operate as gates that open and close to control the input and optimize the freeway throughput and prevent saturation.

b. Utilize Street Capacity. Surface and parallel streets frequently have excess capacity at the time when freeway is overloaded. Diverting some of the short-trip drivers from the freeway to surface

roads can alleviate pressure off the freeway. Caltrans study [4, 1979] outlined several benefits derived from steering short-trip drivers away from the freeway altogether. The freeway capacity is used suboptimally when used for short-trips during peak hours. Traffic flow is interrupted too frequently by short-trip users who increase the rate of merge/diverge and the rate of weaving. For those travelers, ramp metering is supposed to steer good portion of them to parallel surface streets instead.

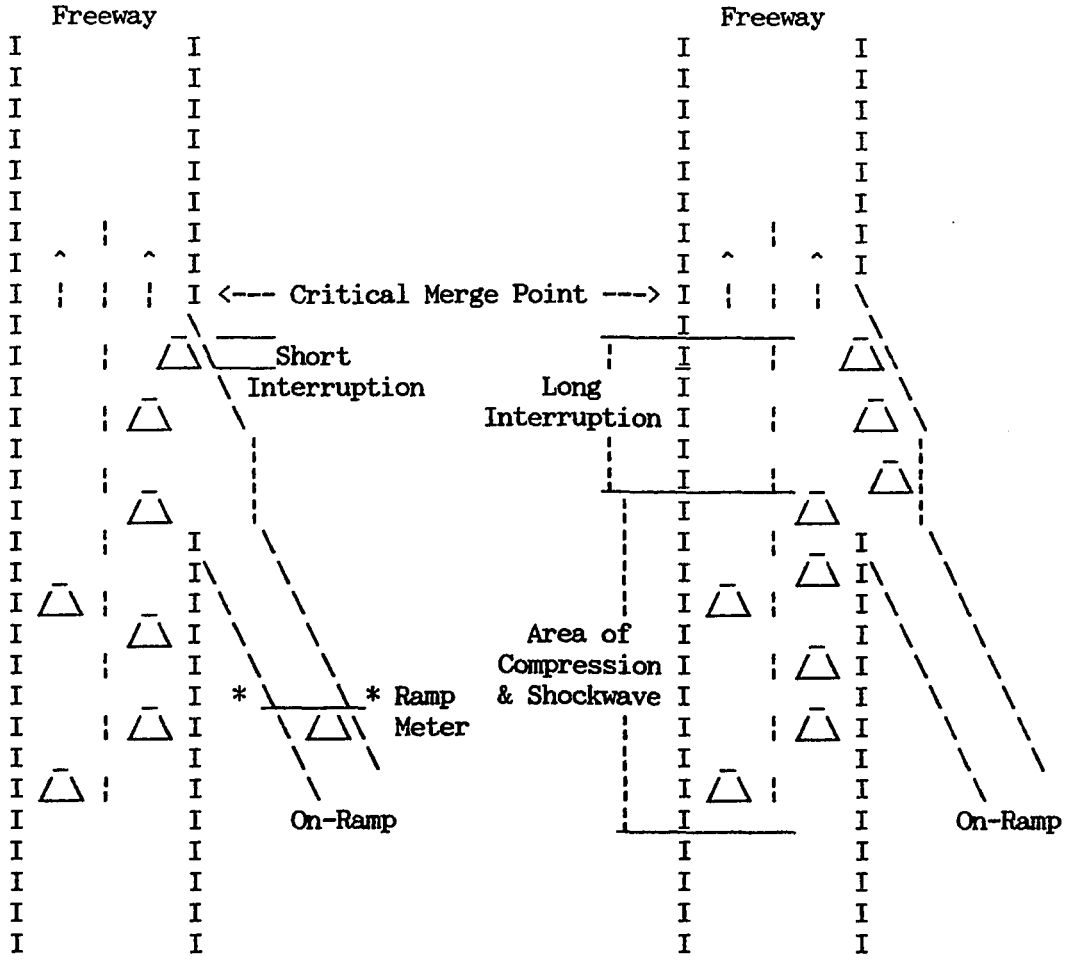
c. Disperse Peak Period Traffic. Ramp metering absorbs spikes or sudden surges in traffic demand by spreading the short peak-periods during which motorists arrive or leave their workplace, and dispersing them over a larger time span [2, 1981].

d. Redistribute Traffic Demand (Uniform Distribution). Uniform traffic density throughout the freeway subsections is highly desirable. A good example of what happens when uniform density is disturbed is the I-405 in Los Angeles. Uniform density is severely lacking throughout many subsections of that freeway resulting in succession of pockets of very high or low density. This leads to local recurrent congestion and high accident rates. Traffic operations experts aim at influencing the choice of route of the traveling motorists so that uniform use of the system capacity is achieved. An inter-connected ramp metering system can help redistribute non-uniform demand throughout the system. A Caltrans study [4] reported that ramp metering helped steer drivers around and away from heavily metered ramps. If metering rates are used properly, it may help divert street traffic wanting to use the freeway from bottleneck spots by tightening up the metering rates at ramps upstream

of the bottleneck area and relaxing them at ramps downstream of the strained section.

One problem with that study is that it assumes that motorists are always aware of the best alternative routes and would use them. That study, and others similar to it, did not discuss the need for a motorist information system to convey messages about freeway conditions, best alternative routes, expected highway delays and so forth. Such information need to be communicated to the motorist before he enters and gets trapped by the on-ramp queue. To insure effective redistribution of demand, a real-time, on-line motorist information system (e.g., changeable message signs, radio reports, etc.) may be very useful and necessary to work in conjunction with the ramp metering system.

e. Regulate Mainlane Merging. A one-by-one release of cars into the freeway provides evenly-spaced intervals between cars and eliminates clustering or platooning that occurs during peak-hours. When only a single car attempts to merge into the freeway mainlane, freeway traffic upstream of the merging point will have to only slightly decelerate to accommodate it. But when a chain of cars attempts to merge into the mainlane as a platoon, interruption time and space become long and mainlane traffic will be forced to slow down to lower speeds for longer intervals. This kind of forced deceleration severely damages the continuity of traffic flow because it results in amplified shockwaves propagating backwards. The worst effect occurs during peak hours when the density of freeway traffic is high. Rear-end collisions are reportedly caused by such interruptions. An illustration of platooned (non-metered) and non-platooned (metered) mergings is given in figure 3.



a) Controlled Ramp (metered)

B) Uncontrolled Ramp (not metered)

Figure 3. Short and Long Merge Interruptions at Controlled & Uncontrolled Entrance Ramps. a) Metered (Non-Platooned) Merges, b) Non-Metered (Platooned) Merges.

Note: Experimental simulation runs were made earlier in this project to study the merging phenomenon showed that "random" injection of ramp traffic into the freeway mainlane caused more severe disruption to the mainlane flow than "evenly-spaced" injection.

Acceleration (recovery) rates are usually much smaller than deceleration (slowdown) rates, particularly when they occur at higher (freeway) speeds. A Caltrans study [7, 1989] concluded that the recovery rate is 3-4 times less than the slowdown rate. It found that each minute saved in responding to an incident on the freeway saves four minutes of recovery time. The rate is greater than four for heavily congested freeways. Figure 3 also demonstrates how ramp metering minimizes the interruption time and space by breaking the platoons of merging vehicles.

f. Save on Total Trip Time and Improve Highway Travel Speeds. Field observations and research reports from around the country do indicate that, in general, ramp control reduces travel time and improves travel speeds. This is particularly true on heavily congested freeways. Such improvements are largely due to *improved operation* and not necessarily suppressed or reduced demand. As an example, figure 4 illustrates volume (throughput) improvements from "before" to "after" metering along a newly metered segment of I-5 (northbound Denver/Delta Ave. & Interstate Bridge) in Portland, Oregon [24, 1981].

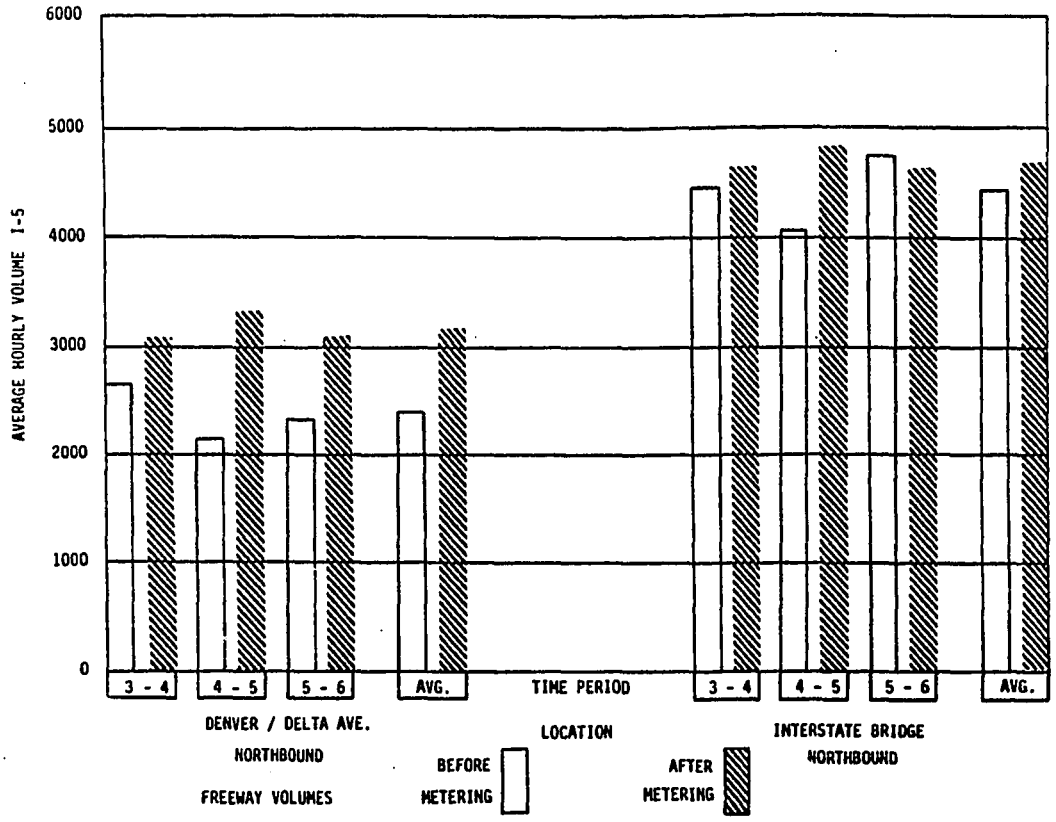


Figure 4. Before & After Average Hourly Volumes (an increase) on I-5, NB, Portland, OR. Source: 24.

Figure 5 shows a significant decrease in travel time on that same segment between the Broadway and Interstate Bridges. The average travel time between Broadway street and the Interstate Bridge (north bound on I-5) in the afternoon rush hour was reduced from about 22 minutes down to about 9 minutes (an improvement of nearly 60 % over a six-mile section). Both figures come from ODOT's *First Two Weeks of Operation -I-5 North Freeway Ramp Metering* [24, 1981], which is an evaluation report on ramp metering on that particular segment of I-5.

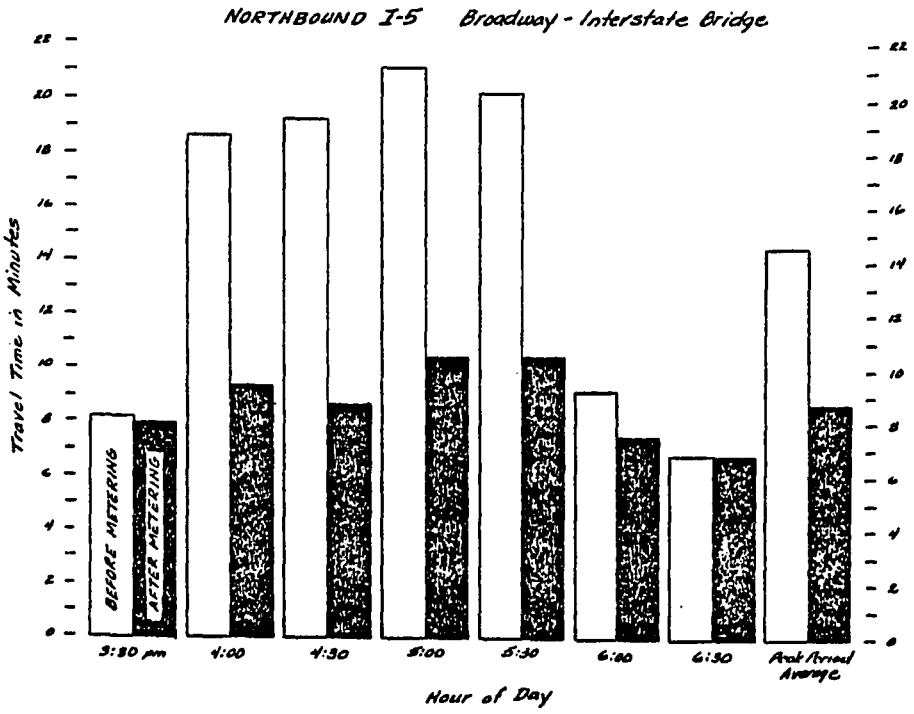


Figure 5. Before & After Travel Time in Minutes (a decrease) on I-5, NB, Portland, OR. Source: 24.

In Seattle, WA, a recently completed evaluation on the I-5 NB metering system shows that between 1981 and 1987, mainline volumes during the peak traffic periods increased 86% northbound and 62% southbound. Before the installation of metering, the travel time on a specific 6.9 mile course was measured at 22 minutes. After metering in 1987, the travel time for the same course was measured at 11.5 minutes [19, 1989] .

Carlson [12, 1989] presented their experience with local freeway traffic congestion in Minnesota and the method of ramp control:

> On I-35W, about a 17 mile section of a freeway, we had travel time of about 30 minutes before we started metering. After the metering was activated, we cut that to about 22 minutes. An evaluation of this project for I-35W after 10 years of operation showed that average peak period freeway speeds increased from 34 to 46 MPH (35% improvement). Over the same 10 year span, peak period volume increased 32%, the average number of peak period accidents declined 27%, and the peak period accident rate declined 38% [21, 1989].

In Denver, CO, the initial metering system consisted of five local traffic responsive metered ramps operated during the AM peak period on a 2.9 mile section of I-25 south of the city. Periodic post-installation evaluations revealed significant benefits. An 18 month post-installation study [8, 1989] showed that average peak period driving speed increased 57% and average travel times decreased 37%. Stop and go conditions on the section were generally eliminated.

In Detroit, MI, an evaluation performed by Michigan State University for MDOT [16, 1988] found that ramp metering increased speeds on I-94 by about 8%, which is not very significant, but still considered an improvement. At the same time, the typical peak hour volume on the three eastbound lanes increased to 6400 vehicles per hour from an average of 5600 VPH before metering. In addition, the total number of accidents was reduced by nearly 50% and injury accidents were down 71%.

In Austin, TX, metering along I-35 NB resulted in an increased vehicle throughput of 7.9% and an increase in average peak period mainline speeds of 60% through the section [18, 1981].

In Long Island, NY, metering along the Long Island Expressway resulted in good savings in travel time and gains in speed. An analysis of the initial metered segment after 2 months of operation in the PM peak [15, 1989], shows a 20% decrease in mainline travel time (from 26

to 21 minutes) and a 16% increase in average speed (from 29 to 35 mph). Motorists entering at metered ramps also experienced an overall travel time reduction of 13.1% and an increase in average speed from 23 to 28 mph.

g. Improve Public Safety. Ramp control is credited with significant contribution to the reduction in the number of traffic accidents on freeways. At every freeway spot where incidents are frequent, it has been helpful to use entrance ramp control [12, 1989]. The 1981 Oregon DOT report [23] stated that ramp metering has been proven, statistically, to provide better safety to the freeway travelers:

> Ramp metering will improve safety and reduce the overall number of accidents on the freeway. By reducing rush-hour congestion and stop-and-go traffic, freeway traffic will be safer. On I-5 northbound for example, reports indicate that traffic accidents have been reduced by 50 % after the 1981 installment of ramp meters.

Later reports [26, 1982] indicated that , overall, there was a 43% reduction in peak period traffic accidents.

Carlson [12] described how ramp metering technology contributed to the safety of the traveling motorists:

> Our evaluation studies have shown that ramp metering is very effective in terms of improving freeway safety. Our large system on I-35W has resulted in a reduction of 120 accidents/year -a reduction of accidents rate by about 38%. We have conducted many studies as we have implemented projects over the years. Taking a look at the before and after conditions as far as accidents were concerned. In every study, there was a significant reduction in the number of accidents associated with ramp metering.

At some Colorado highways, the incidence of rear-end and side-swipe accidents declined 5% [8, 1989]. Lipp [12] observed that metering led to the reduction of *certain kinds* of accidents:

> We found in Denver that we did trade off accidents from more severe to a less severe type accidents. By allowing the vehicles to emerge more efficiently on the freeway system, we eliminate a lot of injury rear-end type accidents.

An unexpected, unintended post-installation assessment of the system performance in Denver took place one day in the spring of 1987. To switch to day light savings time, all of the individual ramp controllers were adjusted one hour ahead. However, the central computer clock was overlooked. The central computer overrode the local controllers and metering began one hour late. As a result, traffic congestion was the worse it had been in years. This oversight turned to be an excellent test of the effectiveness of the ramp metering system. It is reported that the media became more supportive of metering since that incident [8].

Table I summarizes percentages of before/after improvements in traffic volumes (throughput), travel speed, travel time and incident rates during peak-period hours at seven locations in the United States.

TABLE I

PEAK-PERIOD PERCENT IMPROVEMENTS IN TRAFFIC VOLUMES,
TRAVEL SPEEDS, AND ACCIDENT RATES
AS A RESULT OF RAMP METERING
(Before and After Studies)

City, State	Freeway/ Location	% Traffic Volume	% Travel Speed	% Travel Time	% Accident Rate	Year Installed	Study Duration
Portland, OR	I-5 NB	NA	256	60	43	1981	14 months
Minneapolis, MN	I-35 W	32	35	27	38	1974	10 years
Seattle, WA	I-5 NB	86	NA	48	39	1981	6 years
Denver, CO	I-25NB	NA	57	37	5	1981	18 months
Detroit, MI	I-94EB	14	8	NA	50	1988	NA
Austin, TX	I-35NB	8	60	NA	NA	NA	NA
Long Island, NY	LIE	NA	16	20	NA	1988	2 months
San Diego, CA	(Studies are not yet available!)						

Notes: Volume and speed figures are percent increases.
Travel time and accident rates are percent decreases.
NA: Not Available.

h. Assign Proper Traffic Priority. It is sometimes argued that freeway traffic serves *functions of higher priority* than street traffic. Although this theory may be somewhat biased, it has many supporters. Freeway enthusiasts (e.g., state traffic engineers, heavy users of the interstate system, trucking industry, highway and oil industry etc.) would favor highway interest (including right-of-way) over surface street system and local traffic interests.

Many also contend that freeway traffic should have a priority since it was there first. The first in first out (FIFO) principle, should be upheld. Those motorists who are already on the freeway are assumed to have already been metered (taxed) at an upstream entrance.

Another priority argument is that when a vehicle merges into the freeway, it creates a situation where one vehicle (minority) interrupts

many (majority). Priority must be given to the *majority* and the single vehicle must be *regulated* (via a ramp meter).

The safety argument (discussed before) further argues in favor of freeway traffic. Every time a car merges into the freeway, driver's *judgement* is invoked. If the merging point is short, competition for the available spot may be tense. The confusion will be multiplied if a whole platoon is merging, and so the likelihood of incidents increases. Ramp meters minimize competition for merging spots giving some preference to mainlane traffic.

2. SYSTEMS PLANNING: The Transportation Engineering Strategic Argument

Benefits of ramp metering are not limited to the immediate enhancement of freeway flow and traffic safety. As will be shown below, traffic systems planners hope to rely on ramp metering as a strategy to achieve long-term planning objectives as well.

a. Influence Modal Split and Modal Choice. Ramp metering can be used as a method to manipulate the traveler's choice of transportation modes. Through the use of exclusive high occupancy vehicle (HOV) lanes, carpools and mass transit receive preferential treatment. HOV traffic is normally allowed to bypasses traffic queued at single occupancy vehicle (SOV) lanes. Over time, this strategy could reduce the system demand by increasing the number of passengers per vehicle and thus decreasing the overall number of vehicles in the system. This is "hopped" to happen as more motorists try to carpool or switch to mass transit to avoid long delays at the meters. An Oregon study [23, 1981] estimates that, in addition to savings in individual vehicle operating costs and

environmental gains, carpoolers and mass transit users could save up to 90% of the waiting time at the meters.

b. Efficient Infrastructure & Transportation Pricing Policy. One dilemma that faces the transportation planner is the fact that capacity improvements, while they reduce congestion, induce more use of the facility. As figure 6 shows, capacity improvements invite more use of the facility, which results in more congestion, which requires new capacity improvements, which again leads to increased use of the facility and so on. This loop is self-amplifying (with positive feedback at all the nodes) unless the cycle is broken or weakened. One perceived role for ramp metering is to help "contain" demand by inflicting some penalty on it (in terms of delays). Eliminating excess demand which is encouraged by non-metered access to the freeway will allow a better and more "efficient" use of the "existing" capacity.

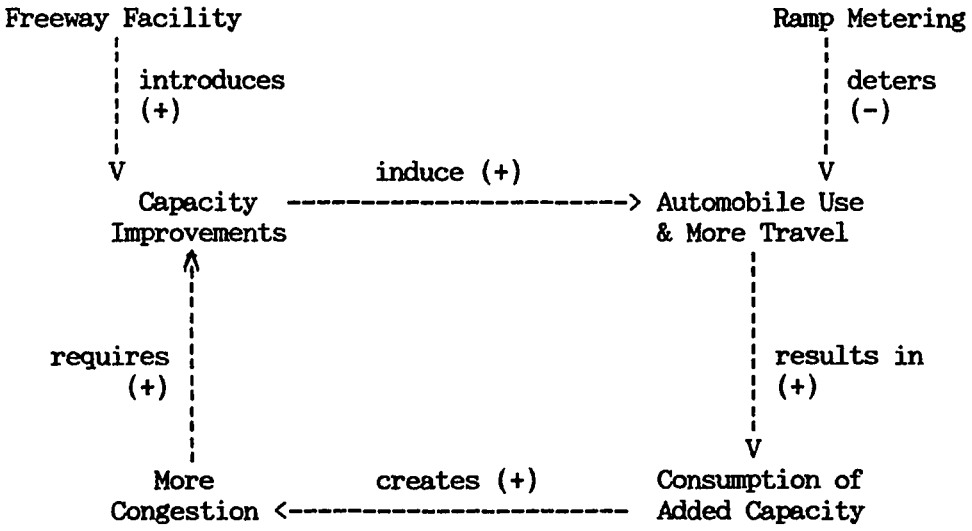


Figure 6. Vicious Cycle of Capacity Improvement and the Role of Ramp Metering in Reducing Demand on the Freeway.

Winston [32, 1991] goes further and points to the fact that users of the infrastructure (the freeway in our case) incur certain costs when they use the facility by contributing to congestion, increased travel time, wear and tear and maintenance expenditures to repair pavements and vehicles. In his view, an efficient infrastructure policy must aim at reducing such costs while increasing benefits to society. He goes as far as proposing a "toll" to help finance maintenance and deter unnecessary use of the facility, thus prolonging its lifetime. It is evident that transportation pricing goes beyond freeway management and would necessarily involve long-range planning and land use considerations. Considerations that are debatable less well-defined and more difficult to measure. The true long-term effect of ramp metering on land use and similar issues is still unclear. More research needs to be done in this area.

c. Increase Energy Conservation and Reduce Pollution. Ramp metering is a measure that can result in energy conservation. When energy is conserved, both the economy and the environment are served. Smooth and continuous movement on the freeway allows maximum utility of burned fuel. Less energy is wasted since vehicle deceleration and acceleration are reduced. The higher freeway capacity generated by the metering strategy allows more long-trip city traffic to either use the freeway more frequently or move more freely with less traffic on city streets. Finally, regardless of acceleration/deceleration, since vehicles are spending less overall time traveling, the amount of burned fuel is reduced accordingly.

Oregon DOT [23] predicted that substantial amount of energy will be saved once the ramps on I-5 in Portland become operational:

> It is estimated that approximately 624,000 gallons of fuel are used annually by motorists on I-5. With ramp metering, fuel consumption will be reduced by 29% a savings of 700 gallons per day.

Ramp metering may also help the environment. If the energy conservation argument holds, less fuel is burned and, consequently, less carbon dioxide, less lead, and less smog is released into the air. Large-scale evaluation efforts are being undertaken (especially in New York) to study these types of benefits and impacts [28].

d. Identify System Deficiencies. Finally, if a good metering system is in place and the system is still experiencing non-uniform traffic distribution and/or persistent bottlenecks, it may be time for other traffic management system (TMS) measures to be introduced. Once the system operation is optimized, system capacity alternatives (widening, new construction projects etc.) can then be seriously considered. Addition of new facilities or the expansion of existing facilities at certain spots would then be more justified [2, 1981].

PROBLEMS WITH RAMP METERING

1. Operational Problems

Ramp metering is not an insurance policy technology against all freeway flow problems. It may serve very well in one area, and serve poorly at another or may have no visible effect either way. The 1979 Caltrans study [4] attests to this fact:

> Various California ramp control projects have slightly increased capacity, slightly decreased capacity, or had no effect on

capacity. This is a function of roadway geometrics and traffic characteristics.

This conclusion was also evident for the case of southbound traffic on I-5 for the morning period in Portland, Oregon. Ramp meters there seem to have produced mixed results at some locations along the I-5 southbound direction. While figure 4 showed significant "after metering" volume improvements at Denver/Delta Av., figure 7 shows that the average hourly volume "AVG" did not change after the ramp meters were installed at the Interstate Bridge. Furthermore, the average hourly volume had actually decreased at the Interstate Bridge and at Lombard Undercrossing from 7:00 to 9:00 a.m., a traffic peak-hour! Oregon DOT experts did not give a satisfactory explanation, but this was attributed mainly to the fact that traffic was already moving at a good level of service prior to the installation of ramp meters [12 & 28]. The author believes that ramp metering warrant item #4, "a freeway recurring congestion," (as will be seen later in the ramp metering warrant section) was not actually satisfied. Evidently, metering works well only during peak-hour traffic. As the traffic gets lighter, meters disrupt an otherwise light and smooth traffic stream.

Performance statistics such as those presented in the previous section are not comparable throughout all locations. Also, measures of effectiveness vary depending on the objectives of the system [28].

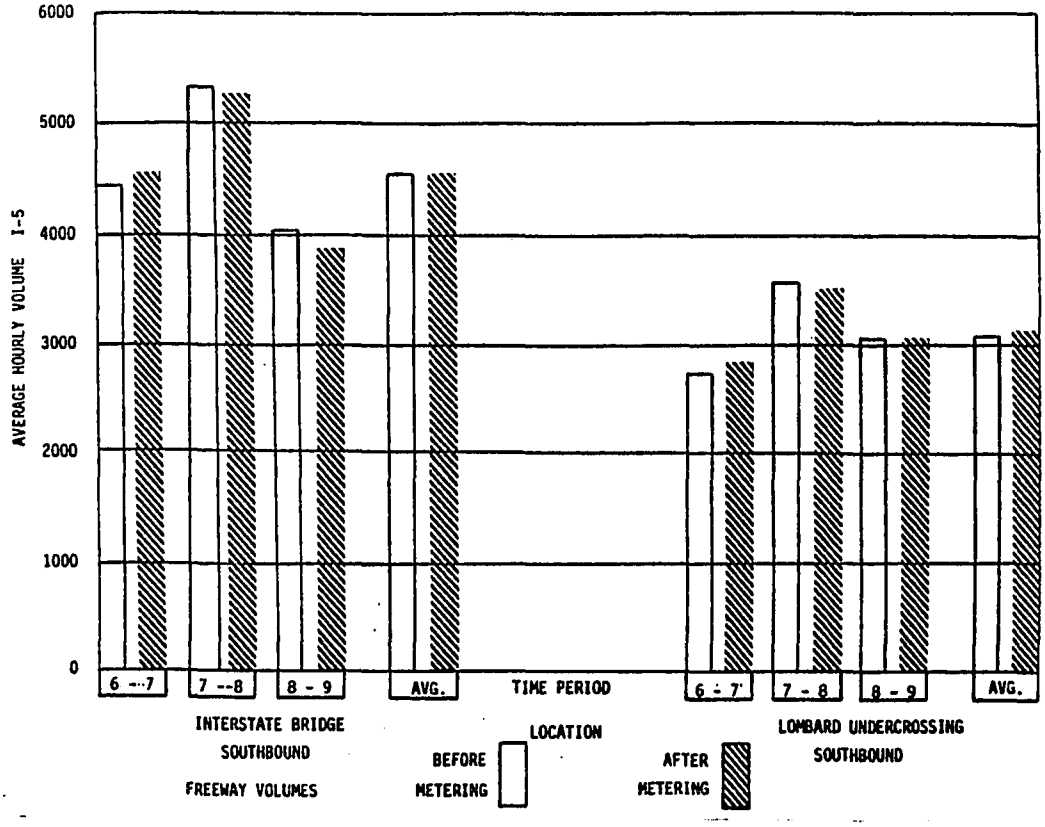


Figure 7. Before & After Average Hourly Volumes (a decrease) on I-5, NB, Portland, OR. Source: 24.

The worst impact of ramp metering is the adverse effect on local streets due to extended queueing. Caltrans study [4] examined the problem and stated that:

> There are situations where the wait at an on-ramp, required to maintain free flow conditions on the freeway, would overflow available storage space and cause city street problems. If this cannot be relieved, freeway flow would have to be sacrificed in order to maintain the integrity of the city street system. Most operational problems caused by ramp meters "backing-up" cars onto the city streets are observed during the first few days of metering. They are usually solved during the adjustment period by the shifting of users travel routes to a more efficient use of the freeway corridor. Some problems with city streets interference are eliminated by adjusting the metering rates at several ramps to better fit the changing traffic patterns.

There are two problems with the above analysis. First, the study does not mention what will or should happen once all ramp meters are operational, metered correctly and problems continue to exist. Second, the study itself is old. Recent studies on ramp metering have not addressed this point adequately. A more recent Caltrans study [6, 1988] came up with similar conclusions to the 1979 study. The new study stated that:

> Ramp metering transfers delay from the freeway to the ramps with some overall reduction in delay due to time and route redistribution of ramp traffic and results in an overall accident reduction. A limitation to ramp metering strategies is the availability of sufficient storage on ramps to accommodate the length of queues necessary for effective metering. Queues from ramp meter installations that block the normal flow of traffic on adjacent city streets are unacceptable. The only recourse is either to obtain more storage or to increase the ramp metering rate to the detriment of the freeway, thus increasing delay.

Nearly every ramp control article in the literature has concluded that, despite some "minor" side-effects, the method is sound and effective. None could be found addressing profound metering problems or dwelling on long-term, potential, and negative impacts of metering such as on the vitality of adjacent intersections or the impacts on land development.

Furthermore, a number of studies have been conducted to examine the technical (operational) impacts of ramp metering at the global (freeway) level. But very few have studied the local (intersection) in any details. Technical research at the local level would have to focus on the diversion of traffic, local congestion, obstruction of surface streets, awkward traffic patterns.

Ramp metering problems will proliferate as traffic demand grows in

many areas. At many location in San Diego, for instance, it is not unusual to have delays of five, ten, or even fifteen minutes at entrance ramps with queues of over seventy vehicles in length during the morning or the afternoon peak-hour.

Figure 8 is an illustration of a typical morning peak-hour intersection problems at the 70th Street and Lake Murray Boulevard intersection, a particularly complex intersection. Certain activities, phenomena, violations, and driver behavior that result from traffic backup at the entrance ramp(s) are highlighted and numbered from 1 to 13 on figure 8. The highlighted activities are described below. Evidently, this example shows that at one selected ramp, at least thirteen types of problems were occurring every time the ramp queue backs up. Although most problems seem operational and temporal, they are expected to produce long-term impacts on business, housing areas, landscape, and landuse and development near the intersection.

Highlighted Activities on Figure 8

(1) Blocked entrance to residential side street.
(2) Blocked entrance to gas station.
(3) Blocked road access from 70th St. to Alvorado Rd.
(4) Trajectory of vehicle bypassing ramp queue tresspassing through the gas station.
(5) Useless street traffic signal due to backed up ramp queue.
(6) Unsafe exposure of on-ramp traffic to fast-approaching, tree-blinded off-ramp traffic.
(7) Blocked access from Parkway Dr. (single lane road) to Murray.
(8) Right-of-way for "long" queue on Parkway Dr over off-ramp traffic.
(9A) Blocked entrance to an apartment complex from 70 NB approach.
(9B) Blocked entrance to an apartment complex from 70 SB approach.
(10) Illegal U-turn to avoid ramp queue.
(11) Trajectory of car going through business lot to break into the middle of the queue.
(12) Trajectory of car going through Denny's lot to break into queue and use of Denny's parking lot by non-customers for traffic maneuvers.
(13) Extreme corner traffic congestion.

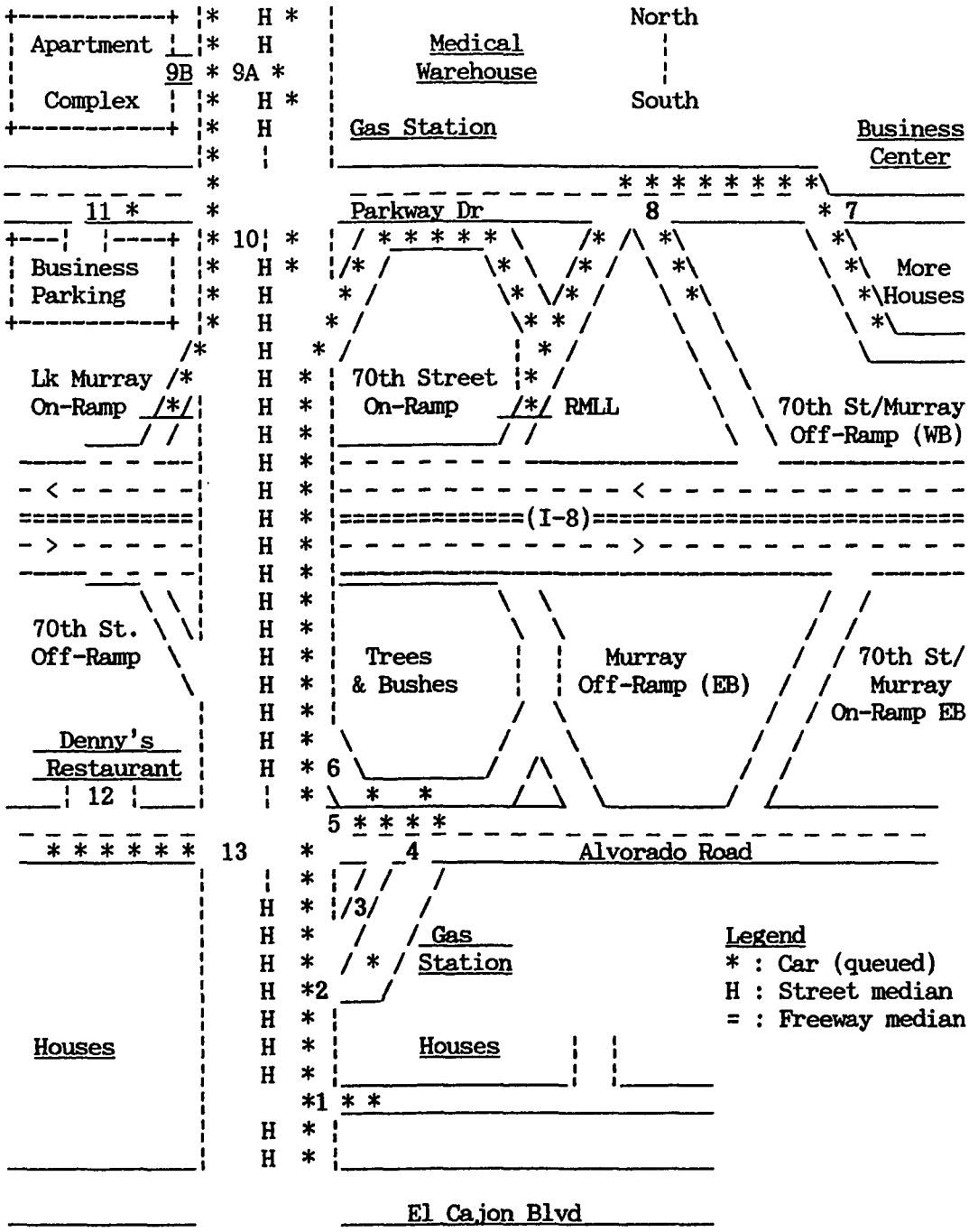


Figure 8. Peak-Hour Trouble Spots at 70th Street & Lake Murray Blvd. Along Interstate 8 in San Diego, California. (8/1989 observations).

2. Strategic Planning, Landuse, and Socio-Economic Problems

Many factors affect land use patterns in the concentrations around freeway access roads and intersections. Factors include urban trends, characteristics of the transportation system, historical factors, and timing of developments [1, 1982].

Certain facilities are particularly vulnerable and more likely to be impacted by traffic activities near the intersections. Baerwald [1, 1982] categorized and ranked those facilities most likely to locate in central urban clusters and those most likely to locate alongside freeways and their interchanges/intersections. Figure 9, expanded after Baerwald, lists types of activities which are likely to locate in central urban clusters, and those which are likely to locate alongside freeways. The later are more likely to be impacted by ramp metering.

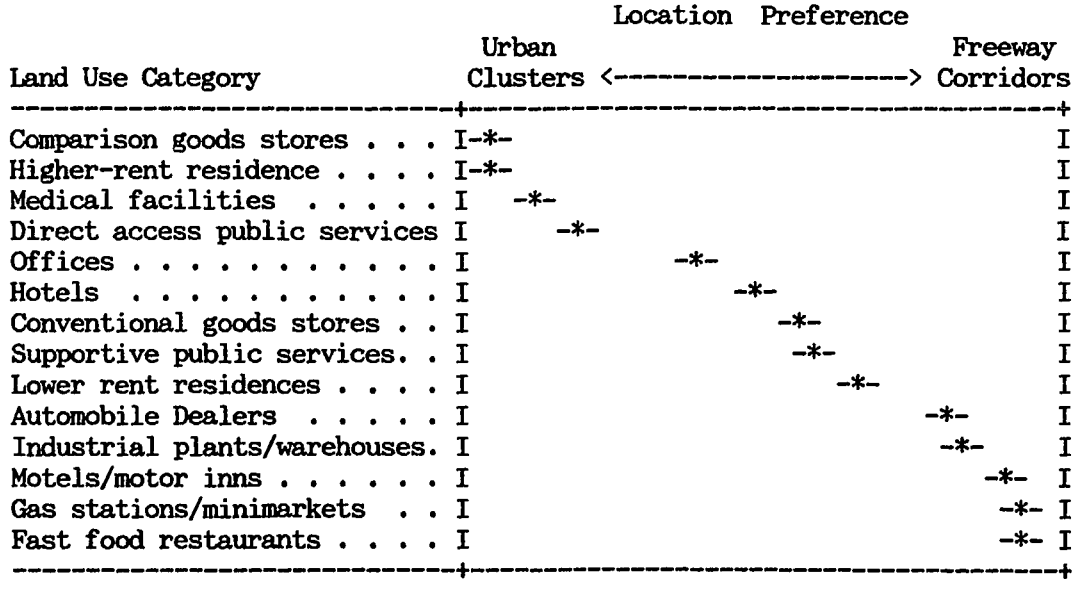


Figure 9. Land Use Preferred Location (expanded after Baerwald [1]).

Finally, time saving, safety, environmental benefits, and an improved overall freeway system performance, are not always evident to individual system users. The motorist is more cognizant of the costs (e.g., queueing delays, vehicle wear and tear, inhaled fumes, increased level of stress, etc.) than he is appreciative of the operational and overall societal advantages of freeway control. He is reluctant to pay an *individual* price for an improved *collective* service.

Questions are raised whether it is justifiable to favor the welfare of the corridor community and system users as a whole at the cost causing some injustice to intersection traffic and nearby inhabitants. Residents in close proximity of ramp meters, for example, are affected disproportionately (in terms of adverse impacts on land development, restricted access, etc.).

Unfortunately, there is very little research on the non-operational, long-term and socioeconomic impacts of ramp metering. The next section examines some suggested theoretical approaches to study overall impacts and the feasibility of freeway control via ramp metering. A systems approach designed by the author will be presented in subsequent sections.

WARRANTS FOR RAMP CONTROL

Despite all the benefits of ramp metering, its use is not always warranted because of its potential negative impacts. The 1978 *Interim Warrants for Freeway Entrance Ramp Control Signals* [2, 1981] provides an operational criterion for metering entrance ramps. According to those warrants, use of ramp meters is recommended when:

1. The expected reduction in freeway delays exceeds the expected delay to ramp users (a); AND
2. There is adequate storage space for delayed vehicles at the entrance ramps; AND
3. There are suitable alternate surface routes (b); AND
4. There is a recurring congestion on the freeway due to the traffic demand in excess of the freeway capacity; OR; there is a recurring congestion and severe accident hazard near the freeway entrance ramp because of inadequate ramp merging area.

THE NEED FOR A SYSTEMS APPROACH AND A MULTIPLE PERSPECTIVE

The above are *operational* considerations. Since there are other less well-defined strategic and socioeconomic aspects to consider besides the operational ones, justification for the installment of a ramp control system at one location and its operation at certain times remains open to discussion and debate among urban planners, transportation experts, traffic engineers, and law enforcement agencies.

(a) Extra travel time for diverted traffic to alternative surface routes is not mentioned in the Warrants and is difficult to measure.
(b) Alternate routes must have a capacity to absorb diverted traffic.

Blumentritt [2, 1981] suggests a more comprehensive "systems approach" to be used for studying *prospective* impacts of ramp metering in order to enhance the decision making process. He recommends that:

> A consideration of the suggested warrants defines several types of studies that may be necessary to assemble the basic information needed to determine the feasibility of entrance ramp control. These studies can be defined as follows:
>
> 1. Bottleneck analysis (location, demand analysis, capacity analysis, and metering rate).
> 2. Geometric analysis (ramp storage, merging areas).
> 3. Traffic diversion analysis (diversion estimates, diversion routes, diversion impacts).
> 4. Accident analysis.
> 5. Enforcement analysis.
> 6. Public acceptance analysis.
> 7. Preliminary cost considerations and cost-effectiveness analysis.
>
> The completion of the foregoing studies in a thorough manner will provide the decision-maker with a solid data base to use in determining the feasibility of entrance ramp control.

The ability to examine ramp metering problems from several perspectives gives the system practitioner more power in decision making. Issues that involve multiple perspectives for decision making have certain common features that distinguish them. According to Linstone [17, 1984], they have in common the following characteristics:

* Ill-structured nature of problem (typically sociotechnical systems);
* Significant policy and/or decision analysis content;

* Significant human aspects (societal or individual).

Transportation systems can be categorized as sociotechnical systems where there are endless and complex interactions between man, the machine (the car), the road, and the land. Three principal players seem to influence the decision making process: the transportation technology, the transportation organization(s), and the individual motorist (the system user). Linstone insists that, in order to enhance the decision making process, the perspectives of all parties must be considered. Our proposed approach to investigate the impacts of an existing ramp metering system or the feasibility of a proposed one takes linstone's view into consideration.

A PROPOSED APPROACH TO INVESTIGATE THE IMPACTS/FEASIBILITY OF FREEWAY CONTROL VIA RAMP METERING

1. Problem Identification

The process of problem analysis and decision making for freeway control via ramp metering could be enhanced significantly if three things are recognized by system analysts and decision makers while conducting the impact/feasibility studies:

1) Two areas of impact.

2) Three impact categories.

3) Three viewing perspectives.

First, the problem of ramp metering appears to be a dichotomous one. On one side, there is the *global* freeway activity and its interests, viewpoints, and patrons. On the other side, there is the *local* on-ramp intersection activity and its interests, arguments and often grievances.

Second, it has been suggested in the literature [30, 1979] that there are three kinds of impacts that an existing or proposed transportation facility has on the urban activity system: *physical*, *economic*, and *social*. Table II has been constructed to present a list of likely impacts of freeway control via ramp metering on the corridor at which it is implemented (a). Impacts on the individual intersections affect "local" system interests and activities. Impacts on the freeway affect "global" system interests and activities.

(a) The word "likely" is used because there is no hard evidence or documented research based on field studies in this regard.

TABLE II

SUMMARY OF LIKELY IMPACTS OF FREEWAY CONTROL VIA RAMP METERING ON BOTH INTERSECTION & FREEWAY SYSTEM INTERESTS & ACTIVITIES

Impact Category	Local (Intersection) Interest/Activity	Global (Freeway) Interest/Activity
(1) Physical Impacts		
Aesthetics & historic value:	Insignificant	Irrelevant
Infrastructure:	Altered function for surface streets.	Less need for freeway expansion
Terrestrial ecosystems:	Increased local smog, dust, & litter.	Reduced overall pollution
Aquatic ecosystems:	NA	NA
Air quality:	More local fumes, exhaust emissions	Reduced overall air pollution
Noise & vibration:	Increased noise and commotion	Less stop-and-go conditions
Damage to adjacent property:	Blocked entrances/reduced access.	Favors outer area users (by better long freeway commutes)
Traffic circulation & parking:	Problem transferred in from freeway	Improved overall mobility
	Queueing	Optimal uniform flow
	Higher demand on surface street capacity	Dispersion of peak-traffic
	Congested intersections	Less bottlenecks
	Illegal traffic movement & maneuvers	Re-routing of short-trips
Public safety:	Reduced safety due to awkward queue overspills and ramp violations	Improved safety/less rear-end, injury accidents
Energy:	More fuel consumption on ramps	Less fuel consumption on freeway
(2) Economic Impacts		
Employment, income, business activity:	Adversely affected due to congestion.	Better freeway trucking service
	or favorably affected due to "exposure"	Better access to suburban businesses
Residential activity:	Higher access & maneuverability costs.	NS
Effect on property:	Potential lower land value	Better land value in upstream areas
	Expand right-of-way (add more ramp lanes)	Initial construction/set-up costs
Regional & community plans:	Obstruction to traffic, school bus etc.	Additional costs for enforcement
Resource consumption:	Vehicle tear/wear increased	Optimized resource utilization
	More time delays/more incidents	Less time delays/less accident costs
	Higher surface traffic management costs	Additional system operations costs
(3) Social Impacts:		
Displacement of people:	May affect urban growth, land use/clusters	Serves suburbanization trends
Accessibility of facilities/services:	Reduced access during peakhours (refer to physical impacts above)	Facilitate suburban commute & urban sprawl due to improved freeway mobility and suburban accessibility
Effects of terminals on neighborhoods:	NS (also see physical impacts above)	NS
System users:	Public complaints about ramp delays	Motorists enjoy better freeway drive & safer trips
	Local residents resent congestion	
	Demand for fairness in meter rationing	Global-local conflict of interests
	HOV violations create citizen frustrations	
Special user groups:	Carpooling & mass transit favored (modal split encouraged)	More privileges to motorcycles, HOV mass transit, and transient traffic

NA: Not applicable, NE: No effect, NS: Needs further studies

Although the impact classifications of table II appear to be simple and proper, it is necessary to make one important distinction. The classifications are for studying the effects of existing or proposed "facilities". Even though ramp metering projects often involve roadway construction and pavement work, ramp metering is viewed as more of an "operation" than a "facility". Furthermore, this operation is a periodic one. Most metering systems operate during peak-hours only. Consequently, most of their impacts are instantaneous and temporal. Detailed investigation of these points is beyond the intended scope of this study. Again, further research is needed to address such questions.

Third, as discussed before, the intersection-freeway activity system is a sociotechnical system by definition. Therefore, freeway control via ramp metering could be investigated through at least three viewing perspectives: *technical, organizational,* and *individual.* Both global and local interests may be viewed separately, proportionally, and simultaneously by all three perspectives.

2. Problem Solving

Figure 10 illustrates the proposed systems approach for analyzing the impacts of freeway control via ramp metering. The two-dimensional "local-global" impacted activity system is suggested by the author. Again, most of the impact categories are adapted from the literature [30]. The T-O-M viewing perspectives are modified after Linstone's multiple perspective approach.

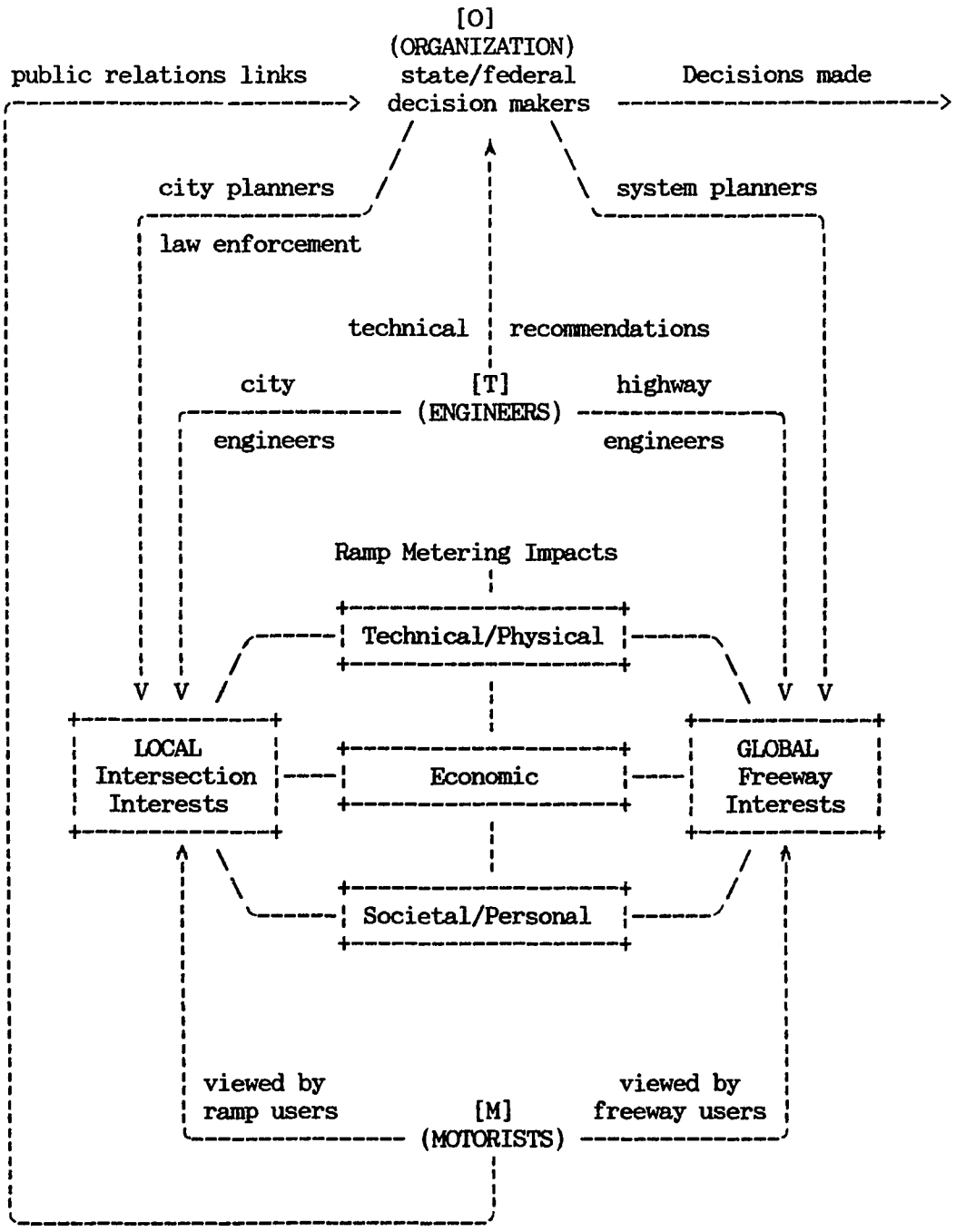


Figure 10. A Proposed Systems Approach for Analyzing the Physical, Economic, and Societal Impacts of Freeway Control Via Ramp Metering on Local and Global System Activities and Interests Using Views of Three T-O-M Perspectives.

Engineers conduct their microscopic analysis from a technical [T] perspective. Highway engineers begin with a comprehensive operational evaluation of the impacts/feasibility of freeway control via ramp metering. As suggested by Blumentritt [2, 1981], four types of impact studies could be performed.

1. Geometric Analysis: which includes ramp design, right-of-way, etc.

1. Bottleneck Analysis: which includes freeway saturation, redistribution of traffic demand, mainlane merging, and travel time delays.

3. Traffic Diversion Analysis (done with local traffic divisions): to include peak-period dispersion, short and long trip assignments, surface street capacity analysis, and traffic priority assessment.

4. Accident Analysis: including public safety studies.

The engineering report is culminated by a benefit/cost analysis which converts to dollar worth at least the following items:

Benefits:

* Travel time delays/savings, say $ RTT.
* Reduced pollution/environmental gain, say $ EG.
* Accident reduction/savings, say $ AR.

Costs:

* Initial system installment/maintenance/enforcement cost, say $ ICM.
* On-ramp user costs, say $ RU.

Then, B/C = (RTT + EG + AR) / (ICM + RU) must be greater than 1.0 to justify metering. The benefit/cost report may be incorporated into the *highway engineering report.*

System planners view the two-dimensional problem from a higher macroscopic stance, accommodating not only operational arguments, but also the strategic "systems planning" ones. Planners evaluate the strategic impacts which include modal split and lane use implications. A similar cost/benefit analysis may be done at this level as well to encompass the dollar worth of at least the following:

Benefits:

* Benefits of modal split and mass transit gains.
* Benefits of trip diversions.
* Corridor development.

Costs:

* Cost of land use impacts and effects on property.
* Costs to residential/business activity and displacement of people.
* Enforcement costs.

The benefit/cost report compliments the final *systems planning report.*

Finally, the social/personal impacts are investigated via a system users' perspective (motorist [M] perspective). All three perspectives are integrated into a well-rounded, comprehensive decision making process which is carried at the organizational [O] level.

The above approach has been outlined to help put the problem of freeway control via ramp metering in perspective. The outline has been used to guide a limited pilot study on this subject. The study is discussed in the following sections.

PILOT FIELD STUDY ON THE [O] & [M] PERSPECTIVES OF FREEWAY CONTROL VIA RAMP METERING

A pilot field study has been conducted by the author in order to further understand and appreciate the organizational [O] and personal motorist [M] perspectives. Three sources of input have been utilized:

* Literature reviews.
* Field interviews with some engineers, planners, decision makers and law-enforcement officers in Portland, Oregon.
* Small-scale system user survey in Portland, Oregon.
* Personal field observations at Interstate 8, San Diego, California.

1. The Organizational [O] Perspective

Systems decision making is made by the upper-level management of the organization (usually a state department of transportation) with state funding and many federal guidelines and regulations. Enforcement, cost/benefit analysis and the evaluation of the strategic arguments of ramp metering could be conducted by systems planners and analysts through a comprehensive organizational [O] perspective. Both local and global system activity interests must receive balanced consideration.

Although transportation and land use planning are *supposed* to be a coordinated single process, they are often carried out by a variety of government organizations (federal, state, metro-county, and city agencies), each of which has its own distinct professional objectives and political stands. The integration of these (government agencies) is often fragile and superficial. The degree of financial and political power held by these groups is also markedly different [27, 1984]. Typical perspectives of each class of agency are discussed below.

a. Perspective of State & Federal Transportation Authorities.

Higher-level transportation authorities (state and federal) are upmost concerned with highway facilities. Their chief concern is that freeway resources are getting scarce [31, 1989]. At this high level, it is important to point out that highway engineering is well-established and receive consistent business and industrial support and political lobbying, something that is not true of planning in general. Moreover, highway engineering represents *a market-oriented approach*, which is carried out by politically powerful state agencies and supported by financially able federal institutions [27, 1984]. Support is thus for the "global" rather than "local" activity system interests.

The perception at the state-federal level is that any measure that optimizes the utilization of these (global) resources should be considered. Since metering is technologically valid and will serve the "global" interest, it *must* be adopted and implemented whenever and wherever it is warranted, even if some local land use functions may have to be sacrificed. This philosophy automatically supports giving freeway operations *higher priority* (including ramp metering operations at points

of entry). Land use developments, densities, and arrangements alongside and around access roads (at the local level) will have to "self-adjust" to provide compatibilities with restricted freeway access [11, 1982]. Some transportation experts argue that if existing land use patterns are incompatible or inflexible with the conditions of freeway proximity, they eventually will have to be eliminated or relocated elsewhere, away from the freeway [9, 1989].

b. Perspective of Local (City) Urban Planning Authorities. The "local" perspective is normally represented by the city planners, engineers, and architects. Dotterrer [10, 1989] who is a chief urban architect from the City of Portland, was interviewed to provide some insights on typical local planning interests and concerns. Dotterrer was quite interested in issues of *fairness* to the city residents and to the welfare of the central business district (CBD):

> The ramps installed along I-5 seem to have made a substantial difference [on traffic]. When the system was first implemented, we had a number of complaints from people who were not used to waiting. The City had a lot of concern about the system when it was first put on. The city had a written agreement to guarantee some protection that the problems aren't transferred from the interstate highway to the arterial system. While it is important that through freeway traffic enjoys continual and smooth flow, it is also equally important that the arterials, which serve other functions, go at comparable smoothness. These functions [of the local streets] could not be eliminated or impaired for the sake of freeway traffic.
>
> The city of Portland has had a regrettable experience when the city had to limit traffic access to the area at near the intersection of Sandy boulevard/Burnside Street. The area used to be full with business activities, fast food restaurants etc. But due to [severe] traffic congestion and consequently reduced access, the entire area was abandoned. It is now totally vacant and business moved elsewhere, except for some parking lots. All businesses were hit with big losses and the land value took a plunge and [the area] it was not serving any purpose any more. This should never happen again. We can't tolerate this happening to a parallel street by putting a ramp meter just to let the

freeway run a little more nicely.

Dotterrer also stressed the fact that ramp metering is not an isolated freeway management issue. The City's *mass transit* program, for example, could be adversely affected by an awkward ramp metering system:

> Metering should not be put on ramps which don't have a bypass lane for carpoolers and buses. Ride sharing plans should not be impeded by meters or ramps with single lanes. If the DOT insists on putting a meter, they will have to redesign the ramp first to include either a bypass lane for HOV's (high occupancy vehicles) or perhaps restripe a usable shoulder.

c. Perspective of Mid-Level (Metro) Authorities. The Portland Metro Service District perspective was solicited as typical mid-level authority with an appreciation for local as well as global interests. This perspective is summed through an interview with Andrew Cotugno [9, 1989], the transportation director at Portland Metropolitan Service District:

> The "good old" days of free access to freeway may be gone forever. One of the most fundamental tasks of the Metro District is to conduct transportation studies for each of the regional corridors to identify the problems, determine optimal solutions, initiate projects, and so forth.
>
> From a purely microscopic point of view, the Transportation Department can not be blamed for limiting the entrance ramps if the freeway capacity is limited and producing substantial (freeway) queueing. When there is a lot more demand on a corridor than what the freeway can handle, all the affected communities need to figure out how to provide additional transportation capacity. It could mean that the freeway has to get bigger, the arterial system has to be bigger, the transit system has to be better and so forth.
>
> The ramp metering conclusions are correct whether the overall transportation system is adequate or not. Traffic problem that appear at the ramps may actually be caused by problems of *overdevelopment* or awkward patterns of land use at a given corridor and not actually caused by the ramp control itself.

d. Perspective of Law Enforcement Agencies. Ramp control relies heavily on voluntary compliance of the public. Therefore, if motorist discipline is lacking, enforcement becomes a prerequisite for effective operations. If enforcement is absent and the violation rate is high, the metering system will result in a net deterioration of the freeway service, safety at the ramp will suffer, and illegal traffic activities at the local intersection will increase.

To inquire about the viewpoint of law-enforcement authorities, the author sought to learn about the perspective of the Portland Police Bureau. It was assumed that the views of this law-enforcement agency would give a sample of police attitudes toward problems of this kind. It is acknowledged however that such views do not necessarily represent all views of all enforcement agencies in all the states.

The view of the Portland Police Bureau was reflected through an interview with Sergeant Michael Foss [13, 1989] of the Traffic Division. He stated their position in a frank, clear, and simple manner:

> Some of the ramp meters do present potential need for enforcement. The meter is as good as the public compliance with it. It is impossible to have full enforcement to any system at any police jurisdiction and therefore, "selective" enforcement is necessary. The potential for violation is higher at ramps with HOV lane(s). Frequently, every third car [this may be exaggerated], there is someone by themselves on the HOV lane. The biggest drawback of this is the frustration, heart burn and ill-feeling that it causes to the guy sitting in the regular lane doing his civic duty trying to obey the traffic laws and trying to cooperate with the system that is supposed to make life easy for every body.
>
> From *the police department standpoint, however, there should not really be any difference between a vehicle with one person in it and a vehicle with three persons in it,* in terms of ramp metering. A single vehicle will occupy the same amount of space and will have the same impact on the freeway flow whether it has one person or three in it. We see no value in this preferential treatment. This double standard doubles the strain on the police officer during enforcement to monitor car occupancy as well as obedience

traffic lights.

Our community [Portland, Oregon] and local courts are quite liberal. There is very little cooperation [with the police] at the Multnomah County District Court, for these types of violations, or any type of violation as far as that goes. Therefore there are no easy situations to enforce. An officer [issuing a citation] has to be able to testify what color the light was, when the vehicle disobeyed, how far back it was when they disobeyed, how and when they pursued the vehicle and so forth. Furthermore judges don't see themselves as law enforcers, but rather, as some type of policy setters and traffic experts and decide, on their own, which ramps are good and must be enforced and which are not, defeating the entire strategy of the ramp control system.

When the ramp meter program first went out, police enforcement was done. However, it was found that it was not cost effective or an optimal use of our resources and therefore it was no longer done. People will see police as deterrent as long as the police is there. Once they leave, things go back to normal and violations continue to occur.

Our experience with the motorists in this area is that you will get voluntary compliance from x percentage of people regardless of enforcement. We may cite the same person three times in three days and still don't get compliance.

On any given day, the city of Portland has only 8 motorcycle officers on duty for the entire city including freeways and surface roads. Traffic incidents get the highest priority. The police department cannot afford to deploy police officers to the ramps to enforce compliance with a traffic regulation that is not related to accident reduction. Our accident statistics do not indicate high incident rates at or nearby those ramps.

Frequently when we have citizen/resident complaints at a certain ramp, we dispatch someone to investigate the problem. It is the "squeaky wheel gets the grease" situation but, otherwise, no regular enforcement is [nor will be] done.

Finally, other transportation experts (Tom Schwab from Oregon DOT and Mikael Bauer from the transportation office of the city of Portland) were interviewed as well. However, their feedback was incorporated into the technical review section of this report.

2. The Motorist [M] Perspective

Ramp metering, like any other freeway management measure, needs to gain public and political acceptance. To the public, ramp meters are perceived as controls to an otherwise "free-way". Although definite

benefits may be achieved as a result of metering, the benefits may not be recognized by individual motorists. Few minutes of delay on the other hand are immediately recognizable [28].

Dotterrer [10] points out that ramp metering is a good example of a sociotechnical system where *technology changes at a faster rate than the society :*

> People and organizations don't like change. Ramp metering is not only a change, it is a change to the worse in terms of free access to the highway. A concept that cannot be explained to people, will not get accepted. When it is not accepted, it will be violated so much that it won't be accomplishing its goals.
> One of the tricky things about ramp metering [a socio-technical problem] is that it is quite technical and the theory of which is not immediately logical. Therefore, it is very undesirable to suddenly switch to the most optimal solution. The system must be introduced gradually one meter at-a-time and with reasonable [car per green] rates until the system is finally accepted.

A different perspective came from Cotugno [9] who believes that the voice of the traveling motorists should be heard, but, *it should not dictate the technical premise.* Education, Cotugno insisted, is very important to let the public know the advantages of employing these techniques and the disadvantages of not using them. He indicated that in no case should the "public complaints" lead to the abandoning of the concept. Cotugno left some room for flexibility: "Technical rules could be bend somewhat, nevertheless, to allow the affected communities participate in the "rationing" [distribution] of ramp meters, although this could present the technical side with numerous challenges."

Evidently, if the people (system users) are unhappy with the system or its operations, the system will have problems no matter how it is well-designed technologically.

PORTLAND SURVEY OF FREEWAY SYSTEM USERS

To probe some system users directly, a limited pilot study survey has been conducted by the author in Portland, Oregon. This survey may serve as a test case in preparation for a large-scale user survey (perhaps 10,000 copies) in San Diego, California. The principal goal of this survey has been to test the idea of opening a direct feedback channel for the system users to voice their concerns and help identify deficiencies in the system, as seen from their perspective. Although the number of questionnaires passed is small (145) and there might be several flaws in the design of the survey, the method remains an interesting and effective tool to work with the societal component of the system.

Due to time and space constraints, only twelve questions (1, 3, 5, 6, 8, 9, 14, 17, 18, 19, 20, and 23) of the original twenty-three in the survey have been selected for use and analysis. Less relevant and/or awkwardly stated questions (which lead to meaningless answers) have been skipped.

1. Sample Population & Distribution of Questionnaire

One-hundred and fifty copies of questionnaire were made and one-hundred forty-five copies were actually distributed. One-hundred copies were distributed during spring 1989, The other forty-five were distributed during the fall 1989. The number of returned copies was sixty-eight, about 47% rate of return.

The sample population was largely well-educated, working people. The sample is not representative of society as a whole but it is a good

segment of the society *to be tested first*. Large portion of questionnaire copies was distributed on Portland State University campus to students, faculty, and staff members. Participating students were mostly part-time graduate students who belong to the working category which uses the transportation system extensively. Other copies were handed out to people at places like First Interstate Bank, Kinko's Copies, Battelle NW Inc, Target store, and Portland Metropolitan Services District.

2. Statistics About the Participants (as obtained through question 23)

SS = 68, the sample size (the number of respondents who returned their questionnaires.

N = the number of people who answered a certain question.

N = 67 (Number of people who provided the personal information)

Average Age = 31.6 years

Male ratio = 46/67 = 69 %

Female ratio = 21/67 = 31 %

Education:

High School	=	1/67 = 2 %
In College	=	8/67 = 12 %
BS / BA (Received/In-progress)	=	10/67 = 15 %
MS / MA (Received/In-progress)	=	35/67 = 52 %
Ph.D. (Received/In-progress)	=	13/67 = 19 %

Let: high school=1, College=2, BS=3, MS=4, Ph.D.=5;

Then: average education = 3.76. (between BS & MS)

3. Discussion and Analysis of Questionnaire

The first half of the questionnaire was designed to gather information on the respondent's routine commute. For example, questions 1, 3, 5, 8, 9, and 17 were to learn about method of travel, trip patterns, and car occupancy. Questions 8 (estimate of trip distance) and 9 (estimate of travel time) measure the extent of freeway usage. The proportion of short and long trips can also be estimated from these two questions. Such information could be used by the traffic engineers to decide if too many short trip drivers are indeed using the freeway. For example, if the definition of short trip is less than three miles, then, according to the answers in Q8, 13 % of freeway trips are short trips. This number is an estimated ratio that could be targeted for diversion to surface streets. According to this survey, aiming at higher diversion ratios would perhaps be unrealistic.

Correlation between question 8 and 9 could be used to check the reliability of the motorist's estimate of time and distance. From question 8, the average *travel distance* is about 10 miles. Question 9 indicates an average *travel time* of about 15 minutes. This means that the average speed is about 10/(18/60) = 34 mph. Using the guidelines of the *Highway Capacity Manual* [14], this speed is very close to the expected average speed (about 35 mph) if a freeway is running near capacity during peak hours. [Note that the majority of respondents use the freeway during peak hours].

The other half of the questionnaire consisted of mostly attitudinal questions designed to solicit a reaction from the motorist (approval, disapproval, expression of anger/frustration, etc.) to a proposed idea.

Questions 6 and 14 were to test whether motorists (especially short trip travelers) would avoid the freeway and choose to travel on city streets because of delays at the ramp meters. Questions 18, 19, and 20 were to probe motorist's acceptance of alternative modes. Follow up sub-questions were intended to clarify why a motorist chose a certain answer or to sub-categorize the answer.

Some underlying concepts and perceptions were tested repeatedly through several questions (in different contexts) to see how clear the respondent's perspective was and how consistent his answers were. For example, question # 3 (trip purpose) may be used to support/contest the motorist willingness or ability to *CHANGE* the time, method, route, or mode of his trip, as asked in questions 5 (travel during peak hours), 6 (willingness to change), 14 (willingness to tolerate ramp delays), 17, 18 and 19 (use of carpooling and mass transit as alternative modes).

Some answers indicated confusion in understanding the questions. For example, questions were about the morning commute but this was not stated explicitly in the questionnaire. Some people based their answers on the assumption that questions relate to the afternoon period perhaps because this is when they experienced more traffic problems. For San Diego survey, only the *morning* period in the direction of *heavy* traffic will be examined and that must be made clear to the participants.

Another source of confusion was the use of the word "ramp" while the word "on-ramp" or "entrance ramp" should have been used. Some answers indicated that drivers were having problems with off-ramps, which is not the focus of this work. This confusion should be eliminated in San Diego survey.

Mostly descriptive frequency distribution statistics are presented. No rigorous statistical analysis has been performed on any observation. Relative frequencies were based on the number of people, N, who answered the question. Although N varied from question to question, it is still possible to check the correlation between answers. If one-to-one correlation test on a pair of questions is desired, the lower N could be used. Comparisons based on ratios between questions is also possible because N varied only slightly (from 61 to 68). Computations of the mean and standard deviation and other statistical parameters for most answers are omitted.

The following is an exhibition of selected questions, the purpose of each question, compilation of received answers, selected respondent comments, and corrections needed for San Diego survey. When San Diego survey is completed, a thorough analysis and comparison could be done. Note that "N" in each question section refers to number of respondents who answered that question.

The statement of each question is presented as it appeared on the original questionnaire. Statistics are presented in a mixed tabular and graphical format for simplicity and clarity. Frequency distributions are presented in "histograms" with each "*" denoting one occurrence.

Q1: Do you have a car? If yes, how often do you use it?

Purpose of Question: Inquire about car availability/usage.

Answers:

	N = 68	Rel Frq
Yes	64	64/68 = 94 %
No	4	4/44 = 6 %

83% use cars at least twice a week.

Needed Corrections (for San Diego Survey): Eliminate the entire question. Survey will be handed directly to freeway users at on-ramps -all metered.

Note: Questions 3, 5, 8, 9, and 17 are to learn about trip patterns and car occupancy.

Q3: What is usually the purpose of your (freeway) trip? (Check Category)

Purpose of Question: Learn about trip generation cause and objective.

Answers:

	N = 66	Rel Frq
Work	**********	37/66 = 55%
School	**********	15/66 = 23%
Social Activity	*****	5/66 = 8%
Shopping	****	4/66 = 6%
Recreation	***	3/66 = 5%
Other	**	2/66 = 3%

Selected Respondent Comments: "School and Social activity are equally balanced."

Q5: Do you have to use the freeway during peak traffic hours?

Purpose of Question: Learn about restrictions on trip generation time.

Answers:

	N = 67	Rel Frq
Yes	43	43/67 = 64 %
No	24	24/67 = 36 %

Needed Corrections (for San Diego Survey): Identify to the participants what the peak hours are.

Q8: Approximately how far (how many miles) do you drive on the freeway between the point you enter and the point you exit the freeway?

Purpose of Question: Determine the proportion of freeway usage for short and long distances (establish some O-D information).

Answers:

	N = 64	Rel Frq	Cum Frq
1 mile or less	***	3/64 = 5 %	5 %
1 - 2 miles	*****	5/64 = 8 %	13 %
3 - 5 miles	********	8/64 = 13 %	26 %
6 - 9 miles	**********	22/64 = 34 %	60 %
10 - 15 miles	**********	13/64 = 20 %	80 %
15 - 20 miles	******	6/64 = 9 %	89 %
20 - 25 miles	**	2/64 = 3 %	92 %
20 - 30 miles	*	1/64 = 2 %	94 %
Over 30 miles	****	4/64 = 6 %	100 %

Mean value = 10.41 miles

Standard deviation = 7.65 miles

Q9: Approximately how long does your trip last on the freeway?

Purpose of Question: Estimate travel time, compute speeds (S=D/t), and validate data from Q8.

Answers:

	N = 64	Rel Frq	Cum Frq
1 - 2 minutes	***	3/64 = 5 %	5 %
3 - 5 minutes	****	4/64 = 6 %	11 %
6 - 9 minutes	*******	7/64 = 11 %	22 %
10 - 15 minutes	**********	20/64 = 31 %	53 %
15 - 20 minutes	**********	15/64 = 24 %	77 %
20 - 30 minutes	********	9/64 = 14 %	91 %
Over 30 minutes	******	6/64 = 9 %	100 %

Mean value = 18.48 minutes

Standard deviation = 7.85 minutes

Needed Corrections: Add "between the point you enter and the point you exit." Highlight (or underscore) the word <u>freeway</u> in the statement of the question above to emphasize freeway time only.

Q17: How many passengers do you ***normally*** have in your car in your daily commute?

Purpose of Question: Measure carpooling rate and high occupancy vehicle (HOV) lane usage.

Answers:

	N = 61	Rel Frq
One	**********	47/61 = 77 %
Two	**********	11/61 = 18 %

Note: Q18, Q19 & Q20 probe motorist's perspective on alternative modes.

Q18: Have you ever considered carpooling?

Purpose of Question: Check if ramp/freeway delays encourage carpools.

Answers:

	N = 65	Rel Frq
Yes	**********	17/65 = 26 %
No	**********	48/65 = 74 %

Q19: Have you ever considered mass transit?

Purpose of Question: Determine if traffic delays would motivate people to consider/use mass transit.

Answers:

	N = 65	Rel Frq
Yes	**********	33/65 = 51 %
No	**********	32/65 = 49 %

Needed Corrections: Ask about the "frequency" of use of mass transit.

Q20: Do you think ramp diamond lanes (high occupancy lanes) provide a service or a disservice to cars entering the freeway?

Purpose of Question: Check people's attitude on preferential lanes.

Answers:

	N = 65	Rel Frq
Service	**********	28/65 = 43 %
Disservice	**********	23/65 = 35 %
Don't Know	**********	14/65 = 22 %

Selected Respondent Comments: (Resp. 40) "Can't regulate who uses them."

Note: Q6 and Q14 test motorist's flexibility to change trip time/route.

Q6: Can you ***change*** your schedule to avoid traffic peak hours?

Purpose of Question: Determine whether ramp delays/freeway congestion would cause people to change trip time/pattern. (Test of willingness/feasibility)

Answers:

	N = 66	Rel Frq
Yes	**********	31/66 = 47 %
No	**********	24/66 = 36 %
Uncertain	**********	11/66 = 17 %

Selected Respondent Comments: "Office closes at 5:00 [no flexibility]."

Q14: Assume ramp control really makes your trip much smoother and faster on the freeway, what would be a ***maximum*** time you are willing to wait at the ramp before you start to think about an alternate route/ramp?

Purpose of Question: Measure user tolerance to time delays.

Answers:

	N = 62	Rel Frq	Cum Frq
0 minutes (+)	*	1/62 = 2 %	2 %
5-10 seconds	*******	7/62 = 11 %	13 %
10-20 seconds	******	6/62 = 10 %	23 %
< 1/2 minute	**********	12/62 = 19 %	42 %
< 1 minute	*******	7/62 = 11 %	53 %
1-2 minutes	**********	14/62 = 23 %	76 %
3-5 minutes	********	8/62 = 13 %	89 %
6-9 minutes	*	1/62 = 2 %	91 %
> 10 minutes	**	2/62 = 3 %	94 %
Open (++)	****	4/62 = 6 %	100 %

(+) Would tolerate no restriction.

(++) Would accept judgment of traffic experts

4. Selected General Comments From Respondents (quotation marks omitted)

Respondent # 5 (desktop publisher): The controlled ramp I use operates efficiently for me. I find more problems with *people* trying to cut in front of me the last minute [perhaps due to permanent absence of police enforcement].

Respondent # 7 (student): Survey should ask: Do you feel that ramp control facilitate the traffic flow.

Respondent # 9 (programmer): I normally travel the Portland freeways at all times of the day and frequently encounter ramp metering in the morning and in the evening. In general, I think it does improve the traffic flow.

Respondent # 33 (nurse): Question 13 will be hard to standarize since you are using % of time travel rather than absolute scale.

Respondent # 35 (scientist/engineer): Questionnaire does not distinguish between origin-destination in round trips.

Respondent # 36 (economist): I have considered moving to be closer to work and avoid I-5. I don't mind waiting for ramp lights, but I DO mind getting stuck behind trucks that spill stuff on the freeway. This has happened three times in two months with delays of 15-45 minutes. I would like to see trucks restricted during rush hours. [Trucks go through Portland as transient traffic. They enjoy good flow during peak hours -due to ramp metering- but they do not share the burden of waiting at ramp meters. This is unfair to local users and truck movement should be restricted during peak hours].

Respondent # 39 (department secretary): Too hard to use mass transit with child to get to and from baby sitter. Freeway must have lanes just

for trucks, or, at least, restrict them from fast lanes.

Respondent # 40 (political science student): Questions should also deal with stops and goes on the freeway.

Respondent # 44 (manager): What does Q13 mean? Is Q15 about on or off ramp? -Unclear as to exit/enter ramps. Questionnaire poorly designed. Too long.

Respondent # 45 (data base manager): I have considered mass transit, but use it only 1-3 times/year. Don't care to associate with riders & dirty seats. Diamond lanes serve good purpose only in principle, not in fact.

Respondent # 46 (corporate manager): [Survey should include the question] should preference be given to HOV vehicles? My answer is no.

Respondent # 58 (army officer): My normal time to begin the morning commute is 0530-0600. Hard to find carpool partners at 0530 a.m. US Hwy 26 is awful, even dangerous. It is literally every man for himself, & only the strong survive.

Respondent # 61 (mechanical engineer): Ramp lights sure beat sitting behind [tailing] an old lady at the end of the ramp w/o any way to accelerate onto the hwy & into the traffic flow. [Ramp meters normally discharge vehicles one-at-a-time, thus preventing such close platooning of vehicles merging onto the freeway].

Respondent # 66 (professional): Wait is on the freeway itself (not on the ramps). [Heavy] traffic on the Sunset Highway is due to not enough access to Wash. County. I don't think ramp control will make any difference.

5. Questionnaire Conclusions

On the average, motorist could be flexible and they are not impossible to manage or satisfy. From question 6 (schedule change), a majority, 47%, said that they would modify their trip patterns to escape congestion. Question 14 asked if the motorist would tolerate ramp meters. Only 2% said no. Most showed that they will accept moderate kind of control on their method of travel if it improves their overall commute. But less than 24% would accept delays over two-minutes.

The majority of freeway users might be limited in their choice of trip time and route. Work and school (compulsory trips) accounted for 78% of trip purposes. 64% of the respondents need to use the freeway during peak hours. Many of them indicated a difficulty to change trip time to avoid freeway delays and ramp troubles due to employer/work restrictions. Reducing peak-hour recurring delays that result from work commutes would require staggered work hours,which requires involvement and cooperation of employers -particularly government and big companies.

Car ownership is very high (94%) and so is auto use. 83% of car owners use them to commute to work at least twice a week. *Carpooling seems to be a very hard idea to sell.* Answers to questions 18 and 20 revealed little interest in ride sharing. Only 26% have ever considered carpooling. Although 44% thought HOV (High Occupancy Vehicle) lanes are a good idea, many expressed doubts about their efficiency. Respondent # 45 (male:38, database manager) said that HOV lanes serve a good purpose "only in principle, not in fact," a vote against giving preferential treatment to HOV traffic (which ramp metering must favor). Of course, most respondents who did not like HOV lanes said they traveled alone.

Respondents seems to prefer mass transit over carpooling. Answers to questions 18, 19, and 20 support this conclusion. While only 26% had ever considered carpooling, about 51% have considered (and seem to like) mass transit. Some indicated that they used mass transit to supplement their automobile or bike commutes.

Some Respondents liked to give "engineering" judgments. For example, respondent # 65 (female:30, industrial engineer) said: "Design of off/on [cloverleaf] ramps is very poor. Ramps should be set up so that exit ramps come before entrance ramps. The opposite is what is popular here. This is inefficient and dangerous." Respondent # 6 (male:34, engineer) said: "The ramp I use is essentially controlled by a traffic light near the entrance. The ramp light is really not necessary." Some of these views may be of value. The safety point which respondent # 65 mentioned is worth consideration. But comments should be reviewed carefully to separate valid concerns from technically incorrect motorist conclusions.

Occasionally, respondents try to divert the focus issue. Many drifted into irrelevant problems. Respondent # 1 (female:44) commented: "On I-5 between Columbia River Bridge and 6th, the emergency lane availability is terrible! We also need a third bridge across Columbia." Other comments complained about transient trucks although truck traffic was not discussed in the survey. Some comments suggested discussing freeway flow on the mainlanes and not merely ramp access to them.

Finally, *respondents were generally polite and respectful.* Only two were aggressive and delivered nasty comments. They denounced the survey, condemned the approach, and criticized the surveyor!

CONCLUDING REMARKS AND QUESTIONS FOR FUTURE RESEARCH

Ramp metering is a proven method to improve the efficiency of freeway operations during peak hours. The method has been demonstrated to help control freeway saturation, utilize street capacity, disperse peak period traffic, redistribute traffic demand, regulate mainlane merging, save on travel time and improve travel speeds, improve motorist safety, assign proper traffic priority and identify deficiencies in the traffic management system. But the *effects of ramp metering on long-term system planning issues, such as modal split, driver's behavior, and land use patterns are not clear.* There is not enough documented research on this topic, particularly on the impacts of ramp control on access roads to freeways and on entrance ramp intersections.

In this chapter, a systems approach for identifying and analyzing the potential impacts of freeway control via ramp metering has been proposed and discussed. It is based literature reviews and the author's own experience. It is tested with a pilot study which included interviews with field professionals, a limited system user opinion survey, and empirical field observations in San Diego, CA.

The proposed system approach suggests that two activity system categories, "local" intersection (residents/motorists) and the "global" freeway, seem to be affected at three impact categories: physical/operational, economic/land use, and social/personal. Both activities and the impacts can be viewed from a systems 3-D perspective.

Many questions remain partially or completely unanswered. Some are

technical and others are economic, political, and social. For example, does the gain in freeway flow continuity always justify time delays and motorist inconveniences at the ramps= Who has the knowledge/authority to decide? What are the trade-offs for the individual motorist? for the traffic system? What are the costs of this technique? operational costs, tear & wear costs, maintenance costs, and pollution costs -if any?

If a one-lane ramp with geometric limitations warrants a meter, will it be metered? If the answer is negative, how can the interruption to the freeway flow be tolerated? If the answer is affirmative, then how can delays to carpoolers and mass transit be justified?

Is the "planned" ramp delay better and safer than "random" delays? What is the maximum waiting time beyond which a ramp user should not be subjected? What is the tolerance criteria? Does the wait at ramp queues aggravate the driver so much that it provokes an aggressive driving mood by the time he merges onto the freeway? How does that affect safety?

Will too many long-trip drivers avoid the freeway and choose to use the city streets instead? Where is the cut-off point? Will this create a danger of transferring freeway problems to parallel streets? In this case, what are the short and lone term impacts on street traffic patterns? On land use development? On access to surrounding neighborhoods? On access to the freeway itself? Who gets hurt? Why? How? and by how much? How much of the decision making process is influenced by pure scientific evidence and how much of it is influenced by political power and self-serving organizational objectives? *Educated answers and convincing arguments could be found only if a systems approach for problem solving continues to be pursued.*

The review of the theory of ramp metering and the proposed systems approach in this report would provide a good first step towards a comprehensive, non-reductionist understanding of freeway control via ramp metering.

---*---*---

CHAPTER III

ANALYSIS OF SIMULATION MODELING STRENGTHS & PITFALLS, REVIEW OF EXISTING FREEWAY MODELS, AND A PROPOSED MODELING APPROACH

CHAPTER PREVIEW

In this chapter, the advantages of simulation modeling as a systems analysis tool are presented and discussed. That is followed by a discussion and examples of the limitations of this tool. Recommendations for developing good models are summarized.

Prior to the adoptation of any modeling strategy, a review of existing freeway models is made. The objective is to examine the state-of-the-art in traffic simulation and learn from experiences of others. Finally, a modeling strategy for developing the ONRAMP model is formulated.

ADVANTAGES OF SIMULATION MODELING (ARGUMENTS FOR MODELING)

The importance of simulation modeling as a versatile systems analysis and problem solving technique in the field of engineering is unquestionable. Only simulation models have the capacity (at least in principle) to incorporate system capacity computations with vehicle-driver characterization, discrete and continuous system event processing, stochastic representation and frequency distributions, statistical computations, queueing and shockwave analysis, if-else-then comparison testing, decision making processes, and perhaps VDT graphical

interfacing, all in one integrated analytical package.

Zeigler [64, 1984] has predicted that "Modeling, in its computerized form [simulation], increasingly will take its place as the key knowledge component in all forms of decision making in modern life." Shannon [60, 1975] classified modeling and simulation as "one of the most powerful analysis tools available to those responsible for the design and operation of complex processes or systems." The continuously advancing computer technology has been giving increased credence to these claims.

Simulation modeling as an *alternative* systems analysis tool is particularly useful for solving many complex systems engineering problems which do not have adequate or exact ***analytical*** solutions. Uncertainty and random behavior often characterize many of these systems, where exact solutions are either insufficient or incompatible with the inexact nature of the problem. Freeway control via ramp metering is a typical example of such a (transportation) system problem.

Physical laboratory models and field experiments have been used as R&D tools in engineering and industry for developing new products or improving existing ones. But creating lab prototypes of the highway system (which includes ramps, freeway lanes, vehicles and drivers) is not very feasible. Equally inconvenient and impractical is the field testing of such systems. Many of the field conditions will not occur at the desired testing time, or may not last long enough or be severe enough for analysis purposes. Shannon [60, 1975] also warned of "Hawthorn effect" during field experimentation that involve people, where those who are being observed may modify their behavior, perhaps

only temporarily, causing the observer to draw false conclusions.

Computer simulation modeling provides an alternative approach which could be used instead of physical models and field experimentation at two phases:

1. Systems Engineering Phase: At the early stages of systems design, modeling and simulation makes it possible to experiment with different *computer prototypes* of a system that is under development and does not exist yet physically.
2. Operations Research Phase: The system may have already been developed and deployed, but a control/optimization plan for the operation is needed. Simulation modeling substitutes for field experimentation which might be unsafe, costly, or too disruptive to system users. May [52] pointed out the unique flexibility of simulation models where "Systems can be studied in real-time, compressed time, or expanded time."

LIMITATIONS OF SIMULATION MODELING (ARGUMENTS AGAINST MODELING)

May and other experienced researchers prefer the use of traditional analytical techniques first. "Because of the complex nature of simulation and the extensive time commitments normally required, simulation should be considered as the technique of last resort." [52, 1990]. Furthermore, "Simulation models require considerable input characteristics and data, which may be difficult to obtain." [52].

Systems simulation modeling has serious limitations especially when the modeled system is not purely mechanical. Linstone [17, 1984] explained that the difficulty is in modeling the human and social

aspects of sociotechnical systems where the reality created by the computer model is only "in the mind of the programmer or user" and "*can never be a duplication of a human or societal reality.*" Linstone points out that the computer *capacity* to store and run large-scale models is sometimes confused with its *ability* to represent the sociotechnical system complexity. As a result, "modeling seems to be fun for modelers, but it also can be a nightmare for real-world problem solvers." [17].

Since modeling is not an exact science and is as much art as it is science, presumptuous modeling projects can, inadvertently, create illusions about the modeled systems:

> In its most extreme form, modeling becomes an end rather than a means. The dedicated modeler reminds one of Pygmalion, the sculptor king of Greek mythology. He fashioned a beautiful statue of a girl and fell in love with it. Responding to his plea, the goddess Aphrodite brought the statue to life, and Pygmalion married his model. Today's modelers, mesmerized by the vast computer capacity, may also become wedded to their creations: *the models become the reality* [17].

Large-scale models representing large-scale systems have unique problems. Transportation systems (and traffic systems) are not only sociotechnical, but also large and complex, and so are their models. Lee [44, 1973] expressed serious doubts about much of the proclaimed representativeness of many of the so-called "comprehensive" urban models:

> Perhaps the least discussed problem in modeling is the deviation between claimed model behavior and the equations or statements that actually govern model behavior... Whatever the components of a city or a model city are, their microscopic behavior is largely unknown. The best information we have has to do with aggregate relationships that include the effects of an unknown but large number of other variables. To assume that these relations hold true in the same form when all other variables are allowed to vary independently has no basis in theory or experience. Multicollinearity (correlation of independent variables) makes the

> statistical evaluation criteria meaningless as well as making the coefficient estimates unreliable, and misspecification (omission of important variables) also makes the coefficients unreliable [44].

Modeling critics have always had legitimate concerns about the ways models are validated. Model verification and validation can work only within a given context. Validation is done to *confirm certain assumptions as hypothesized by the modeler*:

> Experimental design and validation of hypothesis are intraparadigmatic: they operate only *within* the framework of a perspective. They cannot prove that a model gives the most useful or "correct" representation of reality; they cannot give assurance that the variables chosen are sufficiently inclusive or appropriate. They tell us nothing about other perspective(s) [17].

Lee had a detailed discussion of the "seven sins of large-scale models" in his article "Requiem for Large-Scale Models". He identified these "sins" as: hypercomprehensiveness; grossness; hugeness; wrongheadedness; complicatedness; mechanicalness; and expensiveness. Lee's observations and conclusion were that:

> 1. In general, none of the goals held out for large-scale models have been achieved, and there is little reason to expect anything different in the future.
> 2. For each objective offered as a reason for building a model, there is either a better way of achieving the objective (more information at less cost) or a better objective.
> The overly comprehensive structure of existing large scale models has two aspects: (1) the models were designed to replicate too complex a system in a single shot, and (2) they were expected to serve too many purposes at the same time... A multiplicity of goals has surrounded these models, and the failure to separate ends and their associated means from each other contributed to the failure of the models... Too many variables and too much detail are included in the model structure. In practice, every additional component introduces less that is known than is not known [44].

Smith [61, 1975] quoted Yankelovich on the four most common fatal errors that modelers make when they face modeling difficulties:

> The first step is to measure whatever can be easily measured.

> This is OK as far as it goes. The second step is to *disregard* that which can't be measured or give it an arbitrary quantitative value. This is artificial and misleading. The third step is to presume that what can't be measured easily really *isn't very important*. This is blindness. The forth step is to say that what can't be easily measured really *doesn't exist*. this is suicide.

Although Smith warned of oversimplified novice modeling approaches, Lee's basic recommendation was that *in order for models to be usable they should be simple*:

> Probably the most important attribute any model should have is transparency. It should be readily understandable to any potential user with a reasonable investment of efforts... A balance should be obtained between theory, objectivity, and intuition. Excessive concern for theory results in a loss of contact with the policy... Build only very simple models. Complicated models do not work very well if at all, they do not fit reality very well, and they should not be used in any case because they will not be understood. *The skill and discipline of the modeler is in figuring out what to disregard* in building his model [44].

THE NEED FOR IMPROVED TRAFFIC FLOW AND TRAFFIC SYSTEMS MODELS

Despite a plethora of fully-developed and half-developed traffic flow models, continued improvement and innovation is required in two areas:

FIRST, the global approach of systems analysis and problem solving necessitates that a comprehensive, non-reductionist (i.e., not one-sided) system perspective be always adhered to. This helps assure that the modeler will not become wedded to his model, or be mislead by its apparent validity, representativeness, or applicability when the model seem to work well within a given, one-sided context or set of assumptions.

The "global" approach is also indispensable, from a systems point of view, to help understand the environment in which it will have to operate; to facilitate the field deployment of the model; and, eventually, to assist in the implementation of its recommendations.

SECOND, as the literature discussion in the following sections will demonstrate, the need for powerful but yet simple and manageable models will be apparent. Traffic engineers are increasingly interested in model depth and details, especially those which have a capacity to depict vehicle and driver characteristics. In the era of advanced electronics and advanced driver-TOC (traffic control center) communications, improved vehicle-oriented, operational models will be in demand. The following sections will provide a comprehensive review of the principal traffic modeling approaches and preview some models.

BASIC APPROACHES OF TRAFFIC FLOW MODELING

There are three approaches for freeway traffic flow modeling. The macroscopic approach, the microscopic approach, and a third approach, being referred to by this author as "mesoscopic". The first one is the most commonly used by traffic engineers and transportation planners.

1. Macroscopic Modeling

1.a. Initial Development. Macroscopic flow concepts date as far back as 1934. They are based on the assumption that traffic flow is analogous to the one-dimensional, compressible fluid flow. The stream of traffic flow is viewed as one mass "a-whole" and not as a collection of distinct parts (vehicles). Greenshields pioneered these studies [42,

1934]. He suggested that a linear relationship exists between speed of the flow and density of the flow. He proposed the first equation to describe the speed-density relationship:

$$V = Vf * (1-P/Pm)$$

where: V: instantaneous velocity of the flow
Vf: free-flow velocity (maximum velocity under non-congested conditions)
P: instantaneous density
Pm: density at maximum flow (maximum density at saturated flow)

Note that 100% saturation rate means standstill traffic, and thus if P=Pm, then V=0.

Over the years, the macroscopic flow theory was further investigated by many others. Greenberg [43, 1959] proposed a macroscopic flow model based on earlier developments of Lighthill and Witham [47, 1955]. He suggested the following equation to describe the relationship between speed and density of the traffic flow:

$$dP/dt + dQ/dx = 0$$

where: P: density of flow
Q: rate of flow

Drake, May, and Schofer [40, 1965] investigated a bell-shaped model for the velocity of the traffic flow. The following format gave satisfactory results when compared with empirical speed-density measurements:

$$V = Vf * EXP(-1/2 * (P/Pm)^2)$$

where: V, Vf, P, and Pm are as defined above

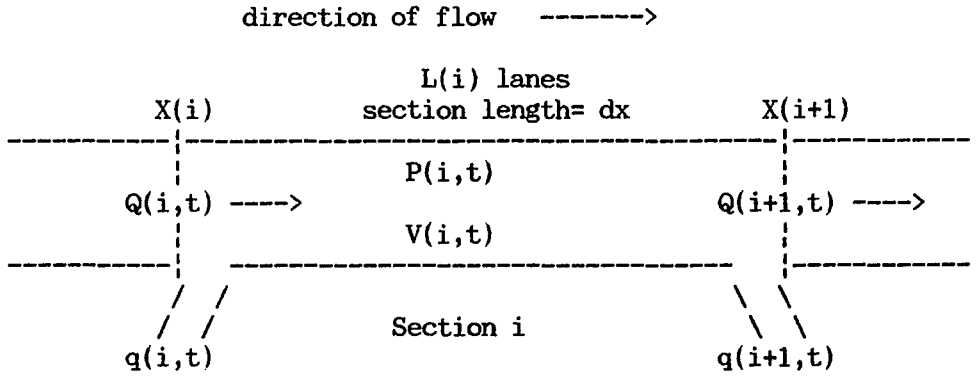
1.b. Paradigms of Macroscopic Modeling. Throughout the 1970's and 1980's, the macroscopic flow theory underwent extensive refinements and many macroscopic computer models were developed. The majority of traffic flow models that are available today are essentially macroscopic. They

gained acceptance and popularity because they require relatively moderate computational time -compared to microscopic models- while providing satisfactory performance evaluation when simulating certain transportation planning alternatives.

Most of the commonly used traffic flow macroscopic simulation models (such as FREQ, FREFLO, and MACK families), although each was developed for its own particular purpose, share the following set of underlying assumptions [36, 1985]:

1. The highway facility is divided into sections or subsections.
2. Time is discretized into small equal intervals.
3. Traffic demand, system supply, and highway performance are constant over the time interval.
4. Traffic flow is viewed as a mass of compressible fluid and the details of individual vehicle (its attributes and dynamics) are inconsequential with regard to the study of the overall system behavior.

The macroscopic approach for describing the flow of the traffic fluid generally requires the definition of three *equations of state* [59, 56, 36, and 41]: 1) The speed-density equilibrium equation; 2) the conservation or continuity of flow (vehicles) equation; and 3) the volume-speed-density equation. Given a freeway section as shown below in figure 1, the equations of state are described in a discrete format suitable for computer application:



Legend:
- t: time instant
- i: section ID
- X(i): upstream end of section i
- X(i+1): downstream end of section i
- q(i,t): on-ramp flow rate
- q(i+1,t): off-ramp flow rate
- Q(i,t): flow rate in veh/hr at time t and point X(i)
- P(i,t): flow density in veh/mi at time t throughout section i
- V(i,t): flow speed in mi/hr at time t at section i

Figure 11. Typical Macroscopic Freeway Section.

1.b.1. The Speed-Density Equilibrium Equation. Several theories (or empirical models) have been suggested -as explained earlier. The most commonly used speed-density equilibrium equation is:

$$V(i,t) = Vf * [1-(P(i,t)/Pm)^a] \qquad (1)$$

where:
- V(i,t): mean velocity of the flow in section i during time step t
- Vf: mean free-flow velocity (in non-congested flow)
- P(i,t): density of flow in section i during time step t
- Pm: density of flow in section i under maximum congestion conditions.
- a: a parameter (equals 1 or 2)

1.b.2. Conservation (or continuity) of Flow (vehicles) Equation. This equation was originally suggested by Lighthill [47] and subsequently used by many others. It states that, for a given section, the rate of change of traffic density over time and the rate of change

of traffic flow over space is equal to the rate of change of ramp traffic flow (all units expressed in veh/hour/mile).

$$dP/dt + dQ/dx = dq/dx \qquad (a)$$

where: P: flow density
Q: mainlane flow rate
q: ramp net flow rate

Considering that density is measured in veh/mile/lane, we must divide the freeway flow by the number of lanes, li, so that dP/dt term is expressed per one lane. Rearranging (a) yields:

$$dP = (dq/dx - dQ/dx) * dt/li \qquad (b)$$

but: dQ = Q(i+1,t) - Q(i,t) , for mainlane flow
dq = q(i+1,t) - q(i,t) , for ramp flow
= q(net)
dP = P(i,t+1) - P(i,t) , for mainlane flow

Substituting in (b) and rearranging the terms yields:

$$P(i,t+1) = P(i,t) + dt/(li*dx) * [Q(i,t)-Q(i+1,t)+q(net)] \qquad (2)$$

Note: The q(net) term will be used to eliminate the need to reverse the signs if the sequence of ramp flow is reversed. Using q(net), we can distinguish three cases:
If q(on) > q(off), then q(net) is > 0
If q(on) < q(off), then q(net) is < 0
else q(net) = 0

1.b.3. The Flow-Speed-Density equation. Finally, the flow-speed-density equation is given by:

$$Q(i,t) = P(i,t) * V(i,t) \qquad (3)$$

where: Q(i,t): flow rate in veh/hr at time t and point X(i)
P(i,t): flow density in veh/mi at time t throughout section i
V(i,t): flow speed in mi/hr at time t at section i

1.c. Special Difficulties with the Macroscopic Approach. A very common deficiency that has been associated with macro models is the discontinuity that results from discretizing the continuous equations of the state for computer implementation with large discrete time

intervals. Several authors [May, Babcock, Ross, Aerde, Yager, and Case (1)] reported model implementation inconsistencies due to this phenomenon. It often is the result of using large interval step, dt, in trapezoidal methods of integration. The resulting unsatisfactory discontinuities are mostly in the speed-density and flow-density curves.

Another common deficiency in macro models is their inability to simulate stochastic behavior. Probability distributions that describe traffic flow have existed for long time but were seldom applied in traditional highway capacity analysis which macro models are largely based on. With the increasing dynamic complexity of traffic networks and recent advancements in computer technologies, modeling stochastic system behavior is becoming a prerequisite for meaningful simulation.

A number of researchers have complained about the inadequacy of macro modeling for many non-standard highway configurations and certain dynamical situations. Cohen and Clark [37, 1987] indicated that these non-standard road configurations included items like closely spaced interchanges, short ramp acceleration lanes, and short weaving sections. Unique dynamical situations include heavy cross-weaving on certain road segments. Cohen and Clark [37] stated that neither FREQ nor FREFLO (both macroscopic models were adequate because they do their weaving and merging simulation based on the Highway Capacity Manual [14], which is particularly weak in analysis of substandard merge or weave sections.

2. Microscopic Modeling

2.a. Initial Development. Microscopic modeling of freeway traffic flow concentrates on detailed modeling of the movement of the individual

vehicle. The overall behavior of the traffic system (or traffic flow) is deduced by aggregating the actions of the individual parts.

The challenge to describe vehicular flow in a microscopic manner led several researchers like Reuschel [58, 1950] and Pipes [54, 1953] to propose a formula to describe the motion of pairs of vehicles following one other. This formula developed into the "Car-following" theory.

A Car-following model is a form of a stimulus-response relationship where the stimulus is initiated by the leading vehicle (on the road) and the response is exhibited by the vehicle that follows it in the platoon. Fox and Lehman [41, 1976] illustrated that the response of the following (trailing) car is the acceleration or deceleration of the driver and is determined by a stimulus function involving: The relative velocity between his car and the car ahead; their relative spacing; the absolute velocity level; and the driver's sensitivity -as well as many other human, mechanical, and environmental factors.

The microscopic approach has developed at a slower pace throughout the years because of its heavy computational demand. Meanwhile, most modeling efforts were spent on macroscopic simulation models.

2.b. Paradigms of Microscopic Modeling. Figure 2 shows the basic building block of a pair of vehicles in the Car-following model. Three basic equations are shared by the majority of microscopic models. They describe the acceleration/ deceleration of the individual vehicle, the position of the vehicle, and its speed.

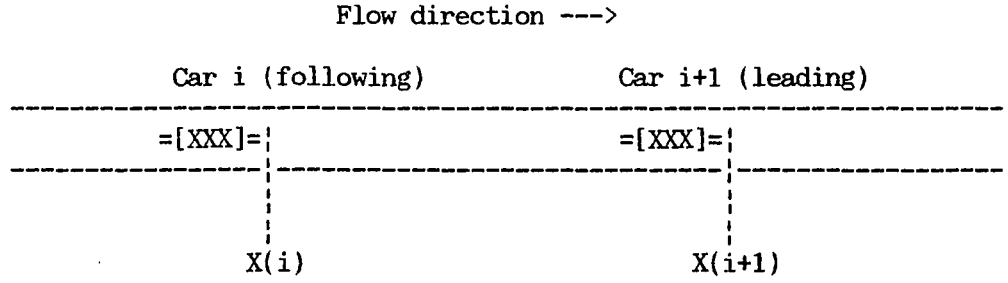


Figure 12. The Microscopic Car-following Model.

2.b.1. Acceleration of the Following Vehicle Equation. Given a freeway section and a pair of vehicles, the acceleration of the following vehicle at a given time t is given by:

$$a(i,t) = dx(i)^2/dt^2 = \text{lamda} * [dx(i+1)/dt - d(i)/dt] = \text{lamda} * [V(i+1,t) - V(i,t)] \quad \text{(a)}$$

(lamda is a Greek letter)

where: V(i+1,t): velocity of the lead vehicle
V(i,t): velocity of the following vehicle
a(i,t): acceleration/deceleration of the following car
lamda: driver sensitivity factor

Therefore,

if V(i+1,t) is > V(i,t), then a(i,t) is > 0 (acceleration)
else if V(i+1,t) is < V(i,t), then a(i,t) is < 0 (deceleration)
else, no change in the speed of the following car.

The speeds of the first pair of vehicles in the system are initialized to a certain value. For subsequent pairs during the following time intervals, (a) is used to determine the acceleration of each vehicle in the system.

The above model was further modified to reflect the fact that the following vehicle will accelerate or decelerate an amount which depends not only on the driver's sensitivity and the velocity of each vehicle, but also on the separation distance between the two cars [41]. To count

for this distance, a new factor, alpha, was introduced:

$$\text{alpha} = V(i,t) \,/\, [X(i+1,t)-X(i,t)]^k \qquad (b)$$

where k: is 1 or 2

Thus, the farther the separation distance and the smaller the speed of the following vehicle, the lower the value of alpha. The final acceleration/deceleration equation becomes:

$$a(i,t) = \text{lamda} * \text{alpha} * [V(i+1,t)-V(i,t)] \qquad (1)$$

(note that alpha is also a Greek letter)

2.b.2. New Position Equation. Given the initial velocity and computing the rate of acceleration in (1), the new position of the following car is computed from the motion equation:

$$X = Xo + Vo*t + 1/2*a*t^2$$

$$\text{i.e.,} \quad X(i,t+1) = X(i,t) + V(i,t)*dt + 1/2*a*dt^2 \qquad (2)$$

2.b.3. New Speed Equation. This is also simply given by the equation of motion:

$$V = Vo + at$$

$$\text{i.e.,} \quad V(i,t+1) = V(i,t) + a(i,t)*dt \qquad (3)$$

2.c. Special Difficulties with the Microscopic Approach. Since Car-following techniques use micro time intervals (seconds or fraction thereof) and require one-at-a-time processing of every single vehicle, the computing time is extremely excessive. It was reported that, during developments of some earlier micro models on the IBM 1620 II computer, that 5 seconds of computer time was required for each car for each second of real-time driving [41]. This means that in order to simulate ten cars for one hour, fifty hours of computer time would be required!

Although these estimates relate to very old computers and vast advancements and breakthroughs in computer technologies have been since

achieved, micro modeling of freeway networks with tens of thousands of cars for several hours with large numbers of computations, decision tests, and micro-analysis, is still a formidable task. Mannering [48, 1989] indicated that his work with freeway incident simulations were bogged down by the input and computational demand of comprehensive (microscopic and macroscopic) models. The impact models he used that were readily implementable provided output that was too general for adequate assessment; whereas those which provided adequately detailed output were "hopelessly complex" in terms of the required input and computations.

3. Mesoscopic Modeling

Several researchers proposed various combinations of microscopic and macroscopic modeling techniques, referred to here as a "mesoscopic" approach. They generally aimed at balancing the sophistication of micro models with the simplicity and low cost of macro models. Davila and Lieberman [38] developed LEVEL I, a hybrid macroscopic-microscopic traffic simulation model for this purpose. Chang, Mahmassani and Herman [36] developed MPSM, which is a combined discrete-continuous model for peak-period traffic flows. It seems to have been an effective method to stay above the extensive microscopic level of details while providing a somewhat acceptable level of sophistication.

It is assumed that combining macroscopic and microscopic approaches into a mesoscopic modeling methodology, coupled with effective use of computer technologies, and the theory of probability, would yield a new generation of better traffic simulation models.

REVIEW OF SOME TRAFFIC SIMULATION MODELS

FREQ10 is the most widely recognized macroscopic model. It has been applied over several freeway facilities [51]. Since 1968, this family of models have been under continuous development at UC Berkeley. FREQ traffic flow logic is based on the principle that bottleneck sections produce shockwaves when demand exceeds the capacity. The model is capable of simulating directional freeways and their ramps using ramp origin-destination (O-D) tables (as external demands). The model can be described as quasi-static [37, 1987] because changes in traffic demands can only occur at fixed time intervals -normally 15 minutes time slices.

The primary function of the model is to evaluate priority (high occupancy vehicle "HOV") and normal (single occupancy vehicle "SOV") entry control on directional freeways. It simulates a given (user-specified) system, evaluates its performance, and predicts the traveler response, all based on the *Highway Capacity Manual* [14] fundamentals.

The most powerful feature of the model is its ability to draw ramp metering optimization plans using linear programming techniques. The general format of the optimization objective function in the model is:

Maximize: Z= SUM(i=1,NORG) [Ci * FLOWi]

Subject to: SUM(i=1,NORG) [Aij * FLOWi] <= CAPj , j=1,2,..NSEC
FLOWi >= MINi , i=1,2, ...NORG
FLOWi <= MAXi , i=1,2, ...NORG

Where: Ci: the relative value associated with each ramp i
FLOWi: the input flow rate from the on-ramp
Aij: fraction of traffic from on-ramp i passing through section j
CAPj: capacity of freeway section j

NORG: number of the on-ramps
NSEC: number of sections
MINi: the maximum allowable metering rate at the on-ramp that produces minimum possible on-ramp flow rate
MAXi: the maximum expected flow rate from the on-ramp

FREQ On-Ramp Queueing. All on-ramps capacities (unless over-ridden by the user) are assumed to have a ramp limit of 1500 vph. The model can use different demand volumes at different time slices. However, demand and capacity are assumed to remain constant over the entire time slice. Therefore, the cumulative arrival and departure curves are linear over the time slice -even though this may not be true in reality. The released ramp volumes are considered direct "demands" on the freeway mainlanes.

Ramp time in queue is computed for all vehicles in the queue throughout the time slice from the horizontal distance between the arrival and departure curves on the queueing graph. The queue length is computed from the vertical distance between arrivals and departures. If ramp demand exceeds ramp capacity during a given time slice, excess demand is stored until the next time slice. As a result, no excess demand is allowed even if queue space becomes available during the present time slice. This is another violation of real-time flow processes.

FREQ Steps of Flow Process (Flow is processed downstream to upstream).

1. Initialize the system to a certain low demand.
2. For each time slice, generate the on-ramp volumes (from O-D) tables.
3. Compute capacity of each section:

$$c=2000*N*W*Tc$$

where: ci: capacity of section i
Ni: number of lanes in section i
W: lane width factor
Tc: truck factor at present time slice

4. For each subsection from downstream to upstream:

If demand > capacity, then

compute and store excess volumes (excess=demand-capacity)
compute speed of the shockwave (from flow-density curves)
compute resulting queue length (length=speed*time slice)

5. If demand <= capacity, then

if previous queue exists, compute reverse shockwave
else compute v/c ratio
{v/c ratio cannot exceed 1. The freeway reaches saturation
at v/c=1 and back-up queue will form upstream of that point}.

6. Compute the service volume:

$$SV = c * v/c$$

7. From the given number of lanes, the theoretical LOS is

determined (from 1965 HCM table 9.1).

8. Determine the operating speed So(i) (figure 9.1, 1965 HCM).

9. The average speed is computed from:

$$S(i) = So(i) - Sd(i)/10 * (1-Vi/Ci)$$

where: S(i): average speed of section i
Sd(i): design speed of section i
So(i): operating speed of section i
Vi/Ci: v/c ratio of section i

10. Compute travel time in the subsection i from:

$$T(i) = D(i)/S(i)$$

where: D(i): length of subsection

Many other macroscopic models were developed. The MACK.II model and the original MACK [33, 1987] are deterministic, macroscopic models that

are basically a set of conservation equations and corresponding set of speed-density equations. An evaluation of MACK.II was done by Derzko et al [39, 1983] who found it to contain certain instabilities and non-conformity to real-world data [51].

FREFLO model [57, 1987, and 33] is a further development of MACK.II. It was developed by Payne [53, 1979] as a macroscopic freeway simulation model. FREFLO simulates the flow of traffic on freeway networks using an aggregate variable formulation based on a modified analogy of fluid flow [57]. Initial work with this model revealed that FREFLO was limited in its ability to realistically simulate congested flows [57]. This problem was traced to the discontinuity in the flow-density relationship at the onset of congested conditions. Derzko et al [39, 1983] also found that FREFLO exhibited the same instabilities as the MACK.II model. Further development of the FREFLO appear to be underway by JFT and Associates, but no reports in the literature are found describing such effort [51].

ENHANCED FREFLO is an improved version of *FREFLO* developed by Rathi, Lieberman, and Yedlin [2 & 24]. The model still employs the basic conservation and flow-density equations of state, but uses the following dynamic speed-density equation:

```
V(i,t+1)=V(i,t)+dt*[-V(i,t)*(V(i,t)-V(i-1,t)/dxi
        -1/T*(V(i,t)-Ve(P(i,t)+a*(P(i_1,t)-P(i,t))/dxi]
```

and

```
V(i-1,t)=V(i,t)      if there is no adjoining upstream link
P(i+1,t)=P(i,t)      if there is no adjoining downstream link
                     (use either equation only)
```

where: V(i,t): speed at section i and time t

P(i,t): density at section i and time t
t: current point of time
i: current section
dxi: length of section i
dt: time step
T: relaxation coefficient
a: anticipation coefficient

FRECON2 is a dynamic macroscopic freeway model that can simulate the flow under normal and congested conditions [51, 1987]. FRECON was developed by implementing the scheme of FREFLO with further refinements. Neverthless, FRECON's ability to simulate severely congested flows was limited [57, 1987].

CORQ (CORridor Queueing) [33, 1987] is a macroscopic assignment model for allocating O-D demands to a time dependent traffic networks. According to May [51], CORCON has not been applied in practice since about 1980. It is anticipated that this model is a proprietor one [51].

UTCS-1 is a microscopic model in which the vehicle is treated as an individual entity as it traverses its path through the network. This model was later integrated into SCOT family.

SCOT (Simulation of COrridor Traffic) family of models is hybrid macro-micro combination which was introduced in the early 1980's [51 and 1]. SCOT is the synthesis of two previous models: UTCS-1 (Urban Traffic Control System-1) and DAFT (Dynamic Analysis of Freeway Traffic) with some modifications [33]. Applications and implementations of SCOT are not documented.

DAFT is a macroscopic simulation model of freeway ramps. Vehicles are grouped into platoons and consequently lose their individual identities. Platoons are processed along the freeway according to a prespecified speed-density relation. Review of the literature and

discussion with freeway modelers [51] revealed that no further work was done on the SCOT model.

ROADRUNNER is another macroscopic freeway model developed at the University of Toronto for Ontario Ministry of Transportation and Communications in 1978 [51]. There is no information about the application or validation of this model [51].

INTRAS (INTegrated TRaffic Simulation) is a microscopic, stochastic, Car-following model developed by KLD and Associates in the late 1970's [51]. INTRAS uses the network theory to inter-relate freeway and arterial traffic [33, 1987]. The program is very large and complex. A few control strategies are incorporated into the model, but it is difficult to allow for new control strategies (or modifications of the logic) because of the "closed" structure of the model [33].

Users of INTRAS have reported problems with some aspects of traffic behavior [33] such as vehicles that merge from acceleration lanes, vehicles at exit ramps, and the method of assigning destinations. Some of these problems relate to the complication of communications between vehicles across link boundaries. INTRAS also uses a constant value of 0.3 seconds to represent the reaction time of drivers [35, 1988] which slows the execution and introduces inflexibility. INTRAS also does not take into consideration the start-up delay of stopped vehicles.

Cohen and Clark [37, 1987] reported that INTRAS provided workable means of traffic operations consequences for freeway reconstruction projects. They also indicated that INTRAS was not completely operational. A considerable effort was required to adopt the model to their particular application on the simulation of reconstruction project

of the Theodore Roosevelt Bridge in Washington D.C. along I-66. The modelers were successful in "going around" the model's deficiencies and made some necessary modifications to it [37].

CARSIM (CAR-following SIMulation) is the most recent microscopic model which was developed (and still is under development) by Benekohal and Treiterer [35, 1988] using SIMSCRIP.II simulation language. CARSIM is intended to take into consideration the the shortcomings of INTRAS and to offer additional features for realistic simulation of stop-and-go conditions. The Car-following algorithm of CARSIM is basically a vehicle-advancing mechanism (similar to INTRAS) that facilitates the movement of vehicles from one point to another along the road [35]. Several acceleration or deceleration rates are computed for different situations, and the most suitable is selected for each vehicle at every time interval. After determining the proper acceleration or deceleration rate, the speed and the position of the vehicle are computed and the vehicle is advanced to its new position. The process is repeated for all vehicles in the system. Section [VII] compares the various features and the logic of CARSIM as a microscopic model.

CARSIM has been validated at both microscopic and macroscopic levels [35, 1988]. At the microscopic level, the speed change patterns and individual vehicle trajectories obtained from CARSIM were compared with those from field data. At the macroscopic level, average speed, density, and volume computed in CARSIM were compared with those from field data. According to the authors (3 and Treiterer) [35], validation process was satisfactory.

FRESIM (FREway SIMulation) is a microscopic simulation model

developed at Ohio State University [Nemeth] in early 1980's, but their has been no evidence that this model has been applied or validated in the real-world [51 & 57].

MPSM (Macro Particle Simulation Model) is a fixed-time (time slice) mesoscopic highway corridor simulation model developed by Chang, Mahmassani, and Herman [36]. It uses the established traffic flow relationships and the same conservation equation and flow-speed-density equations of state as described before to simulate the movement of traffic. It uses the following speed-density equilibrium equation:

$$V(i,t)=(Vf-Vo) * (1-k(i,t)/Ko)^a + Vo$$

Where: V(i,t): mean speed in section i during t-th time step
Vf: mean free-flow speed
Vo: minimum speed on the facility
K(i,t): current flow density
Ko: Maximum possible flow density

For interaction of vehicles, however, the model views traffic as discrete vehicle bunches (or macro-particles) that are moved according to local speeds as defined by local concentrations. New position of each macro-particle is found by computing the necessary advancing distance: d(m,t+1)=dt*V(i,t+1), in section i (m is the particle position)

Input to MPSM consists of the key physical and operational features of the highway facility and information about the usage level and commuter behavior.

MPSM generates individual entities at the on-ramp, but then later groups them into macro-particles each consisting of 5-10 vehicles. Once in the macro-particle, the vehicle platoon is treated as a one piece of log advancing in a stream of water. The wood and water are analogous in their flow to the compressible fluid flow. Details of the merging

maneuvers are beyond the capabilities of the model and, thus, a simple deterministic queueing approximation is used to handle merging.

MPSM's prime feature is the time-series component that allows the investigation of commuter decision dynamics and their interrelations with respect to time dependent congestion patterns.

LEVEL I model [38] is a hybrid macroscopic-microscopic model designed to evaluate transportation system strategies over street (not freeway) networks. Automobiles, buses, carpools, and trucks are modeled individually. The simulation is based on the "event-oriented" methodology of GPSS simulation language. Particles are moved intermittently rather than at every time step. Thus, the model is microscopic in creation of entities, and macroscopic in its treatment of them [38]. The model is still undergoing further development within TRAF group of models.

THE SYSTEMS MODELING STRATEGY

The establishment of a basic building block for a mesoscopic on-ramp simulation model requires that *an integrated and Holistic "systems" analysis and investigation* be undertaken. This is because, contrary to the common perception, systems modeling is not computer programming, although systems modeling may require computer implementation. The modeler has to be a good computer programmer as well as a systems analyst who knows how to analyze and understand the behavior of the system or phenomenon he is modeling. We start by devising a strategy for investigating, understanding, and then modeling the system.

A mesoscopic approach is suggested: a combination of microscopic system decomposition/analysis, mesoscopic model building and verification, and macroscopic model validation. The research plan is to apply Shannon's system decomposition methodology [60, 1975] in conjunction with "properties-of-the-whole" concept by Lendaris [45, 1986] (both are discussed below).

Shannon's decomposition methodology begins by making a distinction between isomorphism and homomorphism. A model's degree of isomorphism is the relative degree of similarity between the model and the thing it represents. According to Shannon, if the model is completely isomorphic, then a) there is a complete one-to-one correspondence between the elements of the model and the item it represents; and b) the exact relationships between the elements are perfectly preserved. In reality, no model can be totally isomorphic.

A model's degree of homomorphism on the other hand is the relative degree of similitude between the model's <u>form</u> (and not necessarily structure) and the thing (the system) it represents [60]. Most models are therefore homomorphic.

To build a homomorphic model of a system, the modeler needs to go through four basic steps of model conceptualization [60]: decomposition, simplification, abstraction, and recomposition.

Given a certain context and specific purpose of the model, first, the system under consideration is broken up into a number of smaller independent (or nearly independent) parts.

Second, the parts and their internal and external relationships are analyzed one-by-one eliminating irrelevant issues, truncating trivial

details, focusing on key elements, and proposing (or assuming) simpler relationships whenever appropriate (i.e., linear rather than non-linear, uniform instead of non-uniform, etc.).

Third, an abstraction of these parts and their relationships is performed to redefine them and manifest their relationships in a manner suitable for building a physical, mathematical, or computer model; a manner that represents the essential properties and qualities of the system but not necessarily in the same form or detail as the original.

Finally, the parts of the system (now the model of the system) are recomposed yielding an "approximate" model of the system --a homomorphic model.

The above methodology suggested by Shannon seems to be a good way to build a microscopic model of a traffic system by breaking it up into its smallest molecular units (or subunits assuming that the system as a whole is a unit) and then recomposing it back into one unit. But the above four-step process seems to contradict the properties-of-the-whole concept in systems theory. That is, the whole is (sometimes profoundly) greater than the sum of its parts.

Lendaris [45] points out that "there are two aspects to this concept: 1) there is no way for an observer to deduce the attributes-of-the-whole by studying the parts and their individual operations, and then somehow 'adding' these up; and 2) the parts, when jointly operating according to some organizing principle, can collectively do something which is 'greater' than a simple collection of their individual uncoordinated operations."

This proposition poses a vexing dilemma in traffic systems

applications. The properties-of-the-whole concept seems to question the microscopic approach and prefer a macroscopic view instead. It seems to challenge the modeler's ability to decompose the system into micro-pieces and then recompose them into a "valid" model without losing significant part of its fundamental qualities. Does this concern limit one's ability to model the traffic microscopically? Is it behind the limitations of some microscopic traffic models?

The systems approach can help resolve this conflict by requiring systems analysts to acquire both "holistic" as well as "wholistic" appreciation of their systems. Lendaris defines the wholistic perspective as that which perceives the system and its attributes-of-the-whole in its (external) environment as one unit. The analyst acquires an appreciation for the whole not by studying the parts, but rather by studying the system as one entity. The holistic perspective on the other hand views the system as a collection of subunits that function together as a whole. The analyst *studies the subunits and their relationships, while being "mindful and appreciative" of the whole as a Gestalt* (*).

Holism philosophy is then the solution. It can be directly applied to systems modeling. System parts (subunits) could be investigated and modeled by "attending the parts while being mindful of the whole." The first step is to distinguish the various levels of subunits and their inter and intra-relationships. Then, simplifying assumptions and

---

(*) Gestalt, as defined by the American Heritage Dictionary, is a physical, psychological or symbolic configuration or pattern so unified and interconnected as a whole that its properties cannot be derived solely from its individual disjointed parts.

abstractions about these relationships are made.

To complete the holistic analysis, the researcher takes a different stance and studies the system relationships with any supra-system relative to it. A mathematical model is next composed. The mathematical model can then be coded -if necessary- into a computer model.

Figure 13 is a tree representation of the hierarchy of the proposed approach. The focus is on the on-ramp system. From the system level at B, ramp lanes, lights, queues etc. at level C are subunits of the on-ramp system. I-8 segment at level A is a supra-system of which the on-ramp is itself a subunit. The ranking of supra-system/system/ subsystem (or subunits) is relative to the viewing level, as suggested by Lendaris.

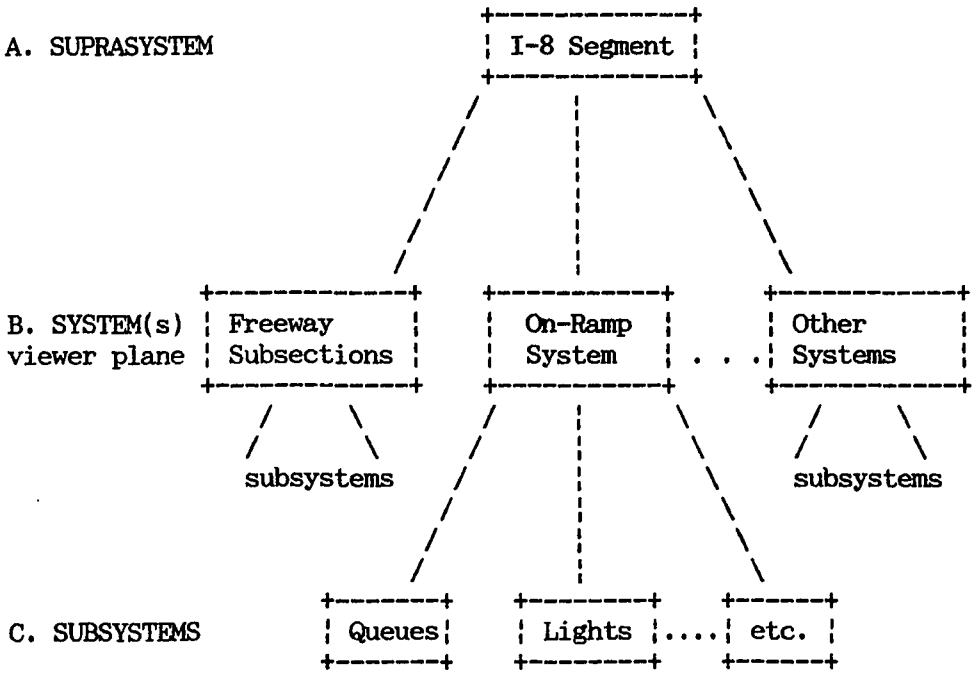


Figure 13. Tree Representation of Systems Hierarchy and the Decomposition into Subunit Blocks.

Pareto law was also useful in modeling the system. According to the Pareto law [60], in every system there exists a vital few and trivial many. It was the *vital few* (i.e., only the key system subunits and components) that had to be identified, understood, and used to manage the problem and furnish the ingredients for the model. The vital few are figured out while studying the reference behavior mode of the real-world system (discussed in later chapters). Trivial details that might weaken or complicate the model were filtered out and ignored. Key system components were studied and modeled microscopically to preserve the accuracy, while less important components were studied and modeled mesoscopically to eliminate redundancy and maintain simplicity. For example, the basic building block of the on-ramp had a microscopic and mesoscopic molecular parts. The part which generates vehicular arrivals has a microscopic structure. Ramp queues have a mesoscopic structure where only queue lengths and delays are examined every minute. The internal dynamics of the queue were considered trivial and thus ignored.

Further details on the system analysis and the design of a model are given in the following chapters. The reference behavior mode (RBM) of the system is established via field observations, data collection and analysis of the system. Subsequently, a computer modeling procedure is devised and it encompasses model input specification, model output validation, and a validation criteria.

---*---*---

CHAPTER IV

ESTABLISHMENT OF A REFERENCE BEHAVIOR MODE OF THE METERED FREEWAY SEGMENT ALONG INTERSTATE-8 FOR THE PURPOSE OF MODEL BUILDING AND VALIDATION

CHAPTER PREVIEW

This chapter describes the collection, transformation, reduction, qualification, and analysis of numerical data and visual observations of a ramp metering system. The objective is to understand the system well and express knowledge about it in an easy-to-infer-from and easy-to-refer-to format, called here the "reference behavior mode" or RBM (*), so that a good model of the system could be built and validated. The work is one phase in the larger systems modeling project that includes a review of the ramp metering theory, a review of existing traffic simulation models, and the development of an on-ramp simulation model and computer program.

The system lies in a 5-mile, heavily-congested, inter-connected segment of Interstate-8 in San Diego, California. The sought RBM form encompasses a quantitative representation of the system's time dimensions and physical components; mathematical abstraction of the

---

(*) RBM abbreviation will be used throughout this paper instead of the expression "reference behavior mode".

system's internal and external relationships; and qualitative characterization of its overall behavior. Maps, graphs, tables, etc., are presented for illustration purposes. The resulting RBM is used to infer concepts for design of the model (model conceptualization), derive input parameters for simulation runs (input formulation), and evaluate the output of such runs (model validation).

Three approaches were available to establish the RBM: theoretical, empirical, or a combination of both. Theoretical establishment of the RBM is accomplished by recognizing a well-established theory regarding the system and then examining the real-world and comparing its behavior with the facts of the theory.

In the absence of a widely-accepted theory, or if the researched phenomenon or system has special qualities or could not be explained well by the theory, the RBM could be established (empirically) by an examining of the real-world system without much predetermined knowledge about it. Quantitative measurements and qualitative observations are made and then transformed into an appropriate knowledge base which is the RBM. Since the ramp metering theory is still under development and since the system under study has some unique geometric and traffic characteristics, this study embarked more on the empirical approach.

Key system components like vehicular arrivals, service times and queueing activities were identified and chosen for study in details. Observations about the system were analyzed and, abstracted, and when possible, transformed into mathematical relationships and probability distributions that could be implemented on the computer. Throughout the process, a balance between *microscopic* system analysis and decomposition

and *macroscopic* system modeling and representation (for simplicity) was maintained. This resulted in an overall intermediate *mesoscopic* approach.

The research work may seem long winded, but reliable models can only be built and validated by a well-established RBM of the system that is representative and accurate as well as clear and simple.

DESCRIPTION OF THE RESEARCH APPROACH

The system under study lies in a large urban area. Urban freeway systems normally have large capacities that serve complex networks and diverse motorists. Despite the numerous developments in the traffic flow theory, the dynamics of such systems are still very difficult to fully understand, let alone to perfectly model and control.

The research work has progressed slowly because of the size and complexity of the problem. The principal approach adopted for problem reduction was the Pareto law -a simple but very important systems concept. It states that each system is composed of some vital few and trivial many [60]. The focus throughout this research was on such "vital few" system elements that determine the overall behavior of the system. Only these elements are recognized, analyzed and will eventually be modeled. The freeway segment is a suprasystem that is composed of two systems:

1) On-ramp system, and
2) Subsection (mainlane) system.

Flow at exit ramps is assumed to be not difficult to model and, hence, exit ramps are not included in our analysis of the RBM. The focus

of this study is on the on-ramp system. Although analysis of the flow on the main lanes is critical in establishing the RBM for the overall suprasystem, the primary emphasis was placed on the activities at the on-ramp system for two reasons:

1) Entrance ramps are the center of action providing the critical link between street traffic and the freeway. The ramp light (meter) is the control element which determine forward flow onto the freeway and backward queueing into the ramp storage lanes and surface streets.

2) Each entrance ramp has some distinct characteristics in geometry, patterns of arrivals, interaction with surface streets, etc., while the freeway traffic stream follows a close-to-uniform global pattern.

Analysis of the random behavior of the system was ignored only when deterministic analysis would be as adequate. Field observations have been used to formulate assumptions about the nature and behavior of each significant system activity. Whenever an activity was clearly random, probability distributions were proposed to represent it. Collected data have been used to estimate the frequency, duration, variability, and the degree of uncertainty for the following activities:

1. On the ramp: inter-arrival times, discharge rates, queue lengths, and queue delays.
2. On the freeway: mainlane speeds and mainlane average volumes.

A number of system activities exhibited erratic patterns and there was no basis for suggesting any particular probability function to represent them. For example, the rate and pattern of arrivals of trucks or commercial heavy vehicles did not exhibit a clear or consistent pattern. Either a uniform distribution or deterministic views could be

assumed for such occurrences.

The scope of discussion is limited to two typical ramps: 70th Street, which is a loop-shaped ramp with limited storage area, that experienced heavy demand and queue blockage and overspills; and College Avenue (North), a ramp that has large storage capacity and experienced no queue overspills. Analysis of the remaining ramps is repetitive and presented only when necessary.

Some of the difficulties which were experienced during data collection and transformation are discussed. Local intersection problems near entrance ramps are discussed and examples of such problems are presented along with some concluding remarks.

LOCATION & PHYSICAL DESCRIPTION OF THE STUDIED SYSTEM

The freeway segment, where data were collected, is a 5-mile stretch of a 4-lane westbound flow along Interstate-8 in San Diego, California. Figure 14 is a map of that segment and surrounding area. The segment has typical highway geometrics, is under heavy demand and is already metered, making it ideal for real-world testing of model development and recommendations. The segment passes through densely populated urban area providing a major east-west link across San Diego for commercial and residential traffic. I-8 is also used by transient traffic between I-5 (west) and Imperial Valley and Arizona (east). Within the segment, there are eight entrance ramps and five exit ones from Jackson Drive (east) to Waring Road (west).

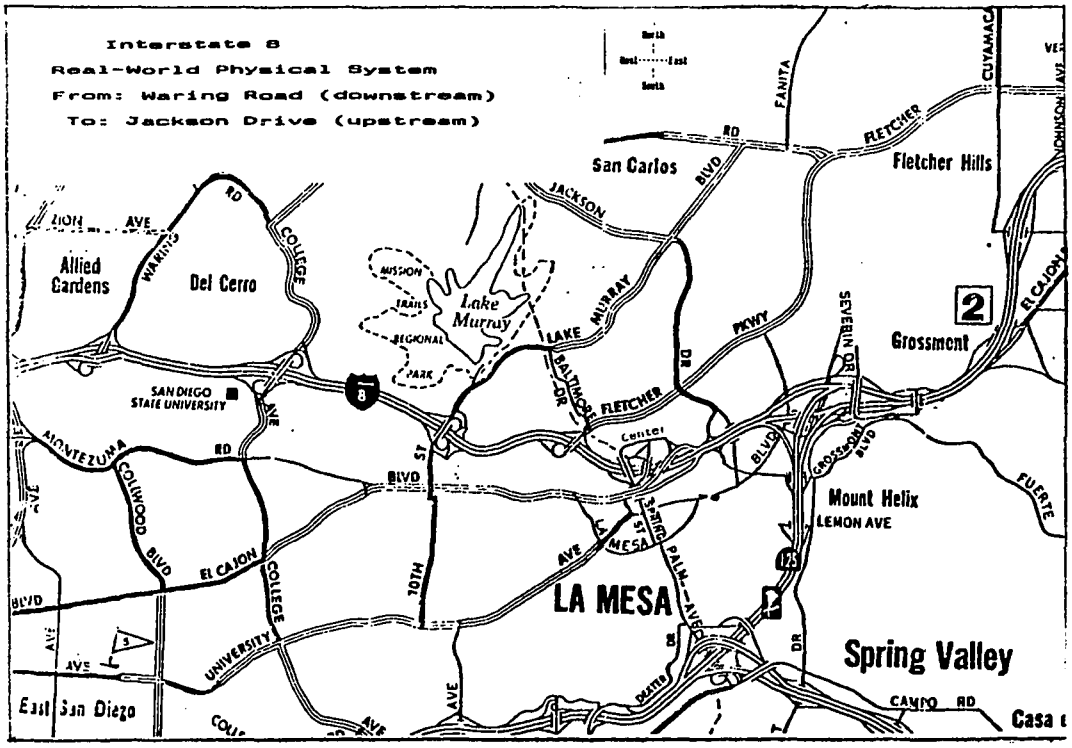


Figure 14. Map of the I-8 Segment of the On-Ramp System

A simplified system diagram is given in figure 15 It shows the on-ramp and off-ramp sequence and all freeway subsections. It also gives the distances between intersections, merge points, and diverge points. All distances are measured along the freeway center line. Merge and diverge points are never well-defined. Theses points were estimated based on repeated visual observations of most frequently chosen points of merge and diverge by drivers.

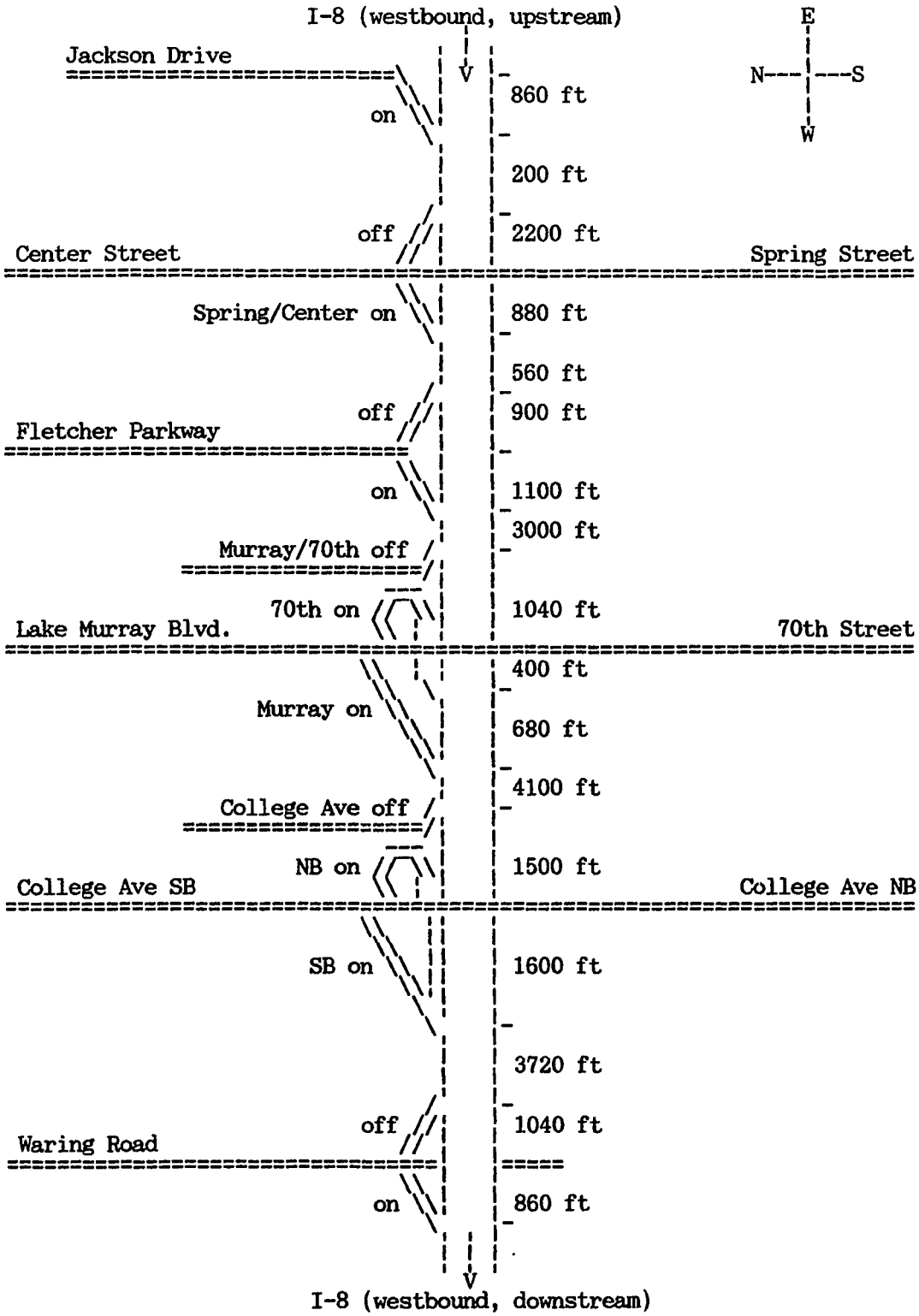


Figure 15. Diagram of I-8 Segment, On & Off Ramps, & Subsections.

Video taping the segment was the primary means of collecting and recording the observations. The duration of observations extended over a period of three weeks from Tuesday August 22, 1989 to Friday September 8, 1989. Only westbound morning traffic for the peak period from 6:30 to 8:42 a.m. was included in the analysis. Traffic was light before and after this period. Table III lists ramp IDs, postmile information, and dates of observations.

TABLE III

RAMP ID, POSTMILE INFORMATION & DATES OF OBSERVATIONS

Intersection/Entrance Ramp	Postmile	Day & Date Videotaped (*)
1. Waring Road. (Downstream)	7.058	Tu 8/22/89 (only)
2. College Avenue South	8.269	We 8/23/89 + Fr 9/8/89
3. College Avenue North	8.269	Th 8/24/89 + Fr 9/8/89
4. Lake Murray Blvd	9.590	Fr 8/25/89 + Tu 9/5/89
5. 70th Street	9.633	Mo 8/28/89 + Tu 8/29/89
6. Fletcher Parkway	10.513	We 8/30/89 + We 9/6/89
7. Spring Street + Center St	10.934	Th 8/31/89 + Th 9/7/89
8. Jackson Drive (Upstream)	11.600	Fr 9/ 1/89 (only)

Note: (*) According to Meyers (a), in transportation system analysis, "suitable historical data for more than one time period are a luxury few planners enjoy." Admittedly, two days of observations on each one of the eight ramps are insufficient to draw final conclusions about the whole metering system over time (weeks, months, years, etc.) or about ramp metering in general. However, for the purposes of this work, limited data is accepted for the lack of better resources. Conclusion drawn in this research are limited accordingly.

(a) M. Meyers, *Urban Transportation Planning*, 1984.

DISCUSSION OF FIELD WORK & DATA COLLECTION

1. Sources of Data

Five different sources of data were available to this study. Each provided an important dimension of information.

a. Video taping the intersections, entrance ramps, freeway merge points, and mainlane activities -simultaneously whenever it was possible. About 45 hours of recordings were obtained. The major emphasis was placed on observing the interaction between entrance ramp traffic and surface street traffic. Only a minor emphasis (at this point of the research) was placed on the interaction between on-ramp and freeway traffic.

b. On-site manual notes and visual observations.

c. Computer data on mainlane speeds/volumes gathered by loop detectors.

d. Historical on-ramp and freeway data from 1985 to 1989.

e. Feedback from staff and personnel of the Traffic Systems branch in San Diego about overall perspectives, system administration and field operation and some on-line control algorithms.

2. Field Work

The field work consisted mainly of video recordings over a period of three weeks, each one of the eight ramps was videotaped twice on two different days. All recordings were done in the morning from 6:30 a.m. to 8:30 a.m. All field work was largely performed by the researcher with the help of an engineering technician. The crew sometimes expanded to four people to help with difficult-to-monitor intersections such as Fletcher Parkway and Spring Street and also to help in observing several

ramp activities simultaneously. The researcher always stayed with the video camera and took notes from that spot. Other crew members helped in setting the camera, taking occupancy counts, observing visually-blocked sides of the ramps, observing mainlane traffic, doing simultaneous reporting (by radio) on the status of other ramps, taking road measurements, taking still pictures and so forth. All crew members, video equipment, and vehicles were provided by District 11 of the California Department of Transportation in San Diego.

An industrial video camera system was used (a Panasonic Digital 5000 HD with 12x zoom ratio and 1:1.6 lens). All recordings were done on VHS format with extended speed. 45-hours of field observations were gathered. The camera was always placed at a high point facing the on-ramp traffic. The many hills alongside I-8 provided some excellent surveillance spots and a good perspective of the ramps.

Except for Fletcher Parkway on-ramp (where even the nearest hill was too far to obtain a good perspective) and Spring Street (where the ramp had a steep slope, too many curves, trees, and several bridge structures blocking sight), the camera captured full view of the ramp from the ramp metering limit line (RMLL) to the first surface street intersection. This positioning of the camera made it possible to to trace a vehicle as it approaches the ramp, passes over the arrival reference point, joins the queue, progresses through the queue, and then departs at the RMLL. It also made it possible to observe queue overspills into the nearby intersections.

3. System Time Dimension

a. 11-Time Slices. It was noted that throughout three weeks of observations at all ramp locations, the morning traffic peak-period lasted from 6:30 to 8:45 a.m. The intensity of traffic activities throughout the peak-period was not uniform. Therefore, this period was divided into eleven time intervals, each 12-minute long, covering a total of 132 minutes from 6:30 a.m. to 8:42 a.m. (or 133 minutes when 8:42 is included). Statistical parameters such as mean and standard deviation of inter-arrival times, queue lengths, and queue delays were computed for each 12-minute time slice.

The 12-minute step size seemed to appropriately delimit the regions of variation in arrivals and queue data over time. The 12-minute time slice was small enough to achieve microscopic accuracy in system measurements and analysis and big enough to maintain macroscopic simplicity for systems modeling. Compatibility between this step size and the standard data block size of the mainframe computer at the Traffic Surveillance Center's was maintained. All computer records at the Center can be grouped into 12-minute parcels.

b. 133-TOTM System Status Checks. A top-of-the-minute (TOTM) check of system status mechanism was employed to monitor the status of metering rates, queue lengths, and queue delays. At each TOTM, a snap-shot was taken to read the metering rate (from computer records), measure the length of the queue (by visually counting the number of vehicles in the queue), and compute the queue delays. Queue delays were measured in the following fashion: at each TOTM, the first arriving vehicle at the end tail of the queue was followed until it departed at

the RMLL. Total time spent in the queue was considered the total delay time for that vehicle and that particular TOTM. Queue lengths and delays will be discussed later in detail.

4. Management of Data and Field Observations Problems

Field observations with respect to time headways (inter-arrival times), TOTM queue lengths and TOTM queue delays were transformed from visual videotape observations into usable numerical format. TOTM metering rates were transferred from the mainframe computer to a PC. Some interpolation and extrapolation work was needed to make up missing points. For example, each set of data has one or more minutes of discontinuity due to pauses to change the VCR battery.

The distant location of the camera from some of the ramps made it next to impossible to "visually" determine the number of passengers in each vehicle. Video observations on car occupancy were based solely on the usage of preferential lanes rather than actual car occupancy. This meant that, regardless of the actual number of passengers, a vehicle was considered high occupancy vehicle (HOV) if it had used the HOV lane. Similarly, it was considered single occupancy vehicle (SOV) if it had used the SOV lane. Since HOV lane violation rate is small (less than 8%), videotape observations on HOV/SOV rates are at least 92% accurate.

HOV/SOV monitoring problems are not unusual in highway studies. A U.S. DOT study [65, 1990] on HOV lane surveillance cited similar difficulties. It concluded that: "Videotape reviewers cannot currently identify the number of vehicle occupants with enough certainty to support citations for HOV lane occupancy violations."

DATA TRANSFORMATION, ANALYSIS OF DATA AND INTERPRETATION OF OBSERVATIONS

1. The On-Ramp Basic Building Block

Entrance ramps are similar -but not identical- in their basic structure. Typically, each on-ramp has three structural components:

a. Traffic source generating entities (vehicles) onto the ramp.

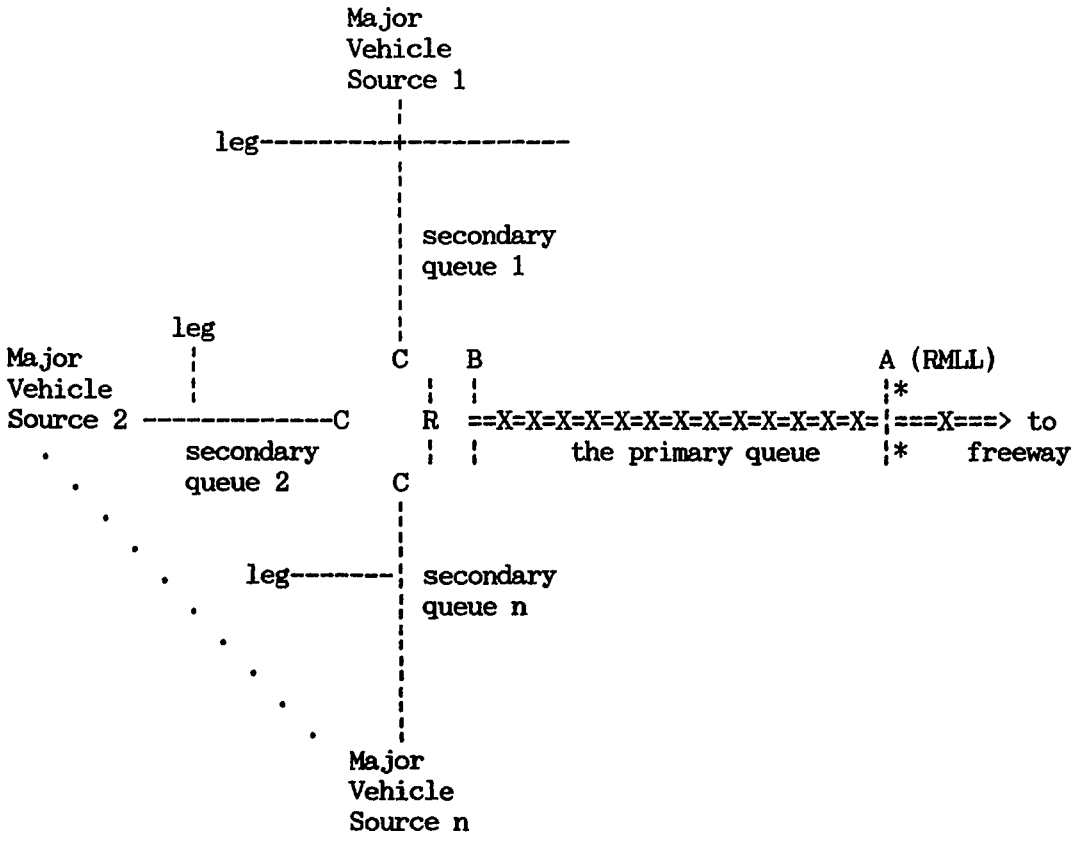
b. Queue storage area with certain capacity.

c. Meters. Those are viewed as single or parallel servers.

Many field observations showed that vehicular queues at each on-ramp were physically split at certain points into primary and secondary queues. As figure 16 shows, a primary queue is a vehicle storage line on the lane of the ramp. A secondary queue(s) is an additional storage on the surface streets (*). Primary and secondary queues will be discussed later.

Four reference points were needed to describe these queues. The reference points are imaginary lateral lines across the on-ramp approach lanes.

(*) Although queue storage on surface streets is highly undesirable and does violate the rules for ramp metering -as stated in the Traffic Control Systems Handbook (FHWA-IP-85-11)- it frequently occurred on at least five of the eight ramps (Waring Road, Lake Murray Boulevard, 70th Street, Fletcher Parkway, and Spring Street). Arguments for and against surface street storage are discussed elsewhere in a chapter two by the author.



Legend:

1) A, the vehicle departure reference point at the ramp metering limit line (RMLL).
2) B, the primary queue capacity limit. A point of maximum storage.
3) R, the arrivals reference point for the primary queue.
4) B-C, Point of split (a gap) between primary & secondary queues.
5) C, the beginning (head) of the secondary queue(s).
6) X, vehicles.
7) *, ramp lights (meters).
8) =, primary queue lanes.
9) -, secondary queue lanes.

Figure 16. The Real-World (RBM) On-Ramp Basic Building Block.

2. Vehicular Arrivals & Inter-Arrival Times (Time Headways)

A computer program was developed to capture this item from the videotapes. The computer time and tape time were synchronized. An arrival reference point (a point of entry for the arriving vehicles) was specified for each ramp. This point is an imaginary lateral line across the ramp lane upstream of the tail of the queue. The "P" key is pressed every time a vehicle crosses over the reference point and joins a queue that is only partially-full. If the queue is at or near capacity, the "F" is pressed to indicate an arrival to a full queue.

At each key stoke, the computer program reads the current time (which identical to the time stamp on the tape), computes the interval since last entry, and then records the time in character and cumulative seconds format, records the interval in seconds, and records the queue status ("P" for partially full, "F" for completely full queue). Table IV shows a typical segment of arrivals file. Once all the inter-arrival times were recorded, the 12-minute arrival rates were computed and grouped into 11 time slices from 6:30 a.m. to 8:42 a.m.

TABLE IV

SAMPLE INTER-ARRIVAL DATA (HEADWAYS) FROM 70TH STREET

Arrival Time	Cumulative Time (sec)	Interval (sec)	Primary Queue Status (P= Partially-full F= Full to capacity)
7:57:15	28635	19	P
7:57:18	28638	3	P
7:57:25	28645	7	P
7:57:38	28658	13	F
7:57:44	28664	6	F
7:57:48	28668	4	F
7:57:54	28674	6	F
7:58:20	28700	26	P

2.a. Choice of Stationary/Moving Reference Point for Arrivals. *An arrivals stationary reference point* was used to count vehicle arrivals. A stationary reference point is an imaginary lateral line across the approach lane of the on-ramp. The best position for such point is point "B" at the end (limit) of the primary queue in figure 16. This point provides an easy and consistent way to register arrivals as vehicles pass over it. However, the stationary point causes some distortion to the measurement of arrivals to the end of the queue itself. That is, as the queue changes in length, time varies between the moment when a vehicle crosses over the reference point and when it actually joins the tail end of the queue. If the queue is short, there will be a long time lag. If the queue is long, time lag will be short. Times between arrivals that are measured from a stationary reference point do not count for such variations. But if a stationary point is used, the prospective model could estimate such time lags based on the difference between the maximum queue length and instantaneous queue length.

If arrivals change the lane on which they arrive originally, after they had been recorded, they cause distortion to the data. This problem was solved by choosing a reference point close to the tail of the queue whenever possible. Note that if the stationary reference point is chosen inside the queue limit and the queue backs up over it, it would have to be moved further upstream. This and the problem of time lags prompted thoughts about a "moving" reference point for arrivals.

A Moving reference point is an imaginary, non-stationary, lateral line located at the end of the *dynamic queue*. It moves backward or

forward as the queue expands or shrinks. Arrivals are recorded immediately upon joining the queue. This method solves some of the time lag problems encountered above for stationary points, but it distorts the true times between arrivals. When the reference point is stationary (at a constant distance from the upstream source), true intervals are correctly measured. When the point is moving, time variations are added to the true intervals between creations of entities. Times between arrivals are affected now not only by how they were generated at the upstream source of vehicles, but also by the current location of moving reference point.

A moving reference point also causes problems for vehicles arriving in platoons. Often, a platoon arrives as one long mass of vehicle. It also stops as one unit at the queue. The moving reference point would have to be shifted across the platoon to enter all the cars at once! Measuring arrivals this way is inaccurate and unreliable, and hence NOT recommended and was not used.

2.b. Description of the Distribution of Inter-Arrival Times.

Analysis of headways and their distributions was done by taking one ramp at-a- time and studying it at time slice 12 (6:30-8:42). This time slice was chosen because it contains *all* data points and more information could be derived from it.

Frequency distribution diagrams for headways were constructed for all eight ramps for this time slice. Inspection of the data distribution in these diagrams resulted in the following conclusions:

1. Each headway distribution is skewed to the left; has a very high mode at an interval time of 2 seconds; drops rapidly near 4, 5, & 6 seconds;

continues towards higher values with more moderate, negative slope and then gradually along a very long and thin "tail". Maximum tail values varied from as low as 44 seconds (at College Avenue S.) to as high as 106 seconds (at College Avenue N.).

2. The second and third highest frequency cells were 1 second & 3 seconds. It was further noticed that the weights of these cells were affected by the arrivals type. Arrivals were divided into three types:

Type APQ: Arrival to Always-Partial Queue. Here, all arrivals (100%) joined a primary queue which was always only partially-full (never reached saturation). No secondary queues were ever present. Examples of this type of arrivals occurred at College Avenue N. and College Avenue S. For this type, the weight of the "1" cell was higher than "3" cell (see figure 17).

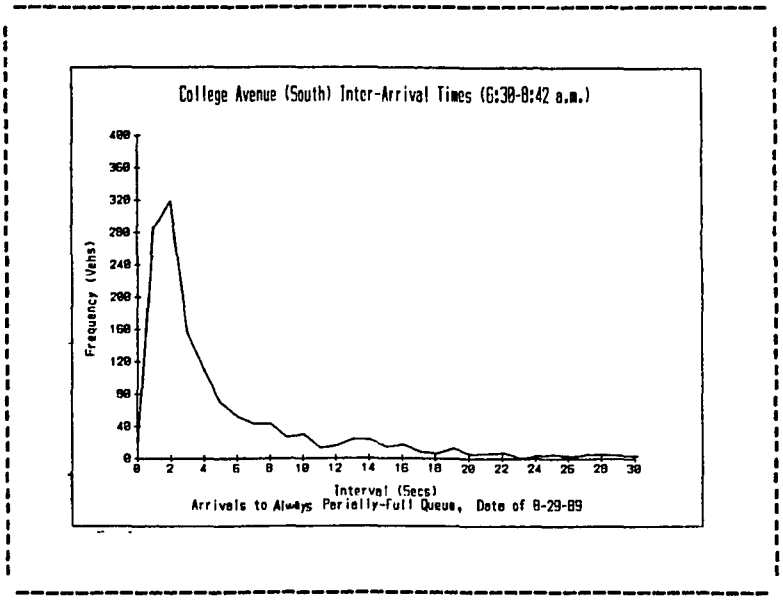


Figure 17. APQ Type of Arrivals at College Avenue South

Type MPFQ: Arrivals to A Mostly Partially-Full Queue. Here, more than 50% of arrivals joined only a partially-full queue. The secondary queue was present (ON) less than 50% of the time. Examples of this type occurred at Spring Street. For this type, the "1" cell was also higher than "3" cell (see figure 18).

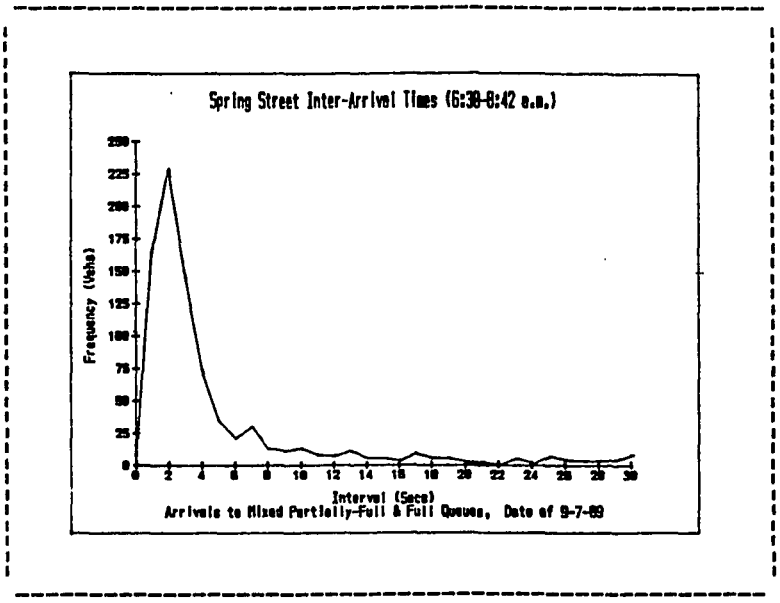


Figure 18. MPFQ Type of Arrivals at Spring Street.

Type MFQ: Arrivals to a Mostly-Full queue. Arrivals of this type indicate that more than 50% of arrivals joined a full (saturated) primary queue. The secondary queue was present (ON) more than 50% of the time. Examples of this type occurred at Fletcher Parkway and 70th Street. For this type, the weight of cell "3" had higher values than cell "1" (see figure 19).

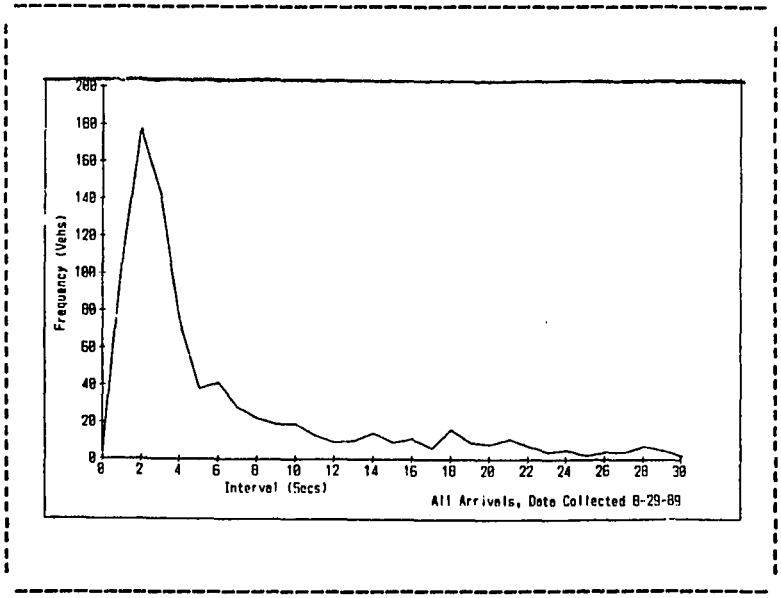


Figure 19. MFQ Type of Arrivals at 70th Street

2.c. Statistical Parameters of Inter-Arrival Times. Statistical parameters of inter-arrival times (inter-arrival times are column 3, table V), were computed and tabulated. Table V shows the predominant primary queue status, the number of arrivals, and the mean and standard deviation of inter-arrival times for all eleven time slices at 70th Street. Traffic mostly encountered partial (P) primary queue, except for the periods of 6:30-6:42, 7:18-7:30, 7:30-7:42, and 7:42-7:54 a.m. where it encountered mostly-full primary queue (Table V).

Data in table V tend to support the hypotheses that primary queue arrivals are controlled more by the availability of space in the primary queue itself than by the true pattern of arrivals. Direct arrivals to partial queues exhibited high variation. The mean value for time between arrivals for "P" status ranged from 5.82 to 11.78 seconds (Table V.a). The standard deviation was also consistently high. On the other hand,

high variation in inter-arrival times seemed to have been absorbed by the secondary queues whenever they existed ("F" queue status).

The mean value of inter-arrival times to full queues ranged from 9.77 to 10.57 seconds (Table V.b). The standard deviation assumed lower values indicating lesser variation. This was particularly evident when the queue was full for several minutes. Longer uniform intervals are due to the "blocking" that occurred at the entry points to the primary queue. Vehicles had to "wait" at secondary queues upstream of the primary queue until a space was available in the primary queue.

TABLE V

INTER-ARRIVAL STATISTICS FOR VEHICLES JOINING MOSTLY-PARTIAL OR MOSTLY-FULL PRINCIPLE QUEUE
At 70th STREET

Time Slice	Time Period	Occurrences (veh)	Queue Status	Mean (sec)	Standard Deviation (sec)
a. Arrivals Which Encountered Mostly-Partial "P" Primary Queue					
2	6:42-6:54	49	P	11.78	18.55
3	6:54-7:06	69	P	9.77	16.96
4	7:06-7:18	43	P	11.14	15.17
8	7:54-8:06	61	P	8.79	9.86
9	8:06-8:18	105	P	6.65	11.94
10	8:18-8:30	124	P	5.82	8.74
11	8:30-8:42	116	P	5.92	8.72
1-11	6:30-8:42	575(*)	P	7.76	12.43
b. Arrivals Which Encountered Mostly-Full "F" Primary Queue					
1	6:30-6:42	23	F	9.87	5.76
5	7:18-7:30	63	F	10.49	9.12
6	7:30-7:42	69	F	10.52	10.37
7	7:42-7:54	68	F	10.57	9.30
1-11	6:30-8:42	298(**)	F	10.36	8.66

Notes: (*) All P occurrences
(**) All F occurrences

Queue blockage was influenced by space availability in the primary queue. Space availability was determined by the metering rates and the

dynamics of the primary queue. Blocking due to street signals was minimal because when one secondary queue was blocked, another was not.

3. Metering Rates (Service Times)

Metering rates are also referred to as service times. Service time is the amount of time, in seconds, which is necessary to process one vehicle by the meter at the ramp meter limit line (RMLL). Parallel servers or 1/2 service time can be employed to process more than one vehicle at-a-time. An on-line algorithm runs the on-ramp meters along I-8 in San Diego. It receives input from the freeway mainlane loop detectors on the volume and density of traffic on the freeway. Then, it uses that information to set the red-green light cycle on the on-ramp, and thus, modulate the influx into the freeway. Service times may be measured by one of two methods:

1. Explicitly by direct reading of the length of one full, red-green cycle of the ramp meter (light) or;
2. Implicitly by calculating the time interval (headway) between the front bumper of a leading car and the front bumper of the following car as they pass over the RMLL.

The second method accounts for variations in driver reaction to the light. Since the dynamics of the queue are influenced by the driver's response to the meter and since these measurements were intended for a simulation model, service times were computed according to the second method. However, there were some gaps in the data for 70th Street, Fletcher Parkway and Spring Street. These gaps were filled using the first method.

Service time is a completely microscopic parameter (i.e., may change every one second). A simplifying determination was made to use a mesoscopic top-of-the-minute (TOTM) time interval check instead of the one-second interval. It is assumed that each measured service time at the TOTM (e.g., 07:26:00, 07:27:00, 07:28:00, etc.) will represent all service times during that minute (each observed service time was considered as an average for the entire minute). As a result, for each ramp observation, there were 132 measurements of TOTM service times (from 6:30 a.m. to 8:41 a.m) plus one reading at 8:42, making it total 133 minutes. Average service times for all 8 ramps were also computed and macroscopically grouped into eleven 12-minute time slices from 6:30 to 8:42 a.m. These parameters are used directly in the prospective model.

4. Queueing Dynamics

4.a. Primary and Secondary Queues. Primary queues can be fully modeled because they are physically contained and easy to observe. On the other hand, detailed modeling of the secondary queues would weaken the prospective model because of the diversity and uncertainty of secondary processes. Secondary queues often have multiple "legs" (refer to figure 16) that channel traffic into and out of the queue in a random fashion. It was concluded that in order to keep track of all traffic activities at the secondary queues, a great deal of personnel, equipment, and time resources must be provided. But even if resources were available, it would have not been cost-effective to employ them at this point.

A compromise was needed to contain the problem and the concept of "secondary" queue was conceived and has been adopted. Secondary queues in this study are an ON/OFF binary parameter. "ON" designates presence of secondary queue(s). "OFF" designates their absence. When the capacity of the primary queue is partly used (i.e., queue only partially-full), the secondary queue is considered "OFF" and arriving vehicles directly join a "partially-full" primary queue. Conversely, when the capacity of the primary queue is fully used, a secondary queue is "ON" and arriving vehicles are first accommodated by secondary queues. When a space is available in the primary queue, front vehicles at point(s) C (see figure 16) join the queue.

Note that in the presence of secondary queues, demand on the primary queues is no longer affected by variations in the actual arrivals distributions. Arrival waves and platoons are absorbed by the secondary queues.

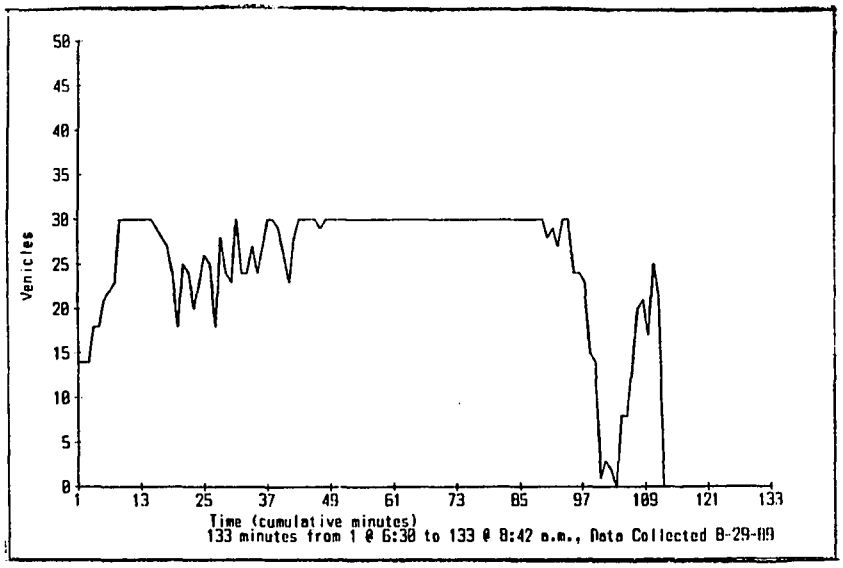
Arrivals to the primary queue from secondary queues are controlled by space availability in the primary queue and also sometimes by blockage from surface street traffic signals. Blockage due to traffic signals was minimal because while one leg of the secondary queue was shut, another leg opened creating continued demand on the primary queue.

4.b. Queue Lengths. Queue length (QL) is clearly a function of two variables: the fashion and intensity of arrivals and the cycle of the ramp meter. Queue lengths may change every second. Like service times, it was neither easy nor necessary to measure this parameter continuously (i.e., completely microscopically). The same mesoscopic TOTM interval for checking system status (which was used for service times too) was

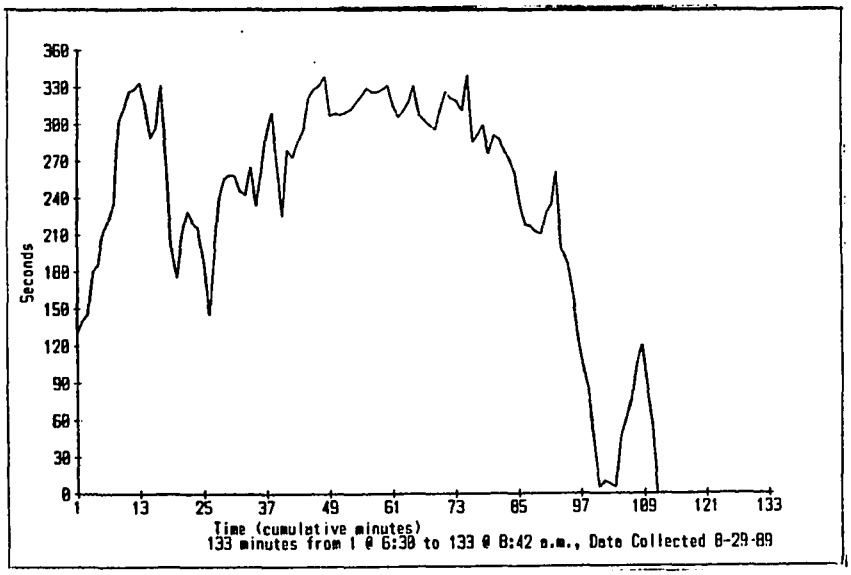
used here. The length of each queue was measured at top-of-minute intervals (e.g., 07:26:00, 07:27:00, 7:28:00 etc.). Measuring the length was done by freezing the video tape at exactly the top of each minute and then counting the number of vehicles in the queue. Queue lengths were obtained for the period from 06:30 a.m. to 08:42 a.m. for all eight ramps. Figure 20 (Plot 1) is a graphical illustration of queue length vs time at 70th Street. Plot 1 looks truncated from above at the 30 minute line. This is because this figure shows primary queue lengths only. The queue reached its capacity (30 cars) from around minute 9 (6:38) to 15 (6:44) and from about minute 42 (7:11) to 90 (7:59).

Average queue lengths for all the ramps were, similar to service times, computed and macroscopically grouped into eleven, 12-minute time slices from 6:30 a.m. to 8:42 a.m. TOTM queue length measurements that result from simulation runs will be compared with RBM measurements.

4.c Queue Delays. Queue delays (QD) reflect how long a vehicle arriving at the queue at the top of each minute is expected to wait. Ramp delays were measured mesoscopically in a similar way to service times and queue lengths. Exactly at the top of each minute (e.g., 7:28:00), a stop watch is activated and the first vehicle that arrives to the queue would be traced until it departs the queue. The total time interval (total delay including the service time at the ramp meter) is read and recorded once the vehicle crosses over the RMLL. The tape is reset (fast forwarded or rewound) onto the next minute in the sequence (e.g., 7:29:00) and the process repeated.



Plot 1. Primary Queue Lengths.



Plot 2. Primary Queue Delays.

Figure 20. TOTM Primary Queue Lengths and Delays at 70th Street.

Queue delays were obtained for the period from 06:30 a.m. to 08:42 a.m. for all eight ramps. Figure 20 (Plot 2) is a graphical illustration for queue delay vs time at 70th Street. These delays are primary queue delays only. Secondary queue delays were not measured.

Average queue delays for all the ramps were computed and grouped into eleven macroscopic , 12-minute time slices from 6:30 to 8:42 a.m.

SIMULATION & SYSTEM IMPROVEMENTS

Some system deficiencies and recurring problems were identified while studying the RBM so that alternative strategies can be tried and evaluated in the future using the final validated model.

1. Traffic Circulation

Field observations of the ramp metering system along the studied segment of I-8 in San Diego revealed several specific problems that should be included during the simulation runs of "what-if" situations. In the on-ramp subsystem, such deficiencies include surface street lane obstruction, light signal interference, abnormal traffic circulation, safety hazards for traffic and pedestrians, blocking some freeway exits (where off-ramps may meet on-ramps), undesirable traffic diversion and so forth. Many of these problems were discussed in chapter two (section: Problems with Ramp Metering).

2. Traffic Delays

At several intersections & on-ramp locations, congestion problems appear to have been transferred in part from the freeway to local streets. Numerous observations showed that freeway traffic was flowing

smoothly while adjacent surface streets were clogged up with saturated queues over the on-ramp lanes and adjacent intersections. In particular, Waring Road, Lake Murray Boulevard, 70th Street, and Fletcher Parkway intersections suffered persistent local congestion. For example, as table VI below shows, on Tuesday morning of August 29, 1989 at 70th Street, freeway flow enjoyed average speeds of 53 mph from 7:18 to 7:30, 56 mph from 7:30 to 7:42, and 54 mph from 7:42 to 7:54 while 70th Street on-ramp traffic encountered fully-saturated ramp queues (30 vehicles) through the same period and experienced average delays of 318 seconds (5.3 minutes), 311 seconds (5.2 minutes), and 292 seconds (4.9 minutes) respectively. Primary ramp queue was continuously used to full capacity. Secondary queues on 70th Street were occasionally backed to El-Cajon Boulevard (south) almost 1/2 mile away.

TABLE VI

SYSTEM PERFORMANCE: AVERAGE MAINLANE SPEEDS VS RAMP QUEUE STATUS (*) AT 70th STREET

Slice Number	Start Time	12-minute Arrival Rates (vehicles)	Ramp Ave. Queue Length (vehicles)	Ramp Ave. Queue Delay (mm:ss)	Freeway Ave. Travel Speed (mph)
1	06:30-6:42	45	23	3:46	51
2	06:42-6:54	63	26	4:16	46
3	06:54-7:06	75	25	3:51	54
4	07:06-7:18	75	29	4:54	48
5	07:18-7:30	68	30(+)	5:18(+)	53
6	07:30-7:42	70	30(+)	5:11(+)	56
7	07:42-7:54	68	30(+)	4:52(+)	54
8	07:54-8:06	86	28	3:26	54
9	08:06-8:18	105	11	0:55	55
10	08:18-8:30	124	5	0:12	55
11	08:30-8:42	116	0	0:00	55

Notes: (*) Primary queue delays only. Secondary delays are difficult to measure & may be higher than primary queue delays.
(+) Primary queue saturated/secondary queue "ON".

CHAPTER SUMMARY AND CONCLUDING REMARKS

In systems modeling, establishing an appropriate reference behavior mode (RBM) for the modeled system is a prerequisite step for building a good model. In this chapter, the supply and demand of a metered highway segment along I-8 in San Diego, CA were both investigated to establish an RBM which will be used to build a simulation model of the system.

To arrive at a balance between simplicity and precision in such a large modeling problem, only significant system aspects were considered. System details were investigated only deeply enough to achieve reasonable accuracy while maintaining an appreciation for the system wholeness from a higher macroscopic perspective. The mesoscopic view, an intermediate step between microscopic analysis and macroscopic conclusions, is deemed as a proper Holistic approach (attending the parts while being mindful of the whole) to systems modeling.

Arrivals were one key demand element on the system. They were analyzed and modeled microscopically. Three types of arrivals were observed: arrival to an always-partial queue (APQ), arrivals to a mostly partially-full queue (MPFQ), and arrivals to a mostly-full queue (MFQ).

Probability theory has been applied to find a composite theoretical probability distribution function which describes arrivals properly. Queueing is the other key system output parameter. Excessive queueing, due to high traffic demand, created considerable queue overspills. Queues therefore had to be separated into primary and secondary. Primary queues formed directly on the ramp lanes and were easy to observe. Secondary queues stretched beyond ramp lanes to surface streets and were

neither easy nor necessary to observe.

The macroscopic 12-minute slice was chosen based on examining the variation in arrivals and queueing data with respect to time. The size of the 12-minute time slice was small enough to achieve microscopic accuracy and big enough to maintain macroscopic simplicity. Top-of-the-minute (TOTM) queue observations (lengths and delays) were detailed enough to be highly representational but still simple to handle.

Congestion problems at several intersections appeared to have been transferred directly from the freeway to local streets. For example, on Tuesday morning August 29th at 70th Street, freeway flow enjoyed average speeds of 53 mph from 7:18 to 7:30, 56 mph from 7:30 to 7:42, and 54 mph from 7:42 to 7:54 while 70th Street on-ramp experienced delays of 318 seconds (5.3 minutes), 311 seconds (5.2 minutes), and 292 seconds (4.9 minutes) respectively. One of the model's tasks is to simulate and evaluate alternative strategies to improve overall system performance.

---*---*---

CHAPTER V

DEVELOPMENT OF A BASIC BUILDING BLOCK FOR AN ON-RAMP, MESOSCOPIC, STOCHASTIC, DISCRETE SIMULATION MODEL

CHAPTER PREVIEW

In this chapter, a modeling framework for model conceptualization and model input data specification will be set. The model itself, ONRAMP, is developed and presented. A criteria for model output validation will also be discussed.

According to Shannon's four-step modeling process [61, 1975] (decompose, simplify, abstract, & recompose), a workplan can be outlined as follows:

1) The on-ramp system is decomposed into its key physical subunits.
2) Its governing & internal/external relations understood & simplified.
3) These relations are mathematically abstracted, or qualitatively expressed.
4) Subunits are recomposed yielding a homomorphic model of the system.

A homomorphic model of the system implies that there is an acceptable degree of similitude between the model's form (and not necessarily structure) and the system it represents.

Steps 1, 2, and part of step 3 were taken in chapter four to produce an RBM, which is the objective of that chapter. In this chapter, steps 3 and 4 will be completed.

MODEL CONCEPTUALIZATION

Field observations showed that each on-ramp could be assinged three basic components:

* A vehicle source.
* A queue(s).
* A server (a meter).

The real-world (RBM) system can be viewed and simplified as shown in the upper part of figure 21. Vehicles arrive at one end of the on-ramp, enter and wait in queues, then depart the system at the other end as fast or as slow as the metering rates allow.

The ONRAMP block is conceptualized after the RBM system using Shannon's four-step process. Three parts are conceptualized in the model. An entry node is modeled after the ramp entrance; a queue node is modeled after the ramp storage area; servers are modeled to depict meters; and service times are assigned after metering rates.

The analogy of the model is that of a "black box". On the left side of figure 21, ONRAMP (the blackbox) receives input from one end in form of entities (vehicles in the RBM) with certain headway distributions. Model headway distributions are validated by their similarity to RBM headways. ONRAMP also receives, computes or simply reads service times (metering rates in the RBM). Service times are currently used as collected from the data. Since this item will always be determined by another algorithm, less effort will be spent to model it. Then in the middle part of figure 21, ONRAMP queues the entities, processes them, and releases them (termination) on the right end.

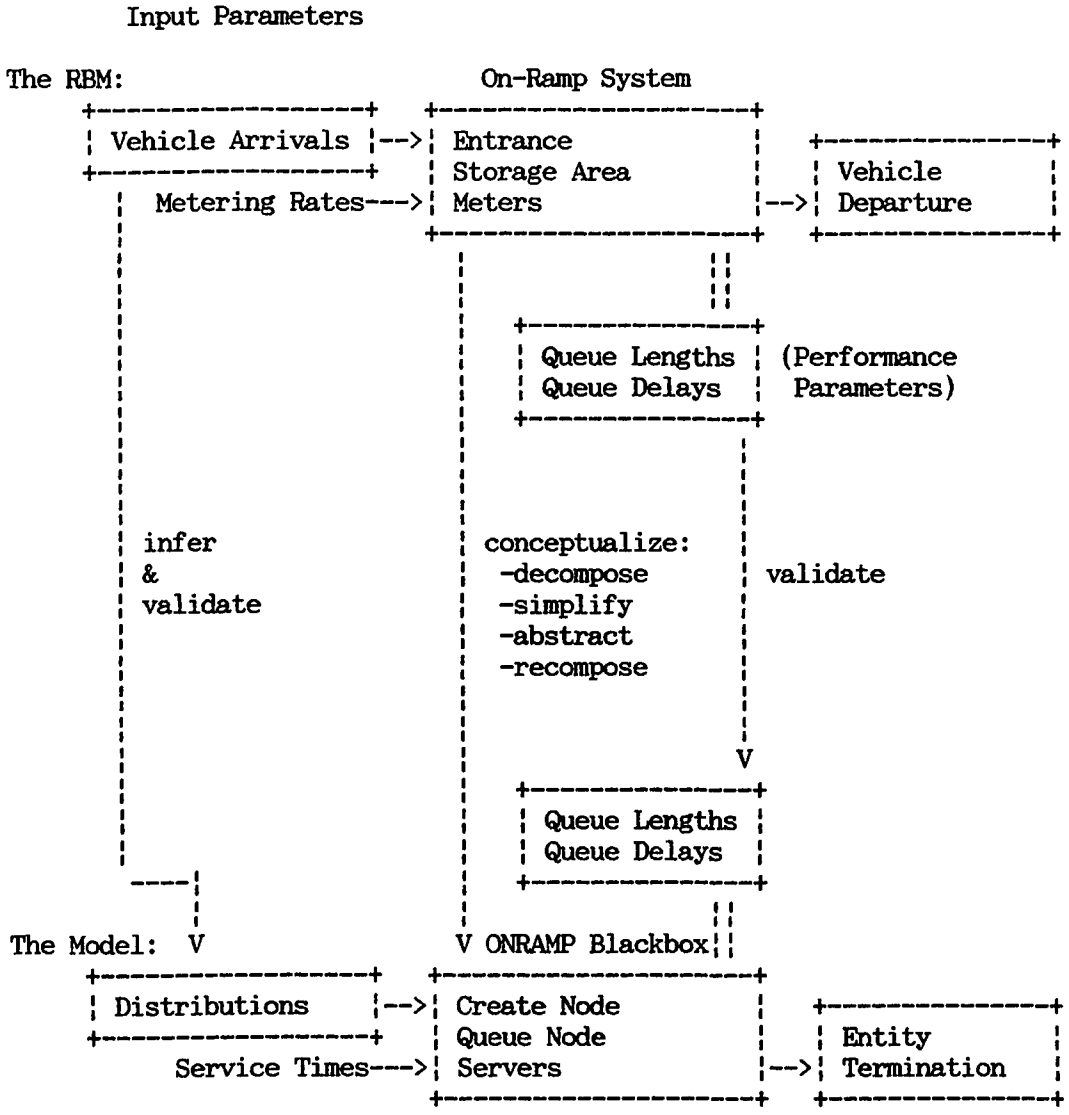


Figure 21. ONRAMP conceptualization from, and validation by the RBM. ONRAMP behaves like a black box. Its input is inferred, and its output is validated based on the real world system.

MODEL INPUT SPECIFICATION

The two most important input data items which will have to be derived from the RBM are the inter-arrival times (headways) and metering rates (service times) for the servers. The service time is the meter's full red-green cycle. Vehicular arrivals to the ramp are clearly random. The service time is also a random variable, because it may change momentarily in response to freeway conditions (*).

To properly assign headways for the model, theoretical or empirical frequency distributions needed to be determined. Gamma, lognormal, Poisson and other distributions have been tested for validity and goodness of fit. The work of other researchers in this regard have been reviewed to determine if there is an appropriate model/method to reproduce headways.

Computed sample means and standard deviations for headways and service times at each on-ramp will be used in the prospective model in conjunction with the chosen distributions for that ramp. These statistical parameters and frequency distributions are the main probabilistic components of the model.

Service times have already been obtained and they are randomized in the model using simple uniform distribution with mean and standard deviations equal to those computed in the sample data.

(*) One of the deficiencies in the current metering system is its inability to either monitor or predict ramp queues and intersection traffic movement to coordinate that with the freeway.

MODEL OUTPUT VALIDATION

Queues form and dissipate within the ONRAMP block based on arrivals and service rates. Queue lengths and delays at the ONRAMP block may be compared with RBM queues and delays to check the model performance. Once the vehicle is processed, it is discharged to the freeway. It no longer affects the on-ramp system and thus is discarded.

Assuming that a) headways and service times were properly measured, interpreted, and modeled, and b) the model was properly built and is behaving properly; then, it should be possible to verify and check the validity of the model and its output -from simulation runs- by comparing that output with RBM data with regard to headways distributions, queue lengths, and queue delays.

Since only queueing dynamics will be used for the analysis of the system performance, queue formation and dissipation had to be monitored and measured by two parameters:

1. Instantaneous queue lengths QL: the actual queue length (in vehicles) as observed at the top of each sequential minute. QL is a critical parameter because of its impact on the surface street.
2. Instantaneous queue delays QD: the total delay (in seconds) for a vehicle arriving at the queue at the top of the minute. It represents an average delay for all vehicles arriving during that minute. QD is also a critical parameter because of its impact on the total travel time and the choice of trip route. The two parameters are necessary and adequate to measure the system performance and satisfy the model objectives. They will be used to validate the model.

MODELING ARRIVALS DATA

1. Use of Simple Theoretical Distribution to Describe Arrivals

In order to model the distribution of the RBM inter-arrival times, five singular (i.e., not composite) candidate theoretical distributions and the monte carlo method were selected. Each was used to generate N random data points on the computer to compare with N data points obtained from the RBM at each ramp. Distribution parameters (mean & standard deviations) for each distribution were estimated from the data.

Part (a) of figure 22 (plots 12.1-12.3) shows the shapes of three of these distributions vs. the actual distribution of collected data at College Avenue N. Visual inspection of the plots and the chi-square tests lead to the rejection of the fit hypothesis. The lognormal distribution exhibited high mode around 2, but it was too high and the tail was too short (plot 12.1). The gamma distribution (plot 12.2) produced high mode around 1 but the curve dropped quickly after 1 (0 values were shifted to 0.5). Monte-Carlo (empirical) distribution, as expected, gave a good fit. But since the objective was to model a phenomenon and not merely to replicate past data, a theoretical distribution was still desired.

Part (b) of figure 22 (plots 12.4-12.6) shows the plots of three more distributions. Poisson distribution (plot 12.4) was too skewed to the left with no tail. The exponential distribution (plot 12.5) provided long tail (although thin) but had too many 0 points and low mode around 1. The triangular distribution (plot 12.6) was skewed to the left and had a short tail.

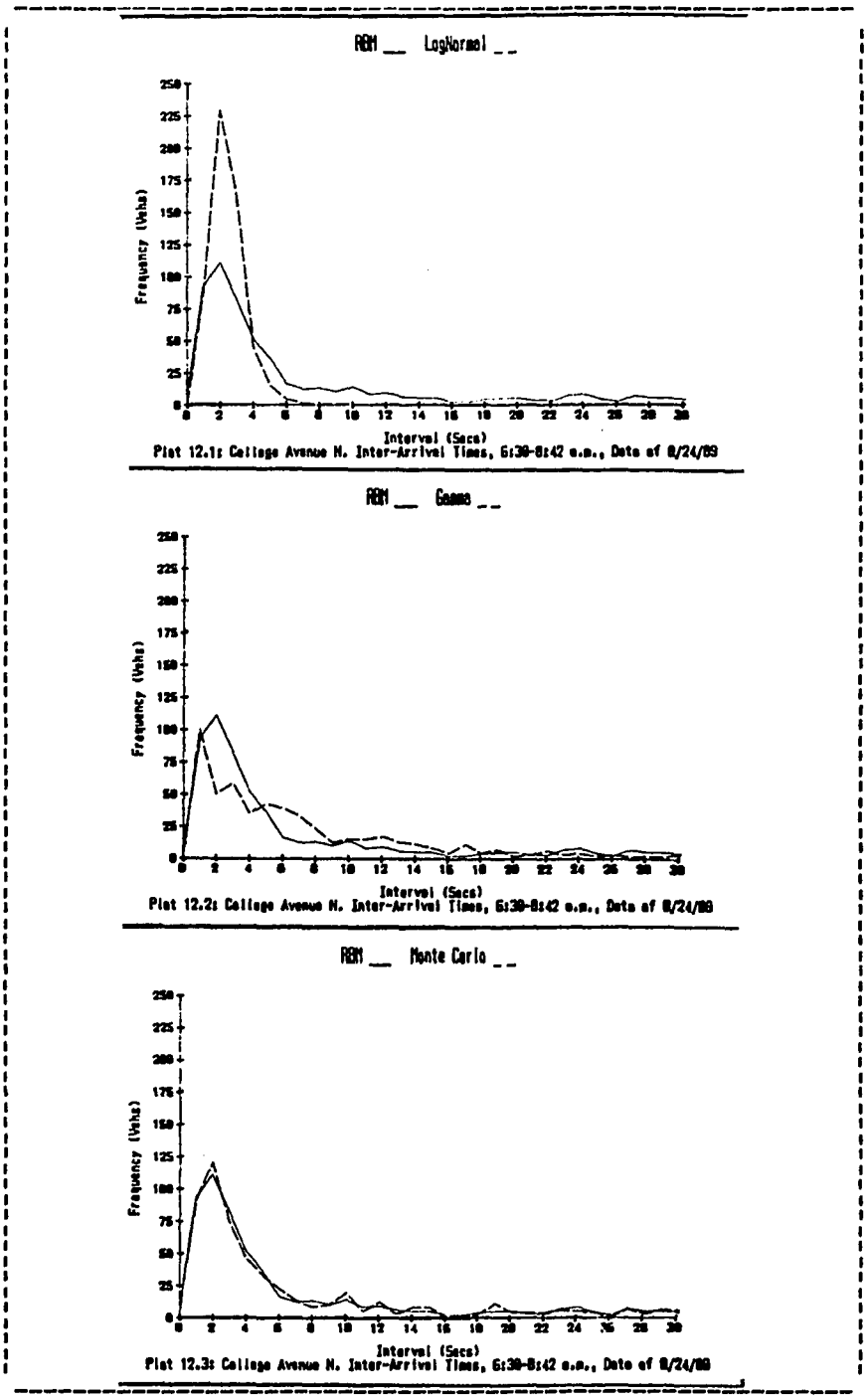


Figure 22.a. RBM Headway Distributions vs. Headway Distributions of Selected "Singular" Theoretical Probability Distributions. ("12" refers to time slice 12, 6:30 - 8:42 a.m.)

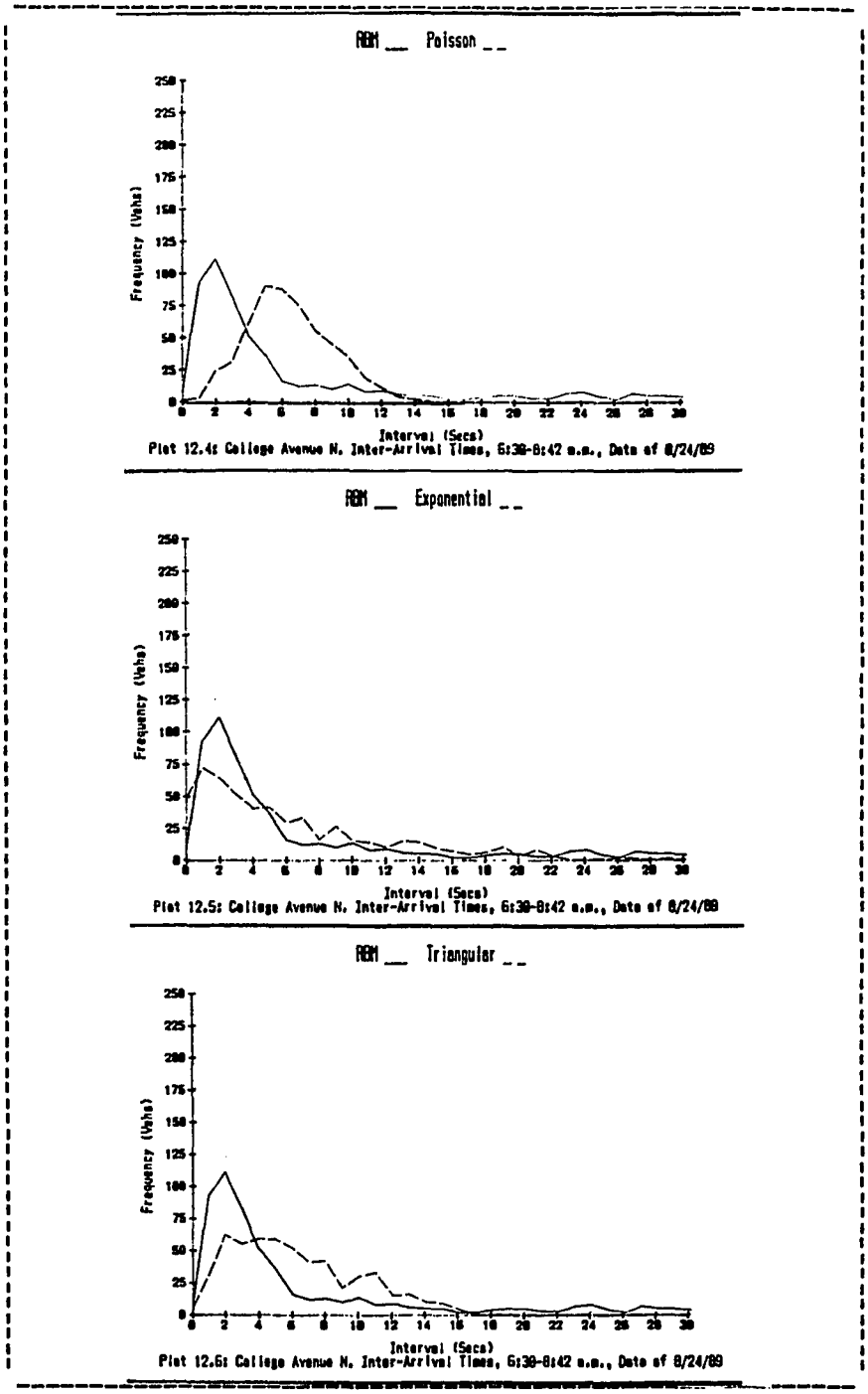


Figure 22.b. RBM Headway Distributions vs. Headway Distributions of Selected "Singular" Theoretical Probability Distributions. ("12" refers to time slice 12, 6:30 - 8:42 a.m.)

2. Use of More Complex Theoretical Distributions To Describe Arrivals

No microscopic-stochastic model could be found in the literature to describe and/or investigate vehicular arrivals at entrance ramps. However, May [52, 1990] and others investigated more general headway distributions on highways and roads.

According to May, inter-arrival times (headways) can be classified into three categories: random, constant, and intermediate. Random arrivals occur under very light traffic conditions. They are considered independent from each other. On the other hand, under very heavy flow conditions (e.g., freeway flow running near capacity), almost all vehicles travel at the *same speed* and maintain a near ***constant spacing***. Time headways in this case are considered constant. The third category is an intermediate state in between the previous two. Vehicles are not completely independent from each other, but their headways are not near constant either. This seems to be the category which describes ramp arrivals. May admits that "this is the most difficult to analyze."

May presented a generalized mathematical model, Pearson type III, as a possible way to describe these intermediate headways. The probability density function of this distribution is given by:

f(t)= lamda/phi(K) * [lamda*(t-alpha)]^(K-1) * EXP(-lamda*(t-alpha))

where:

K: User-specified parameter between 0 and infinity which affects the shape of the distribution.
alpha: User-specified parameter that affects the shift of distribution.
lamda: A parameter determined based on K, alpha, and mean time headway.
t: time headway.
phi: The gamma function of K.

Eight different sets of computations need to be performed when applying this distribution. The mean and standard deviation of simulated headways are required. Alpha is determined by trial-and-error approach. K is also approximated based on the empirical data. Lambda is estimated based on K, alpha, and the computed average of simulated data. Once K is found, phi(k), the gamma function of K, must be computed. Next, f(t) is solved for each sequential value (0.0, 0.5, 1.0, 1.5, etc.). The probabilities for interval headways are then computed using the equation:

$$P(t <= h <= t+dt)=[(f(t)+f(t+dt)) / 2] * dt$$

Afterwards, headway frequencies are calculated using the equation:

$$F(t <= h < t+dt) = N * [P(t <= h <= t+dt).$$

The above approach has several limitations. According to May, the theoretical probabilities are almost always less than the corresponding measured (real-world) distributions for values 1 or less. Also, theoretical probabilities are almost always less than the real-world ones for values higher than 4 seconds. The discrepancy between theoretical and measured distributions is most evident for values between 1 and 4 seconds. Additionally, alpha, K, lamda are approximated with a compounded degree of uncertainty. Finally, the gamma function phi(K) can not be easily computed if K is a non-integer.

As an alternative, May presented a normal-exponential model with seven parameters to estimate and involves look-ups in a graph and the normal distribution tables. The exponential is a negative exponential, but referred to here as exponential for abbreviation. The normal part is to describe vehicles in car-following (or platoon) mode. The exponential

is to represent free-moving vehicles.

This model is also limited because it applies mainly to fast-moving freeway traffic rather than traffic approaching ramp queues. It assumes that the mean value of car-following headways is closer to 1 than 2 seconds. It has several inconsistencies because the probability of theoretical distribution is "*almost always*" greater than corresponding real-world distribution for values greater than 4 seconds. Theoretical probabilities are "*almost always*" less than real-world ones for values 2.5 to 4.0 seconds. According to May, even larger discrepancies were observed when headways were between 1.0 and 2.5 seconds, particularly under low-flow conditions. Low-flow conditions are exactly what could frequently happen at certain time slices within the period from 6:30-8:42 a.m. at the entrance ramps. Finally, like the Pearson distribution, the model is somewhat difficult to implement on the computer because of table and graph look-ups, but it is a substitute to the singular distribution functions which give very inappropriate fit.

Other researchers investigated the intermediate headway state, but none more recently nor more adequately and concisely than May. Gerlough, Barnes and Schuhl [68, 1971] investigated the application of Poisson distribution. However, Poisson's implicit independence assumptions do not apply to intermittent, clustered ramp arrivals. Also, the long tail could not be generated using Poisson -as has been shown above. Daou [66, 1964], Greenberg [69, 1966], and Tolle [70, 1971] proposed the lognormal distribution for certain applications of platoon distributions. As has been shown above, the lognormal distribution provides high mode (too high for ramp arrivals) and very short tail. Lognormal distribution

would be more appropriate for freeway headways with no distribution tail and very high mode. Dawson [67, 1968] discussed a hyperlang probability distribution for generalized headway model, which is also oriented toward freeway traffic applications.

3. The Proposed Modeling Hypothesis

Clearly, one of the approaches above had to be adopted or new one formulated. Because of the limitations of the models above, a new approach has been proposed in this research. The proposed approach was developed in light of the modeling strategy introduced earlier in chapter three. Holistic view of attending the parts (e.g., analyzing individual time slices) while being mindful of the whole (i.e., the whole time period 6:30-8:42) was applied. Simplicity, versatility, flexibility, and ease of computer implementation and real-world deployment and application were conditions that guided and sometimes restricted and delayed model development.

The proposed model is a combination of three probability distribution functions. The first describes the *Exponential Creations*, the second describes *Platooned Processes*, and the third describes *tail pauses*. It is assumed that all arrivals are generated exponentially and certain portion of them undergo the platooning process with extended pauses in between.

a. Exponential Creations. The proposed hypothesis is that, starting at certain hour (6:30 a.m.), drivers begin to depart their places (origination sources: residence, nightshift work, etc.) and arrive at the on-ramp in a discrete fashion. Four assumptions are made

regarding the origination of vehicles:

1. Arrivals, originally, are non-clustered, independent events.
2. The probability of an arrival increases as time step increases (i.e., P(dt), probability of occurrence during dt, is proportional to dt).
3. Events are independent in time as well. This means that P(dt1)= P(dt2) within each time slice. This is sometimes called Markov "forgetfulness" property, where the next occurrence is oblivious to the proceeding one.
4. Arrivals, also originally, are irregularly spaced with large degree of variability. Times between arrivals could be as small as 0.5 second and/or as large as 120 seconds (and sometimes higher).

If the above four assumptions are true, then times between arrivals (inter-arrival times or headways) *are identically, independently exponentially distributed.*

b. Platooning Processes. A close examination of data and video observations showed that a certain portion of arrivals do undergo a "clustering" process often referred to as "platooning" in traffic engineering. Platooning seems to take place somewhere upstream of the arrivals reference point. The most likely place for platooning is upstream traffic signals at the approach surface roads/interactions. These signals hold a certain portion of the exponential arrivals, pack them, and release them in "batches" or "platoons". Platoons also form at random as traffic merges into main arterials and roads from side streets, parking lots, etc.

Evidently, a criteria was needed to discern between platooned and non-platooned or broken-platoon arrivals. May [52] suggested that 2.0

seconds "might be an appropriate minimum time headway" for vehicles not in platoon. This value applies to higher speeds, not lower ones as the case is with vehicles approaching queues on ramp lanes.

After reviewing thousands of arrivals in the data set, it was possible to assume a threshold for broken platoons and non-platooned ramp arrivals at 4 seconds. Our assumption is that any inter-arrival time (headway) of 3.5 seconds or less (always rounded to 3 seconds) can occur only in a platoon. Therefore, all arrivals with headways greater than 3.5 seconds (4 seconds if rounded to nearest integer) were considered broken platoons or non-platooned arrivals. The smallest platoon size of course is 2 vehicles.

The most common headway value within platoons was 2 seconds. The second and third most common values were 1 & 3 seconds --as explained earlier. It is important to point out that arrivals to the primary queue from a secondary queue (that is controlled by a traffic signal) were also considered platooned, but evidently with larger headway due to lower speeds. Platooning was considered to have taken place at the secondary queue. In this case, the platoon size was as small or as big as the space available in the primary queue. If only one space is available at-a-time, time headways would depend on the progression speed of the primary queue. With secondary queues present, the mode continued to be 2.0 seconds but arrivals distribution curve was shifted rightwards and the second highest value became 3.0 seconds.

Once the criteria for platooned arrivals was established, computing the ratio of platooned arrivals was simple. The ratio of exponential to platooned arrivals was considered E:P. E is the number of headways

greater than or equal to 4.0 seconds, whereas P is the number of headways less than 4.0 seconds. For example, if E= 40 & P=60, it means that 60% of the arrivals underwent a platooning process. The rate of platooning certainly varies from ramp to ramp and it could also vary from day to day depending on the activities at the traffic signals upstream and on the density of traffic volumes.

c. Compound/Long Tail Pauses. Each headway distribution in the data set had a long tail. The long tail represents long pauses in arrivals. Obviously, the tail end and thickness depend on the type of arrivals. For type APQ arrivals (refer to chapter four), the tail was generally long (up to 108 seconds). For type MPFQ, the tail was shorter and no extended headway pauses occurred. The shortest tail was observed in type MFQ arrivals. Shorter tails indicate more constant demand on the primary queue because when the secondary queue is present, cars in it will move on to the primary queue as soon as there is a space available, except when they were held temporarily by traffic signals. This created a continuous demand and eliminated any long tails.

4. A Proposed Composite Distribution

As was demonstrated before, no single theoretical distribution was sufficient to describe the total phenomenon of the arrival process. This indicated that a *composite distribution* may provide the right mix which would represent all arrivals. Several combinations were examined and some gave good results.

A three-part composite distribution is proposed and will be explained in the following sections. It seems to represent the

combination of random inter-arrival data nicely. The first part is a triangular distribution which describes the platooned arrivals. The second is a truncated exponential distribution which describes broken platoons or non-platooned arrivals. The third is another triangular distribution which describes observed "pauses" in arrivals (the long tail).

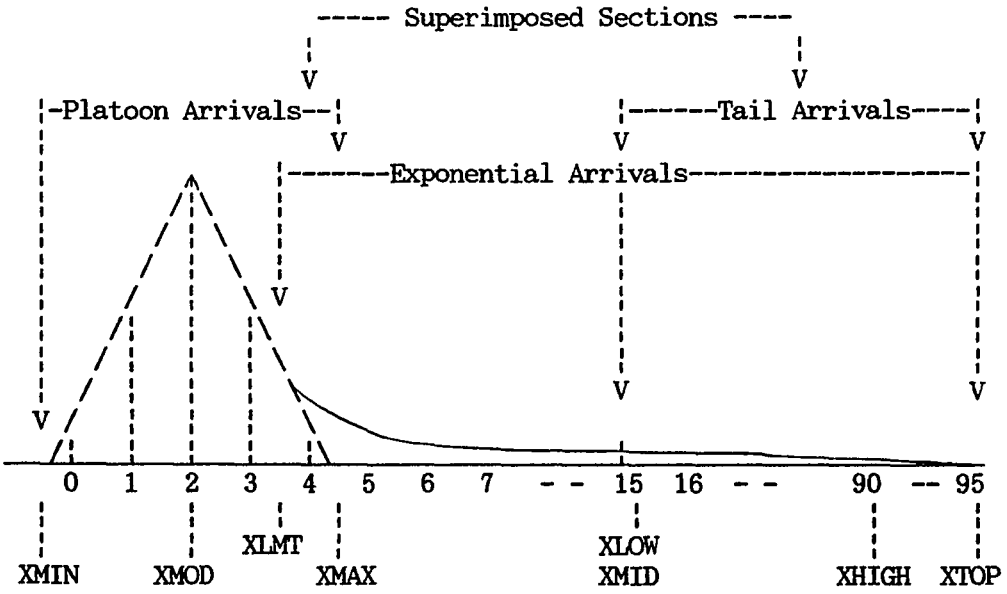
a. Triangular Distribution I: (Platooned Arrivals). This describes the "platooning" process that occurs upstream of the on-ramp. Three parameters are needed to construct this distribution: A minimum value (say, XMIM), a mode (say, XMOD), and a maximum (say, XMAX). These three parameters can be determined from headway data at each on-ramp. The probability density function for the triangular distribution is:

$$f(x)= 2*(X-XMIN)/[(XMOD-XMIN)*(XMAX-XMIN)], \quad \text{for} \quad XMIN < X < XMOD$$

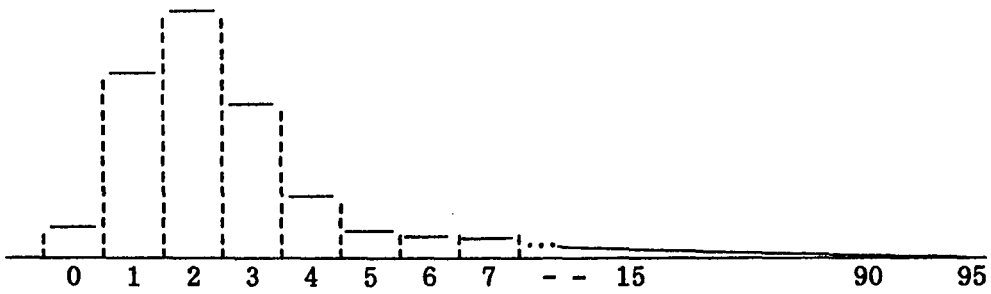
$$\text{or} \quad 2*(XMAX-X)/[(XMAX-XMOD)*(XMAX-XMIN)], \quad \text{for} \quad XMOD < X < XMAX \qquad (1)$$

Values of: XMIN= -0.5 sec
XMOD= 2.0 sec
XMAX= 4.5 sec
are used as default and are recommended and will yield an equa-lateral triangular shape. Figure 23 is a graphical representation of the proposed composite function with parameters set to their default values.

If all values between -0.5 & 0.45 are shifted to "1" and all values between 0.4 & 4.5 are rounded up/down to the nearest integer, a 5-cell step distribution function will result (figure 23).



(a) Real-Value Composite Distribution



(b) Integer-Value Composite Distribution (values between cell boundaries in (a) are added and rounded to nearest integer).

Figure 23. Three-Part Composite Density Function With All Parameters Set to Their Default Values. (a) Before Rounding & (b) After Rounding.

The shifting of values < 0.45 decreases the value of the "0" cell and increases the value of the "1" cell. Notice that rounding values between 0.45 and 0.50 will yield a zero. The choice of 0.45 or other cut-off values for the lower zero boundary is determined by examining the data and computing the weight of "0" inter-arrivals. Data observation suggested the use of 0.45 or 0.40 as a default value across all the ramps and throughout all time slices.

b. Truncated Exponential Distribution: (Broken Platoons). This distribution starts at a certain threshold value and it describes broken platoons or non-platooned, identically, independently distributed inter-arrivals. It is assumed that any headway value > 3.5 seconds (4 seconds if rounded to the nearest integer) represents a broken platoon, or non-platoon arrival. This assumption is based on field observations of vehicular platoons and on the visual inspection of the RBM distribution curve of inter-arrival times. Only one parameter, the mean (AVG), needs to be estimated.

The probability density function of the standard exponential distribution is given by:

$$f(X) = u*EXP(-uX) \quad , \quad \text{for} \quad 0.0 < X < \text{infinity}$$

The probability density function used in the composite function is a "truncated" exponential distribution and is given by:

$$f(X) = W * [u*EXP(-uX)] \quad , \quad \text{for} \quad XLMT < X < XHIGH \qquad (2)$$

where:

u: The inverse of the exponential mean u=1/AVG.
W: Compensation factor for the truncated part below XLMT.
XLMT: Threshold for non-platooned arrivals.
XHIGH: An upper bound for inter-arrivals. This boundary is set to eliminate very long tails or "pauses" in arrivals which are unrealistic but the exponential distribution may produce.

As an illustration, using AVG=5.0 and integrating (2) between XLMT=3.5 & XHIGH=90 & equating the integral to 1, it was found that W=2.0138.

C. Triangular Distribution II: (Tails). Most observed data had a "thicker" tail than that which can be provided by the exponential distribution only. This proposed part of the composite function is superimposed on the exponential tail to give the additional needed thickness (see figure 23). Based on data observations, this superimposed tail is assumed to start at the mean value of the exponential distribution + 10 (i.e., AVG + 10). Again, the probability density fraction for the triangular distribution is given by:

$$f(x) = 2*(X-XLOW)\ /[(XMID-XLOW)*(XHIGH-XLOW)], \text{ for } XLOW < X < XMID$$

$$\text{or} \quad 2*(XHIGH-X)/[(XHIGH-XMID)*(XHIGH-XLOW)], \text{ for } XMID < X < XHIGH \quad (3)$$

where: XLOW: Lower boundary for the triangular tail.
XMID: Mode of the triangular tail.
XHIGH: Upper and maximum allowable end of the tail.

Values of:
XLOW= AVG + 10
XMID= AVG + 10
XHIGH= XLOW + UNIFORM(60,90).

are used in the proposed model. Notice that the tail's upper end is randomized by adding a uniformly distributed value between 60 and 90 (seconds) to XLOW. The values of 60 and 90 are suggested based on observations of the data and may be modified if desired.

As an illustration, if the mean of the exponential distribution is AVG= 5, and UNIFORM(60,90)=75, then the three parameters become:

XLOW= 5.0 + 10.0 = 15.0
XMID= 5.0 + 10.0 = 15.0
XHIGH= 15.0 + 75.0 = 90.0

Finally, the composite density function takes the format:

$$
f(X) = \begin{cases} \text{alpha}*f(11) & , \quad \text{XMIN} < X < \text{XMOD} \\ \text{alpha}*f(12) & , \quad \text{XMOD} < X < \text{XLMT} \\ \text{alpha}*f(12) + \text{beta}*(1-\text{gamma})*f(2) & , \quad \text{XLMT} < X < \text{XMAX} \\ \text{beta}*(1-\text{gamma})*f(2), & , \quad \text{XMAX} < X < \text{XMID} \\ \text{beta}*[(1-\text{gamma})*f(2) + \text{gamma}*f(3)] & , \quad \text{XMID} < X < \text{XHIGH} \end{cases} \quad (4)
$$

Where:

f(11)= 2*(X-XMIN) / [(XMOD-XMIN)*(XMAX-XMIN)]

{left half of platoon triangle}

f(12)= 2*(XMAX-X) / [(XMAX-XMOD)*(XMAX-XMIN)]

{right half of platoon triangle}

f(2)= W*u*EXP(-uX)

{Exponential part}

f(3)= 2*(XHIGH-X) / [(XHIGH-XMID)*(XHIGH-XLOW)]

{tail triangle}

5. Parameter estimation of the Proposed Composite Distribution

Triangular Portions (platooned & tail). Estimating the triangular parameters for the platooned parts depends on the desired shape of headway distribution. If a generalized case is sought, no parameter estimation for the triangular distribution (platooned arrivals or tail arrivals) would be required. The minimum, mode and maximum parameters for these two distributions are set to default values in the model as shown below:

1. Default values for platoons:

XMIN= -0.5
XMOD= 2.0
XMAX= 4.5

2. Default values for the tail:

```
XLOW= AVG + 10.0
XMID= XLOW
XHIGH= XLOW + UNIFORM(60,90)
```

If it is desired to manipulate the model to produce arrivals to resemble a particular shape, XMIM, XMOD, and XMAX can be varied to manipulate the platoon section and XLOW, XMID, and XHIGH can be manipulated to play with the tail. For example, to shift it the centroid of the platoon triangle to the left, XMIM can be decreased or both decreased. To shift it to the right, XMIN or XMAX could be increased or both increased. To reduce the height of the triangle, XMIM could be decreased and XMAX could be increased. To increase the height of the triangle, XMIM could be increased and XMAX decreased.

Exponential Part. Only one parameter, the mean, is needed for the exponential distribution. This parameter was estimated experimentally by the modeler. Initially it was done by visually inspecting the curve past 3.5 value and assuming a reasonable mean value. If the assumed value is a good estimate, the resulting exponential distribution from point 3.5 and up would be a good fit for the arrival data at 3.5 seconds and higher.

One way to approximate this parameter was to compute the mean of all arrivals less than 30 seconds. (to eliminate skewness toward the tail). The following equation has also been formulated empirically to approximate the sample mean. The formula has been developed by performing a regression analysis:

AVG= 5.75 + 0.5 (Erate * Smean)

where: AVG: Estimated mean for (model) exponential arrivals.
Erate: Percentage of all arrivals greater than 3.5 seconds.
Smean: Computed sample mean of all arrival data.

Alternatively, May [52] suggested the following formula to find either the mean value of non-platooned headways knowing the mean value of all headways, or vice versa:

Pp = 1.5 / Tb

where: Pp: Percent of platooned arrivals
Tb: Mean value for time headways (all arrivals)
1.5: Mode of all platooned arrivals (estimated empirically)

Rearranging and substituting XMOD for 1.5 for use in our model yields:

AVG = XMOD / Erate (*)

(*) The above methods for estimating AVG are a source of weakness to the model. They are meant only to find an "approximation" for the mean and more research is needed to find better a estimation method.

THE COMPUTER IMPLEMENTATION

The computer implementation for generating random variates from the proposed composite is done by using separate subroutines for the triangular and the exponential distributions. Calls are made to these subroutines in a manner proportional to alpha (or PRATE, weight of platoon part), beta (or ERATE, weight of non-platoon part including tail and is equal to 1.0 - PRATE), and gamma (or TRATE, weight of the tail in the non-platooned arrivals). For example, if N arrivals are to be generated, then the following calls are made:

1) alpha * N calls to TRIAG(XMIN,XMOD,XMAX).

2) beta*(1-gamma) * N calls to EXPON(AVG).

3) beta*gamma * N calls to TRIAG(XLOW,XMID,XHIGH).

Calls would be made in cycles according to the proposed platoon size. For example, for one cycle, if the platoon size is 6 and the platooning rate (alpha) is 0.40, and the tail rate (gamma) is .333, then for each round of calls, the following takes place:

```
1) PSIZE=6, the size of the platoon ==> 6 calls to TRIAG(XMIN,XMOD,XMAX)
2) ESIZE=PSIZE/alpha - PSIZE
        =  6 /  0.4  -  6 = 9
3) TSIZE=gamma*ESIZE
        = 0.333*9 = 3  ===============> 3 calls to EXPON(AVG)
   ESIZE=ESIZE-TSIZE
        =  9 - 3 = 6   ==============> 6 calls to TRIAG(XLOW,XMID,XHIGH)
```

The three calls to tail values are not made in sequence but rather they are scattered within the calls to EXPON to prevent consecutive tail values (and prevent long pauses). The above loop is repeated until N number of calls (total) has been made.

EMPIRICAL VERIFICATION OF THE PROPOSED COMPUTER IMPLEMENTATION SCHEME

The implementation method explained above was first applied to data for the College Avenue N. ramp. The resulting distribution provided a good fit for all eleven time slices for that ramp as well as the entire period 6:30-8:42. Chi-square test was performed and all values were below the critical ones for 0.975 level of significance. The method was then generalized to all eight ramps and tested at time slice 12 (6:30-8:42 a.m.). Plot R.2 below illustrates one excellent sample of the simulated distribution (the dashed line) vs. the actual headway distribution of observed data at College Ave. S.

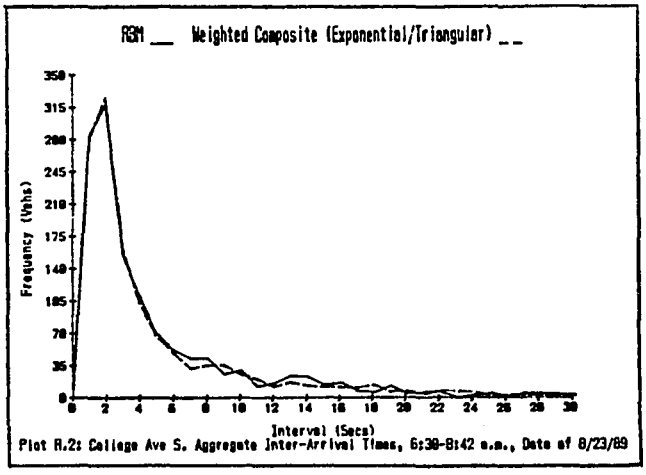


Figure 24 shows four more computer-generated distributions vs. the distribution of the observed (RBM) data at Waring Road, College Avenue N., Spring Street, and Jackson Drive. The method seems to provide not only an excellent visual fit, but also a mathematically acceptable one. None of the chi-square test results at the 95% level of confidence did indicate that the hypothesis of good fit should be rejected.

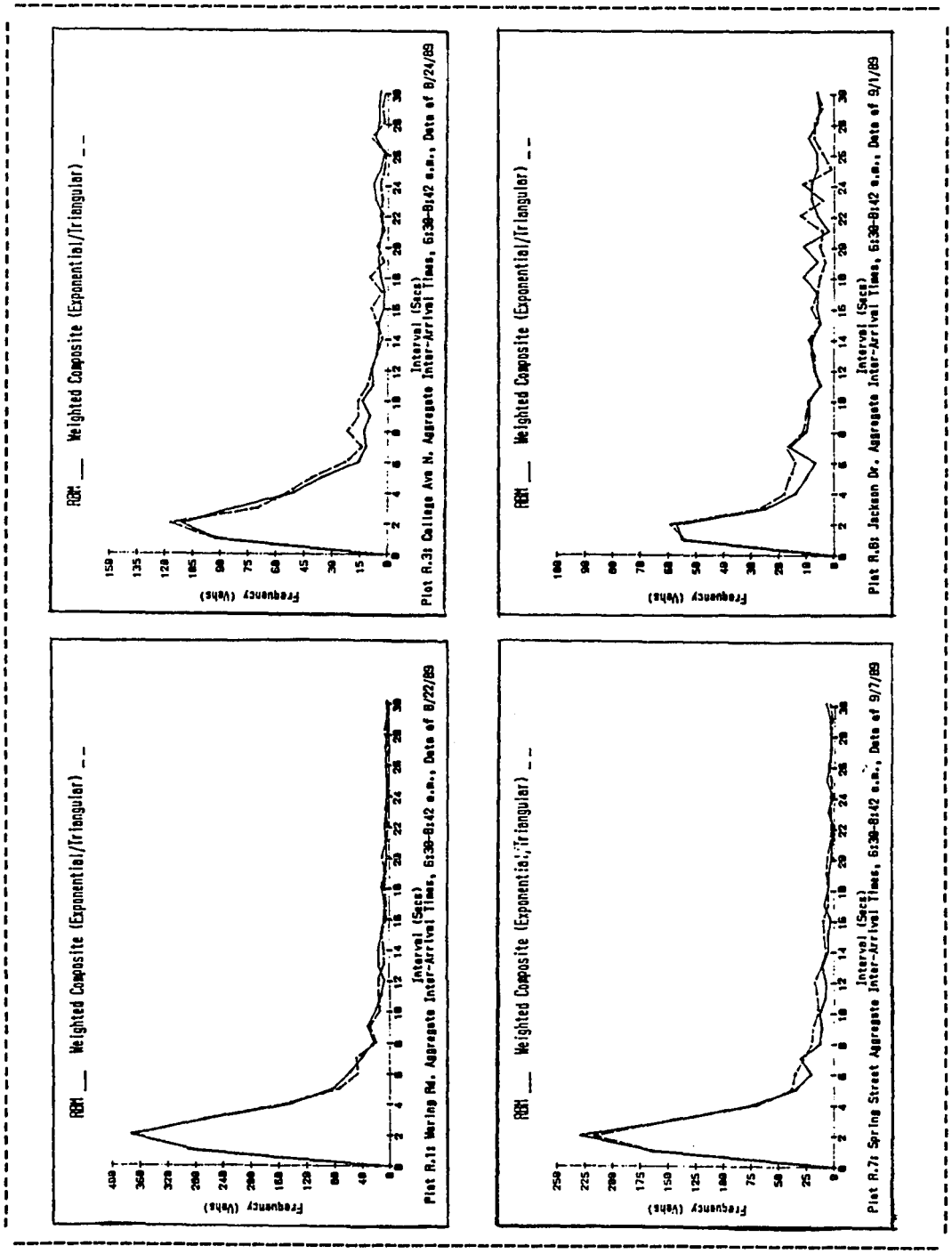


Figure 24. Real-World (RBM) Headway Distributions vs. Simulated Computer Composite Headway Distributions at Four Entrance Ramps. (Simulated Period: 6:30-8:42 a.m.)

MATHEMATICAL VALIDATION OF THE PROPOSED COMPOSITE FUNCTION

In order to accept the proposed computer implementation of triangular/exponential/triangular and the mathematical composite function f(x) as a valid probability density function (pdf), its cumulative distribution function (cdf or F(x)) should add up to 1. To verify this, the cumulative distribution of f(x), F(X), was computed by integrating f(X) in equation (4) along 16-pairs of data points (representing 16 cells from XMIN to XHIGH). Frequency computations were performed using the composite function with default values for XMIN, XMOD, XMAX, XLMT, XLOW, XMID, XHIGH parameters, as suggested for use in the model. Parameters AVG, alpha, beta, and gamma were estimated using the arrivals data set at Waring Road. Therefore, letting:

```
     XMIN= -0.5
     XMOD=  2.0
     XMAX=  4.5                          ,   for the platooned section
      AVG=  5.0 (u=1/5.0)
     XLMT=  3.5
        W=  2.0138                       ,   for the non-platooned section
     XLOW=  AVG+10  =  5.0+10.0= 15.0
     XMID=  AVG+10  =  5.0+10.0= 15.0
    XHIGH=  XLOW+75 = 15.0+75.0= 90.0    ,   for the tail section
&
    alpha= 0.68, rate (weight) of platooned arrivals
     beta= 0.32, rate (weight) of all non-platooned arrivals (truncated
                 exponential and triangular tail). (beta=1-alpha always)
    gamma= 0.15, rate (weight) of superimposed triangular tail within
                 non-platooned arrivals
```

And substituting in equation (4) yields:

```
f(X)= 0.68*[2*(X+0.5) / 12.5]                        ,  -0.5 < X < 2.0
      0.68*[2*(4.5-X) / 12.5]                        ,   2.0 < X < 3.5
      0.68*[2*(4.5-X) / 12.5] +
      0.32*(1-0.15)*[2.0138*0.2*EXP(-0.2*X)]         ,   3.5 < X <  4.5
      0.32*(1-0.15)*[2.0138*0.2*EXP(-0.2*X)]         ,   4.5 < X < 15.0
      0.32*[(1-0.15)*(2.0138*0.2*EXP(-0.2*X)) +
      0.15*(2*(90-X) / 5625)]                        ,  15.0 < X < 90.0   (5)
```

Integration is now done using f(x) in equation (5). Pairs of points were chosen to coincide with consecutive cell boundaries (i.g., 2.5 to 3.5 for cell "3", 3.5 to 4.5 for cell "4", etc. -see figure 23). Integrating between each two boundaries yields a value, a relative frequency, for that cell. If the composite function is a valid pdf, then adding relative frequencies of all consecutive cells must yield 1. The integral value from -0.5 to 0.4 was shifted "added" to the integral value between 0.5 to 1.5 to increase the value of "1" cell.

Table VII shows the results of this integration and compares them with the results of the adopted computer model. Column 1 is cell limits from 0.0 to 90. Column 2 and 3 contain the relative and cumulative frequencies as obtained from integrating the composite function. Column 3 indeed adds up to one. The 0.0009 fraction is a rounding/fraction error. Columns 4, 5, 6, and 7 are data points for each cell as obtained from field observation, mathematical pdf, and two computer outputs respectively.

The two computer outputs were obtained from the compound triangular-exponential-triangular distribution model for headways as implemented in ONRAMP. Each run used a different seed for the random number generator on the AMIGA 1000 computer.

TABLE VII

FIELD-OBSERVED, MATHEMATICALLY-COMPUTED, AND COMPUTER-GENERATED FREQUENCY DISTRIBUTIONS OF 1480 INTER-ARRIVAL TIMES (HEADWAYS) AT WARING ROAD

Headway Cell	Composite Histogram		Compared Data Points			
	Relative Frequency	Cumulative Frequency	Field Data	Mathematical Composite	Computer Model Seed 1	Seed 2
0.0-0.4	0.0443	0.0443>	(these values are shifted to 1 cell)			
0.4-0.5	0.0104	0.0547	17	15	19	18
0.5-1.5	0.1642	0.2189	282	309	283	295
1.5-2.0	0.1231	0.3420+ (a)				
2.0-2.5	0.1231	0.4651	372	364	370	385
2.5-3.5	0.1642	0.6293	265	243	267	242
3.5-4.5	0.1034	0.7327	143	153	150	151
4.5-10.5	0.1537	0.8864	257(b)	228	231	222
10.5-14.5	0.0365	0.9229	53	54	48	62
14.5-15.5	0.0060	0.9289	11	9	11	10
15.5-16.5	0.0057	0.9346	8	8	5	7
16.5-20.5	0.0158	0.9504	28	23	28	21
20.5-30.5	0.0186	0.9660	37	28	30	31
30.5-90.0	0.0319	1.0009	9(b)	47	40	38

Notes: (a) Cell 2 values came from two different integrations for X < XMOD and X > XMOD (refer to the triangular "platoon" part of the composite function.

(b) The denser tail produced by the composite function and the model is due to the generalized assumption of the model that the tail end could be as high as 90, whereas, for this particular data set, the tail was short. The highest observed headway value was 49 seconds. The assumption is that values which should have belonged to the tail were skewed and distributed over the exponential range 4.5-10.5.

Table VIII shows the results of a chi-square test which was performed to check the degree of agreement between three distributions:

A. The observed (RBM) data.

B. The expected data from the mathematical composite.

C. The expected data from the computer model.

The first chi-square test was performed to further verify the

hypothesis that sample data obtained from the proposed mathematical composite function does not differ significantly from observed data. Test results (column A-B in table VIII) gave a chi-square value of 11.249. Another test was performed to check if the method of computer implementation would differ significantly from the observed data. Test results (column A-C in table VIII) gave a chi-square value of 3.927. The third chi-square test was to check the degree of agreement between the proposed mathematical composite and the computer model. Test results (column B-C in table VIII) gave a chi-square value of 5.603.

Since only three parameters AVG, alpha and gamma are needed in the composite distribution (both the mathematical & computer models), the degree of freedom was:

$$df = k - 1 - m$$

$$= 12 - 1 - 3$$

Where: k: the number of cells

m: the number of parameters

At 95% confidence level, the clinical chi-square is equal to 15.5, which is greater than all three computed chi-square values, and thus, none of these three hypothesis of agreement above can be rejected.

TABLE VIII

CHI-SQUARE TEST OF FIELD-OBSERVED, MATHEMATICALLY-COMPUTED AND COMPUTER-GENERATED FREQUENCY DISTRIBUTIONS OF 1480 INTER-ARRIVAL TIMES (HEADWAYS) AT WARING ROAD

	A	B	C	Chi-Square Computations		
				---	---	---
Headway Cell	Field Data	Mathematical Composite	Computer Model(*)	A-B o e	A-C o e	B-C o e
0	17	15	18	0.266	0.055	0.500
1	282	309	289	2.359	0.169	1.384
2	372	364	377	0.175	0.066	0.448
3	265	243	254	1.991	0.476	0.476
4	143	153	150	0.653	0.326	0.060
5-10	227(+)	228	226	0.004	0.004	0.017
10-14	53	54	55	0.018	0.072	0.018
15	11	9	10	0.444	0.100	0.100
16	8	8	6	0.000	0.666	0.666
17-20	28	23	25	1.086	0.360	0.160
21-30	37	28	30	2.892	1.633	0.133
31-90	19	45	19	1.161	0.000	1,441
				---	---	---
				11.249	3.927	5.603

Notes: (*) The average of the two computer runs was used.
(+) 30 points were shifted back to tail values.
o: Observed.
e: Expected.

THE ONRAMP COMPUTER MODEL

The concept of the proposed composite function is integrated into a basic building block of the on-ramp model ONRAMP. ONRAMP is built based on the modeling strategy of chapter three, the real-world observations and the RBM in chapter four, and the modeling hypothesis in chapter five.

1. Model Variables

The model variables introduced here are all SLAM.II global (system) variables and entity attributes. Table IX contains a variable glossary for the proposed model.

TABLE IX

VARIABLE GLOSSARY FOR THE ONRAMP MODEL

Variable Name	Variable Definition
Entity Attributes:	
ATR(1):	Entity arrival time (ARTIME).
ATR(2):	Ramp meter ID, where entity was generated (RAMPID).
ATR(3):	Slice ID, when entity entered the on-ramp system (SLICID).
ATR(4):	TOTM ID, when entity entered the on-ramp system (TOTM).
ATR(5):	Inter-arrival time (headway between entities) (HEADWY).
ATR(6):	Messenger entry code (MSNGER). 1= messenger entity. 0= regular entity.
ATR(7):	Primary queue length when entity arrived to the on-ramp (PQL).
ATR(8):	Secondary queue length when entity arrived to the on-ramp (SQD).
ATR(9):	Total delays for the entity including primary and secondary queue delays and the service time.
Global Variables:	
XX(1)=AVG:	12-minute inter-arrival mean value.
XX(2)=ST:	12-minute average service time at the meter.
XX(3)=SD:	12-minute standard deviation of meter service time.
XX(4)=TSN:	Number of creations/time slice.
XX(5)=STL:	Lower bound of uniform distribution of meter service time. XX(5)=XX(2)-XX(3)-0.1
XX(6)=STH:	Upper bound of uniform distribution of meter service time. XX(6)=XX(2)+XX(3)+0.1
XX(7)=ESIZE:	Exponential portion of vehicular arrivals.
XX(8)=TRATE:	Tail weight of superimposed triangular tail.
XX(9)=PSIZE:	Platooned portion of vehicular arrivals.
XX(10)=TLSTEP:	Tail step counter for the next tail arrival
XX(11)=XLOW:	Minimum value of triangular tail.
XX(12)=XMID:	Mode value of triangular tail.
XX(13)=XHIGH:	Maximum value of triangular tail.
XX(14)=XTOP:	Maximum allowable sum of two successive entities.
XX(15)=XMIN:	Minimum value of triangular platoons.
XX(16)= XMOD:	Mode value of triangular platoons.
XX(17)=XMAX:	Maximum value of triangular platoons.
XX(20)=XCALL:	Dummy variable (to call USERF).
XX(21)=PSTHDY:	Last computed headway.

TABLE IX

VARIABLE GLOSSARY FOR THE ONRAMP MODEL
(continued)

Variable Name	Variable Definition
XX(22)=PLONG:	User-suggested platoon length (determined from field observations).
XX(23)=PRATE:	Platoon rate (determined from field observations & data).
XX(24)=CUMHDY:	Cumulative headways so far. A work pointer.
XX(25)=VEHCUM:	Cumulative counter for all created entities.
XX(26)=XLMT:	Platoon headway limit. Also a threshold for exponential arrivals.
XX(27)=ECOUNT:	Temporary counter for exponential entities.
XX(28)=TLNEXT:	Pointer to the next tail entity.
XX(29)=PCOUNT:	Temporary counter of platooned entities.
XX(30)=CONHDY:	180-second-incremental headways counter. A work pointer.
XX(31)=TOTM:	Minute timer (pointer to top-of-the-minute).
XX(32)=SLICID:	Pointer to time slice ID number.
XX(33)=SLICSZ:	Time slice size (normally 720 minutes).
XX(34)=TSCUM:	Cumulative time slice (720, 1440... etc).
XX(35)=TOTMQ1:	Primary queue length at TOTM.
XX(36)=TOTMQ2:	Secondary queue length at TOTM.
XX(37)=CALINC:	Calling increment. A work counter.
XX(87)=TOTMNW:	Actual time in seconds at TOTM.
XX(88)=AFLAG:	Signal to flag messenger entity. 1= assign entity as messenger. 0= regular entity.

2. Model Assumptions and Notes

The following assumptions are made for the model:

- Platoon breakdown threshold valve is 4 seconds (all <= 3-second inter-arrival times are considered platooned).
- All vehicles are created exponentially, but go through platooning process at an upstream point (i.e., surface street traffic light.)
- All platooned arrivals have mode value = 2, but this may be varied.
- The sum of two successive headways can not exceed tail end + AVG. (to prevent excessive pauses).
- About 1% "zero" intervals are permitted to represent fast arriving

vehicles which are platooned extremely closely.

- All floating-point values of randomly-generated headways are converted (rounded up/down) to the nearest integer in order to be compatible with the collected data. Headway measurements in data came from the video-camera timer and the computer timer. Both provide 1 second increments only.
- Zero headways come from the triangular distribution only. When the lower limit of the triangle (for platoons) is between 0.45 and 0.5, it is rounded down to zero. Triangular lower limit may be less than 0.45, but all values less than 0.4 are shifted to 1.

3. Model Input Parameters

Below are the model input parameters which are currently required to run the ONRAMP model:

1. N: Number of data points (vehicles, or, entities to generate.)
2. PRATE: Platoon ratio. This ratio can be determined from the collected data by computing the percentage of platooned headways (i.e., headways less than 4 seconds). Once this fraction is found, it can be used as a platooning ratio. For example, if the platooning fraction is 47%, then an PRATE = 47% & ERATE= 53%. Note that any PRATE and ERATE ratios may be used as long as PRATE represents logical/appropriate platoon proportion and the sum of PRATE and ERATE equals 1.

   [Default=.5]
3. TRATE: Trail rate. This determines the weight (percentage) of the extended tail of the exponential distribution.

   [Default value= 0.15]

4. Triangular Distribution Parameter (for platooned arrivals):

XMIN: Left end of platoon triangle.
[Default= -0.5]
XMOD: Mode of all platoons.
[Default= 2]
XMAX: Right end of platoon triangle.
[Default= 4.5]

5. AVG: Mean of the Exponential Distribution (non-platooned arrivals):

[Default= 1/2*(observed average for headways > 4 seconds)]

This parameter may be approximated by interpolating between the mean value of all arrivals > 3 seconds and the mean value of all arrivals. This parameter can also be determined by generating several sets of exponential random variates only and comparing their distribution with the distribution of the data. The mean value of exponential distribution that gives the best fit can be used as an estimate for the mean value of exponential distribution.

6. Triangular Distribution (Extended Tail):

XLOW: Lower end of tail.
[Default= AVG + 10]
XMID: Mode value of tail.
[Default= XLOW]
XHIGH: Upper end of tail (highest possible value for headway.)
[Default= XLOW + Uniform(60,90)]

7. XTOP: Maximum allowable sum of any two successive arrivals.

[Default= XHIGH + 5]

Figure 25 is a SLAM network diagram of the proposed model ONRAMP. ONRAMP is composed of two subsystems: an upstream subsystem and the on-ramp subsystem. At the upstream subsystem, vehicles are generated at a create node ORIG (see figure 25). Vehicles go next to a "go-on" node STRT. A decision is made at STRT whether to branch the entity to one

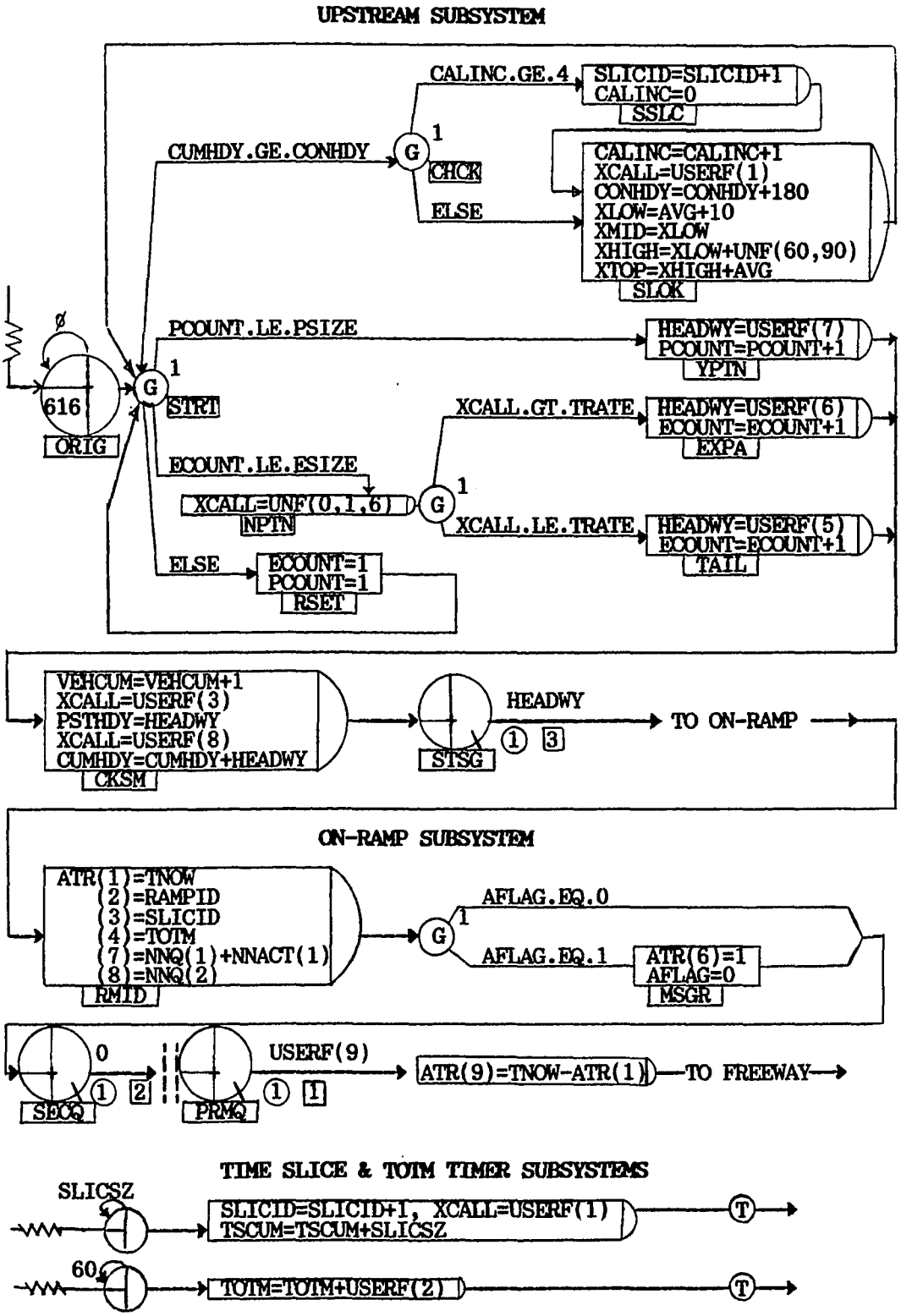


Figure 25. SLAM.II Network Diagram of ONRAMP Basic Building Block

(and only one) of four routes. The routing decision is made based on the need to update the timer, size of platoons PSIZE, and non-platoons ESIZE. When an entity goes to a particular branch, its counter is updated by 1. Possible routes from STRT are:

1. Entity is routed to CHCK to update the system timer and reassign new values to PRATE, ERATE, TRATE, AVG, PSIZE, and ESIZE. This update is done every 3 minutes (i.e., 4 times for each one time slice). The variable CONHDY is updated every 3 minutes (180 seconds). The 12-minute time slice is updated every 4 increments of CONHDY.
2. Entity is routed to a platooning node YPTN (yes platoon) where it is assigned a triangular headway distribution.
3. Entity proceeds without platooning to NPTN (no platoon) node where it is assigned a value for its headway from an exponential distribution at node EXPA or tail distribution at node TAIL. The assignment of exponential or tail values is randomized in proportion to TRATE. A uniform distribution value between 0 and 1 is generated and if it is < TRATE, entity is routed to TAIL and assigned a tail headway. Otherwise, it is routed to EXPA and assigned an exponential headway.
4. Entity is routed to RSET node to restart the exponential counter ECOUNT and the platoon counter PCOUNT and then routed back to STRT to begin the branching selection over-again.

Once an entity is assigned an appropriate headway, it goes through a check point at node CKSM to update the entity counter VEHCUM, shift headway by .5 to represent middle-cell position (e.g., 3.5 to represent cell from 3 to 4), update a headway histogram and update PSTHDY for use in subsequent checks. The entity then proceeds toward and gets held by a

queue STSG (street signal) where it is released based on its headway value. This process releases entities to the on-ramp according to the proposed probability distributions for headways.

When the entity arrives at the ramp at node RMID (ramp ID), its attributes are marked to distinguish entity's arrival time TOTM, ramp of origin RAMPID, time slice SLICID, the arrival minute TOTM, the number of vehicles in the primary queue NNQ(1)+NNACT(1), and finally the number of entities in the secondary queue NNQ(2) (see SLAM diagram figure 25).

Next, if the TOTM check flag AFLAG is 1, it means that it is an entity arriving near top-of-the-minute and the entity must be flagged as a messenger entity to check the system status (queue length and queue delays). The entity is routed to the MSGR node and flagged in attribute (6). Otherwise it bypasses the MSGR node to the secondary queue SECQ (queue # 2). If the secondary queue is "off", the entity passes through it without any delay to the primary queue at PRMQ. There, the entity progresses in the primary queue until it reaches the ramp meter which discharges entities at the rate of UNIFORM (STL, STH).

After the entity is released from PRMQ, it goes through a series of 11 collect nodes (EX01 through EX11, not shown here), depending on the current time slice SLICID. Information is collected on time in system, primary queue length PQL and secondary queue length SQL.

The "Time Slice Timer" is a SLAM subsystem to generate 11 time slices and advance the system by 12-minute every time. The "TOTM Timer" is another SLAM subsystem to generate 133 minutes and advance time by 1-minute at-a-time for system status checks at TOTM's.

Finally, if the entity is a messenger, it is called by USERF(8)

(also not shown in the diagram) to print TOTM status on service time ST; primary queue length PQL; secondary queue length SQL; and time in system. Otherwise entity is terminated. SLAM source code for the ONRAMP model and the FORTRAN USERF (user function) are presented in appendix A.

SENSITIVITY ANALYSIS & ONRAMP MODEL CALIBRATION

Over one-hundred simulation runs were executed to test the ONRAMP model and fine-tune it. Runs were initially made using the data set on College Avenue N., but then later other ramp data were used as well. The results of four sample runs on College Avenue North are presented below.

1. Results of Simulation Runs

Run I. All model parameters were set to their default values. AVG (exponential average) was set to 6.5 seconds. Suggested platoon size of 5 cars was used. Plot 1.1 of figure 26 shows the resulting distribution of headways of this run (the dashed line) vs. the distribution of RBM data (in solid line). The fit is not as good as previous ones obtained during model development. One reason is that not all "default" values were used exactly as suggested. A change in these parameters will change the shape of the distribution curve. Furthermore, the distribution shown is an aggregate one for the entire period 6:30-8:42 which results from using individual parameters (e.g., platooning rates) for each time slice as computed at each time slice individually. The Holistic view requires that the parts (individual time slices in this case) be modeled while being mindful of the whole (time slice 12) as one. The resulting "model whole" differed, but not significantly, from the "observed whole."

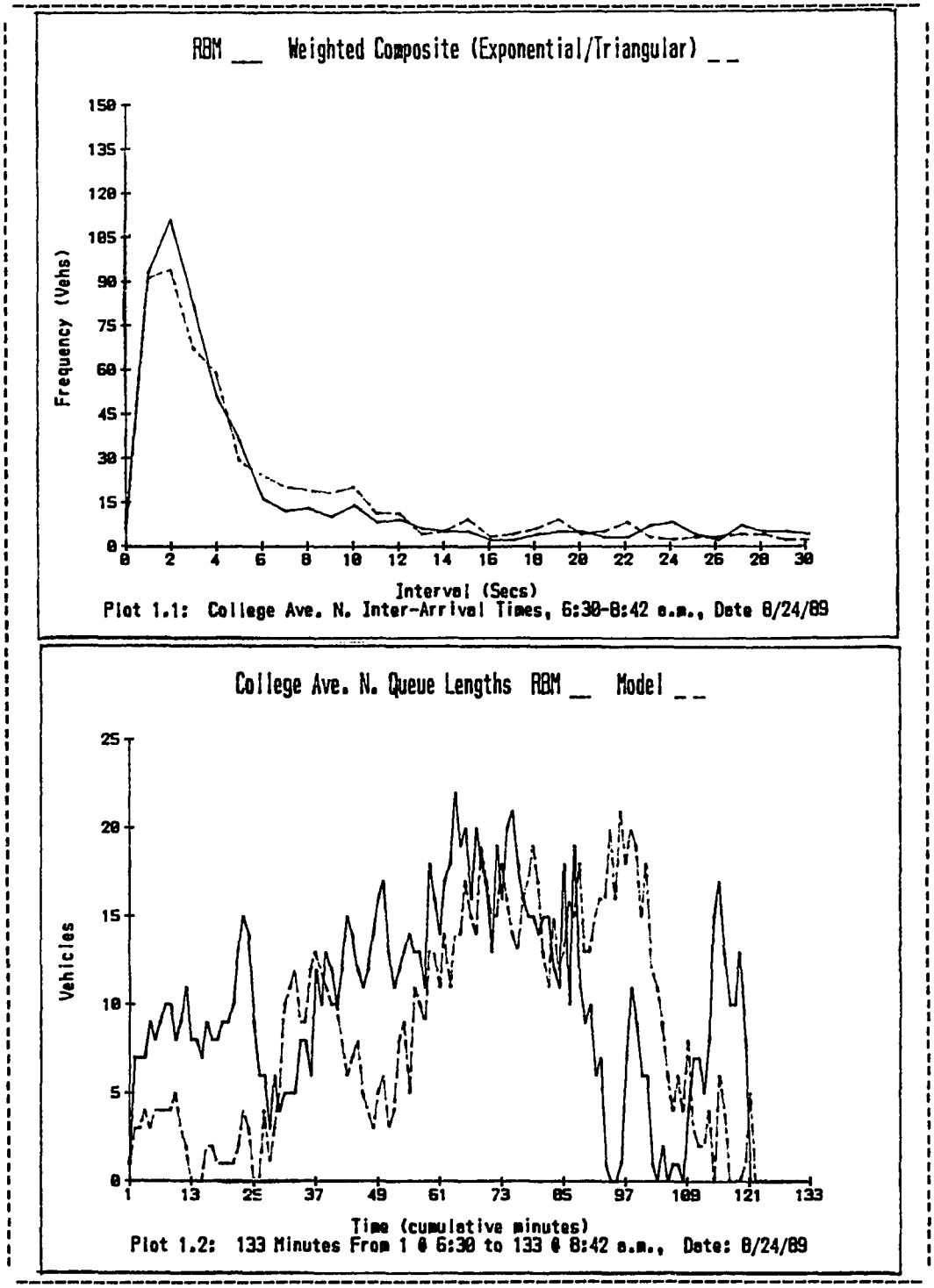


Figure 26. System Performance: Output of Run I for RBM and ONRAMP Headway Distribution (Plot 1.1) and Queue Lengths (Plot 1.2).

Plot 1.2 (figure 26) shows how the system performed in terms of queue lengths. 133 RBM (actual) TOTM (top-of-the-minute) queue lengths (solid line) were plotted along with 133 simulated queue lengths (dashed line). The figure shows a "lag" in the build-up of simulated queues. The simulated queues are shifted to the right in time. The peak in the actual queue length around the minutes 20-25 (6:49-6:54) seems to be shifted right towards the 37th minute (7:06).

The depression in RBM queue around minute 30 (6:59) also seems shifted in the model queue towards the 49th minute (7:18). RBM queue seems to have reached a global peak of about 23 cars around the 65th minute (7:34) whereas the model queue reached its peak of 22 cars around the 95th minute (8:04), a difference (lag) of 30 minutes.

Finally, the RBM queue length dropped to 0 around the 95th minute (8:04), while RBM queue dropped to 0 almost 20 minutes later at the 110th minute (8:19), a difference (lag) of about 15 minutes.

Run II. For this round, all model parameters were also set to their default values but the user-specified platoon length, PLONG, was set to 3 instead of the recommended value of 5, making the platoon size much smaller. AVG was left at 6.5 seconds. As shown in figure 27 plot 2.1, the smaller platoon size resulted in a closer fit for the headway distribution, especially for values between 2 and 6 seconds. However, the smaller platoon size also resulted in a more intermittent queue length (the dashed line plot 2.2 of figure 27), which differed considerably from the observed on (solid line).

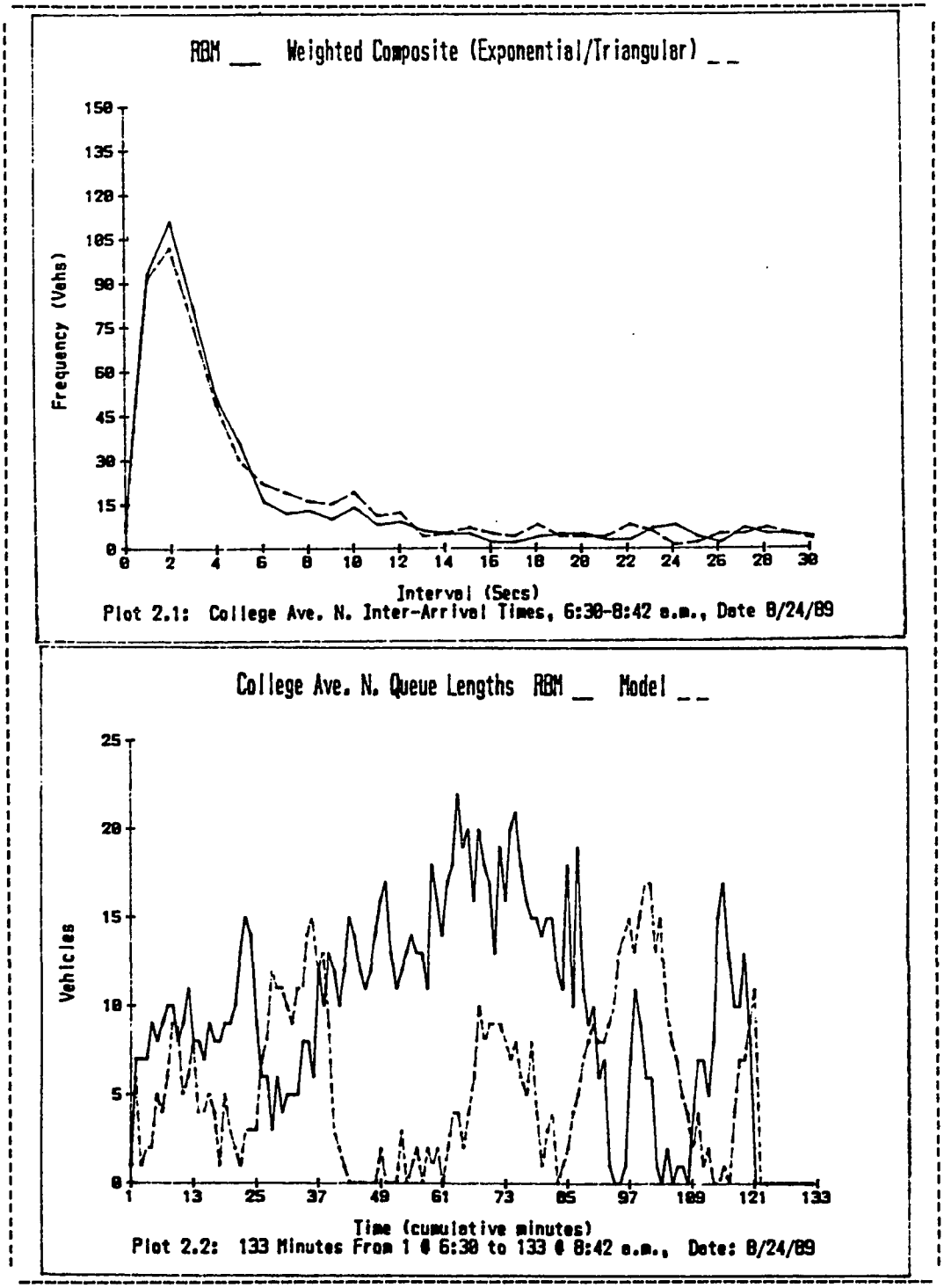


Figure 27. System Performance: Output of Run II for RBM and ONRAMP Headway Distribution (Plot 2.1) and Queue Lengths (Plot 2.2).

Run III. For this run, several changes were made. Parameters were set to the following values:

AVG= 7.5 seconds, as suggested from the data set at College N.

PLONG= 5 cars (platoon size as recommended) for all time slices.

TRATE= 0.35, default value= 0.3, but College data set suggests .35.

XMIN= -0.35, default value= -0.5, but data suggests -0.35.

PRATE= 0.5, as computed using the entire time period from 6:30 to 8:42 and not at individual time slices. Generally, no default values should be assumed here.

A very close fit for the headway distribution resulted from this run. Both the triangular (platoon) part and the exponential part of the simulated distribution curve were close to the RBM curve (see figure 28, plot 3.1). However, the nice fit did not necessarily result in a one-to-one correspondence in queue lines between the RBM and the ONRAMP.

As figure 28 (plot 3.2) shows, the observed "lag" was still apparent and excessive fluctuation in model queues (dashed lines) are evident. This is especially true with the rapid drop from 20 cars near the 37th minute to 0 near the 44th minute. Overall, this run resulted in best headway distribution and closet queueing to that of the observed data.

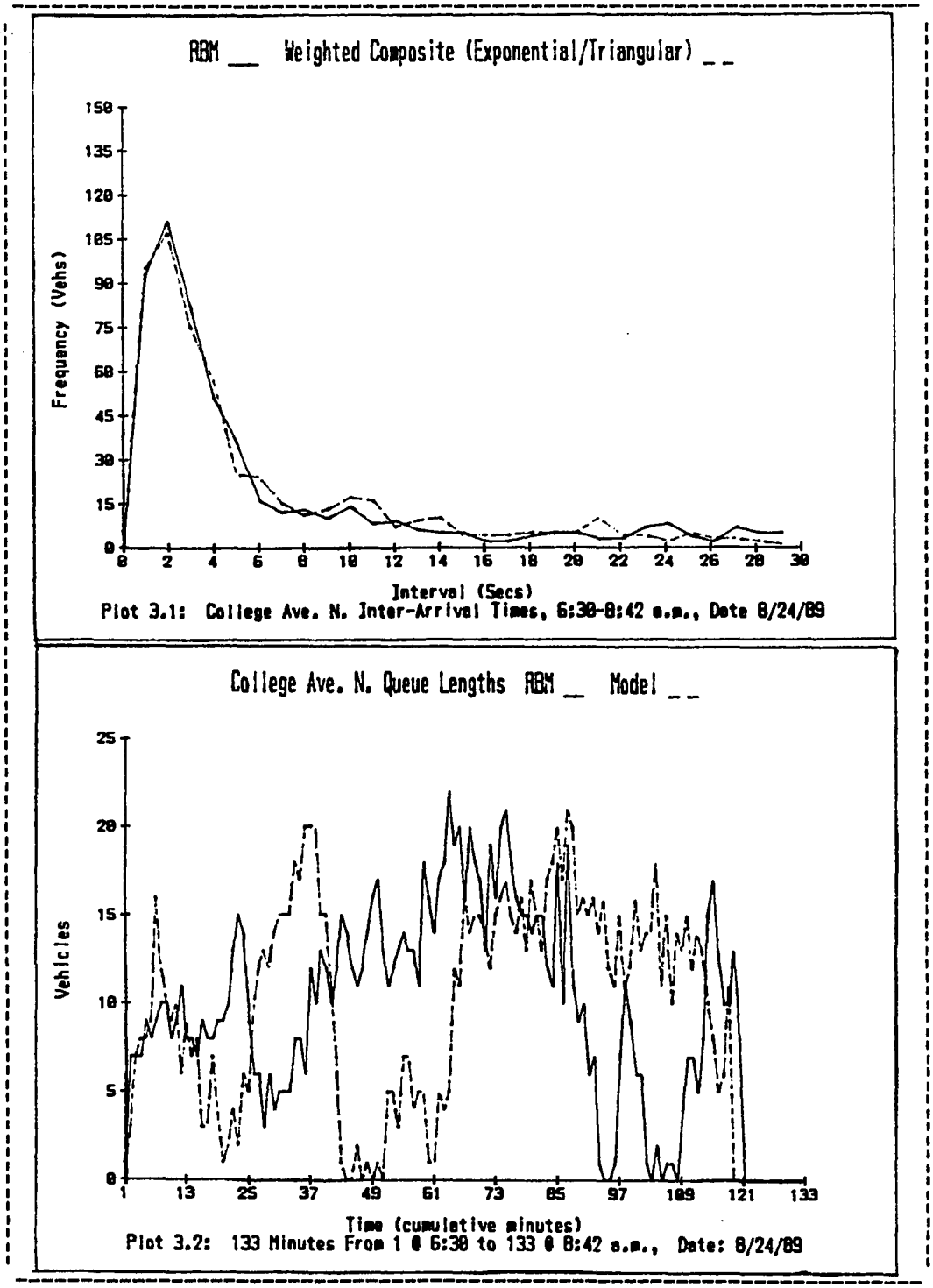


Figure 28. System Performance: Output of Run III for RBM and ONRAMP Headway Distribution (Plot 3.1) and Queue Lengths (Plot 3.2).

Run IV. Finally, all parameters were set back to their model default values, but AVG was reduced to 5.0, less than 1/2 of computed sample means for arrivals >= 4 seconds for College Avenue. The resulting headway distribution (figure 29, plot 4.1) shifted to the left and had a higher density between the values of 6 and 9 seconds. The tail part seemed slightly thinner too.

As expected, lowering AVG resulted in a heavier demand on the ramp. As plot 4.2 shows, the queue built up constantly and consistently. No correlation between RBM queues and the model queue was observed here.

Many more runs were made to test and evaluate ONRAMP. Simulation runs showed that the queue build up/dissipation is very sensitive not only to platooned arrivals, but also to tail arrivals as well. Tail arrivals provide the necessary "pause" to allow the queue to dissipate.

Most of the emphasis during model development was put on the platoon arrivals, their ratios and platoon size, and on the exponential distribution of headways less than 30 seconds. The distribution of inter-arrival times of over 30 seconds exhibited extreme variability. But in general, the higher the headway value, the lower its probability would be. Therefore, a triangular tail was assumed. The critical role of the rate and fashion of tail arrivals was not realized until the sensitivity analysis phase. A change as in as little as 5% in the tail rate could cause visible effects on queueing and delays.

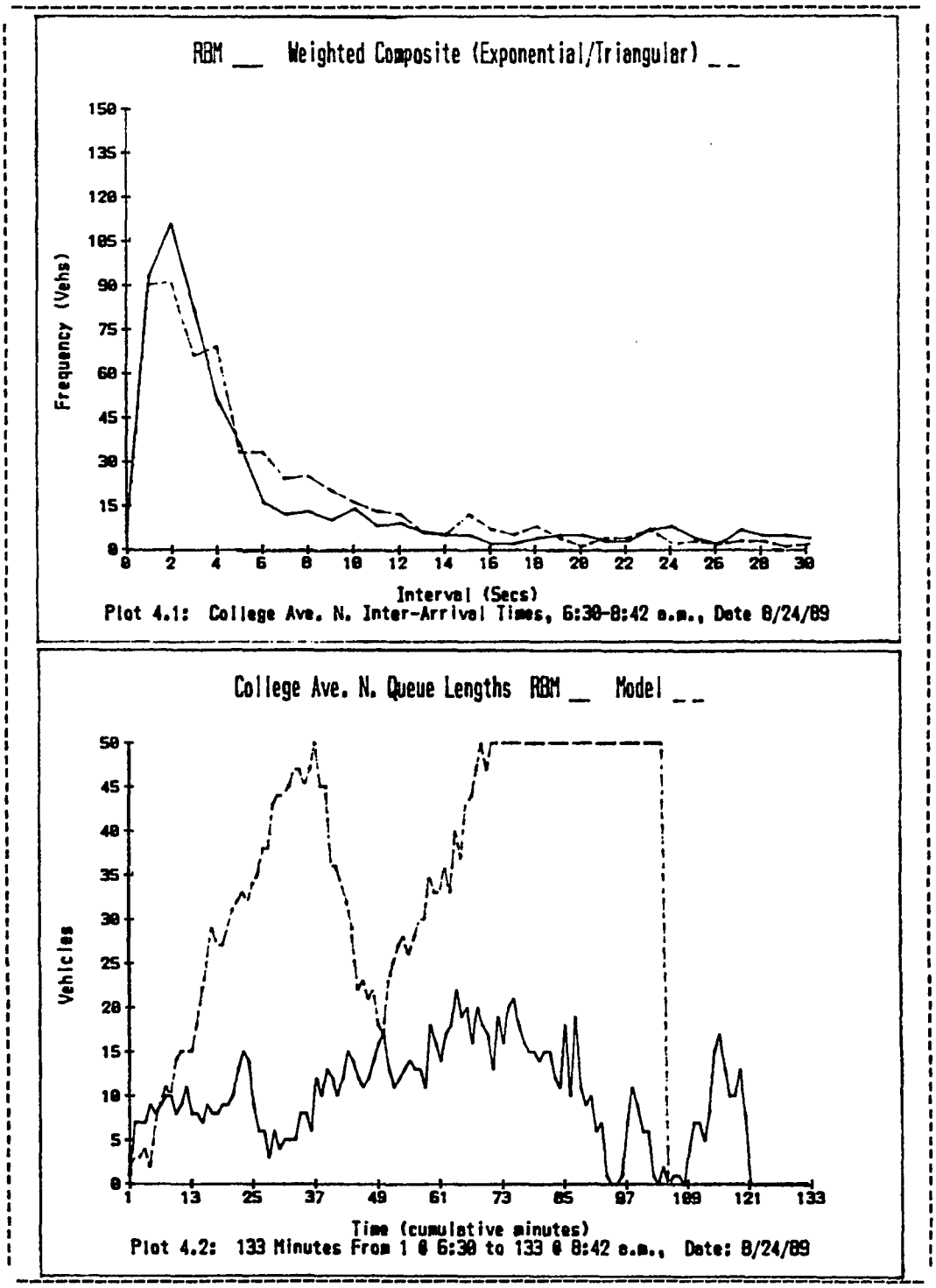


Figure 29. System Performance: Output of Run IV for RBM and ONRAMP Headway Distribution (Plot 4.1) and Queue Lengths (Plot 4.2).

2. Statistical Comparison Between RBM Queues and Simulated ONRAMP Queues

In order to test the model's performance in terms of queueing activities and provide more insight into the similarities/ dissimilarities between the real system performance and the model performance, the basic statistics of RBM queues and ONRAMP queues were computed and compared.

Tables X and XI contain the computed values of statistical parameters of the observed RBM queue lengths and delays and those which resulted from ONRAMP Run I and Run III above. Although figure 26 (plot 1.2) and figure 28 (plot 3.2) showed visible displacement between RBM and ONRAMP queues, the simple comparisons in tables X and XI did not indicate significant differences. The RBM and ONRAMP performance is especially similar in Run III, in which all suggested default parameters were used in ONRAMP. This run also produced very good fit for the distribution of headways (figure 28, plot 3.1).

The assumption here is that ONRAMP may be behaving essentially similarly to the real-world system, but with persistent "lag", or displacement in time. No further conclusions could be made until further model refinement and sensitivity analysis is performed on this particular point. No such improvements will be made in this report and none are contemplated at the present time.

TABLE X

STATISTICAL COMPARISON BETWEEN RBM QUEUE LENGTHS AND ONRAMP QUEUE LENGTHS AT COLLEGE AVENUE N.

	Queue Length (vehicles)		
	RBM	ONRAMP	
Statistical Parameter	Observations	Run I	Run III
N (number of observations)	115	110	110
Mean Value	10.92	9.41	11.10
Standard Deviation	4.91	5.73	5.33
Pearson skewness coefficient	0.257	-0.048	-0.512
Lower quartile	7.0	4.0	7.0
Median	10.5	9.5	12.0
Upper quartile	14.5	14.0	15.0
Inter-quartile Q75-Q25	7.5	10.0	8.0
90th percentile	18.0	17.0	17.0
Highest Observed Value	22.0	21.0	21.0

TABLE XI

STATISTICAL COMPARISON BETWEEN RBM QUEUE DELAYS AND ONRAMP QUEUE DELAYS AT COLLEGE AVENUE N.

	Queue Delay (seconds)		
	RBM	ONRAMP	
Statistical Parameter	Observations	Run I	Run III
N (number of observations)	115	110	110
Mean Value	139.6	115.0	132.0
Standard Deviation	62.0	67.0	59.0
Pearson skewness coefficient	0.488	-0.190	-0.277
Lower quartile	93.5	55.0	89.0
Median	130.0	119.0	138.0
Upper quartile	195.5	173.0	180.7
Inter-quartile Q75-Q25	98.0	119.0	92.0
90th percentile	222.0	207.0	204.0
Highest Observed Value	275.0	237.0	245.0

Note: All zero values were omitted in the above computations in both tables.

3. Further Testing of the Internal Relationships of the Model

The internal relationships between queue lengths (QL) and queue delays (QD) of both the RBM and ONRAMP were checked. This was to see if the model's QL-QD relationships were comparable to the RBM's QL-QD relationships. They are called "internal" because QL-QD relationship is tested within the RBM or within the ONRAMP itself.

The QL-QD relationship was assumed to be a linear one. This assumption was not only based on the logical relation between QL and QD but also on inspecting the scatter plot of QD vs. QL. Figure 30 is a scatter diagram of the observed (RBM) QD vs. QL. Based on the shape of the scatter plot, the linearity assumption could not be rejected.

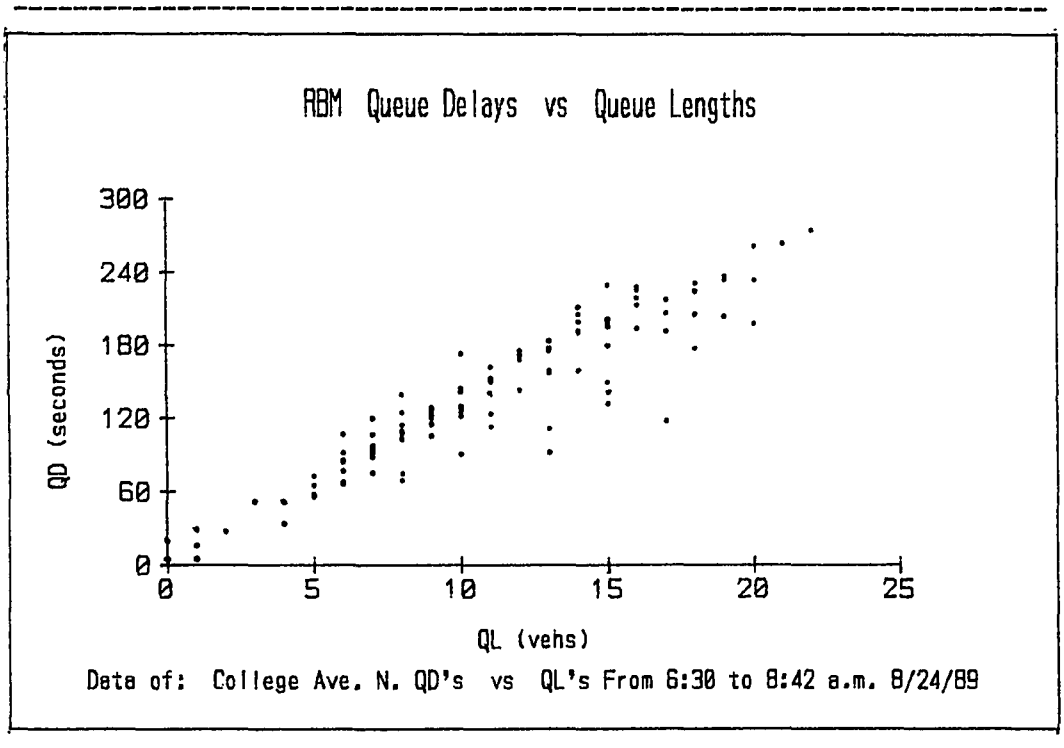


Figure 30. Scatter Diagram of RBM Queue Delays vs. Queue Lengths

Figures 31 and 32 are scatter diagrams for the QD vs. QL which resulted from ONRAMP Run I and III respectively. Although the points seem less scatered in figure 31 and 32, scatter in both diagrams does not differ remarkably from that of the RBM diagram in figure 30. Like the plot of figure 30, both plots in figure 31 and 32 seem to be scattered along a straight line. Neither one seems to contradict the linearity assumption between QL and QD.

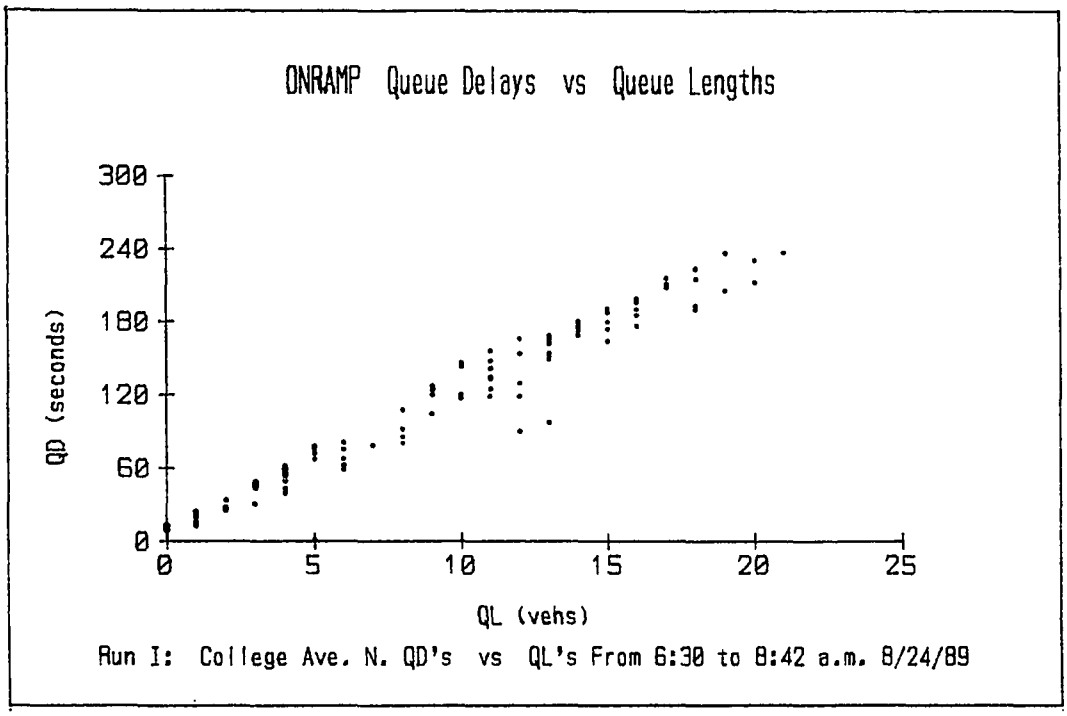


Figure 31. Scatter Diagram of ONRAMP Queue Delays vs. Queue Lengths From Run I.

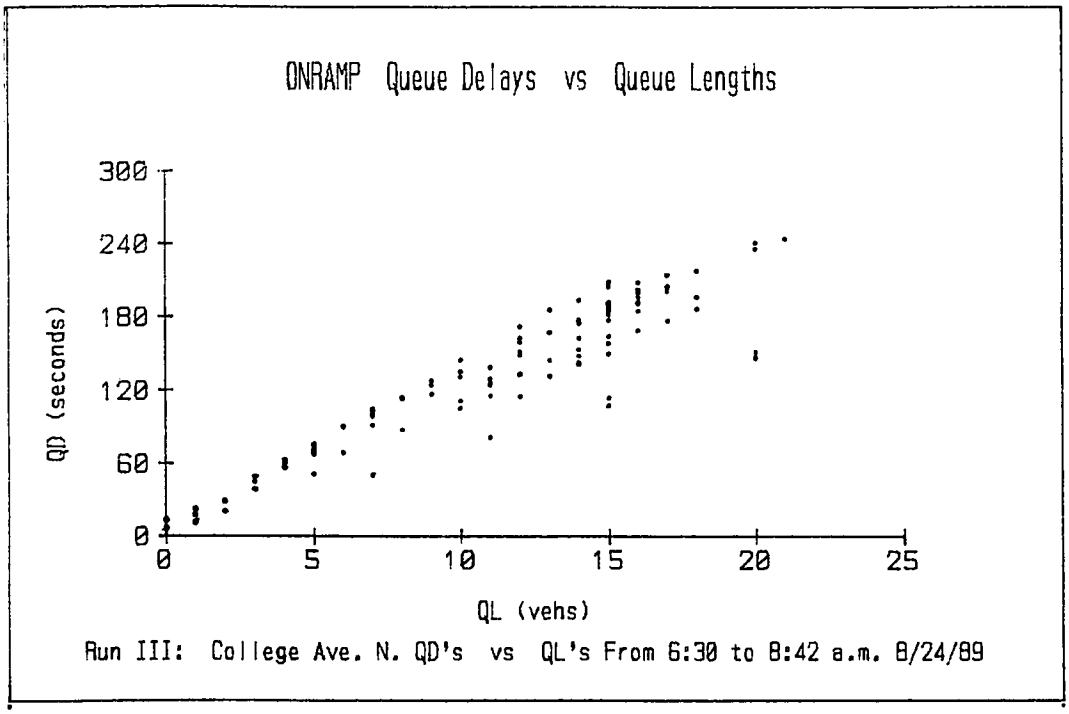


Figure 32. Scatter Diagram of ONRAMP Queue Delays vs. Queue Lengths From Run III.

To test the effect of QL on QD, a linear regression analysis was performed using QL as the independent variable and QD as the dependent one. The results of regression analysis are presented in table XII. Column two contains the results for the observed real-world data. Columns three and four contain the regression results of Run I and III above.

The assumption here is that QL-QD relationship is a simple and linear one, and that QD could be inferred from QL by the regression equation:

$$QD = C1 + C2 * QL$$

While C1 values were different for each column, C2 values were very

close. The results indicate very high correlation -as expected- between the queue length and the queue delays. This was true for the RBM (r=.92) and Run I (r=.977) and Run III (r=.929).

The linearity assumption between QL and QD is not completely verified. The QL-QD relationship could still be curvilinear near high values of QL. Once this point is determined with good degree of certainty, ONRAMP QL-QD relationship could be better verified based on RBM QL-QD relationship.

TABLE XII

RESULTS OF REGRESSION ANALYSIS OF RBM QUEUE DELAYS ON QUEUE LENGTHS AND ONRAMP MODEL QUEUE DELAYS ON QUEUE LENGTHS AT COLLEGE AVENUE N.

Statistical Parameter	RBM Queues	ONRAMP Queues Run I	ONRAMP Queues Run III
N (number of observations)	115	110	110
QL mean value	10.9	9.41	11.1
QD mean value	139.0	115.0	132.0
Total residual Sr (dispersion about Y=C1+C2*X)	68300.0	22352.0	52575.0
Sum of squared differences	447343.0	484631.0	382324.0
Standard error Syx	24.6	14.4	22.0
Regression equation coef. C1	12.6	8.1	17.7
Regression equation coef. C2 (for use in QD=C1+C2*QL)	11.6	11.4	10.4
Correlation coefficient	0.920	0.977	0.929

Note: All zero values were omitted in the above computations.

4. Comments on Sensitivity Analysis

Evidently, the ONRAMP model's initial simulation runs did produce arrivals very close to what is expected, especially in the platooned region. However, queueing seems to present some problems (shifting and quick queue formation/dissipation). The model needs calibration in this

area. The observed lag in ONRAMP queueing with respect to RBM queueing might be corrected by setting special initial conditions (i.e., number of cars initially in the queue, no initial pauses in arrivals etc.).

The model appears to be very sensitive to AVG (exponential mean) and TRATE (weight of extended tail). Even a slight change in TRATE (from 0.333 to 0.30) seems to considerably skew the queue positions forward or backward in time. A change in AVG from 6.3 seconds to 5.0 seconds (with every thing else constant) produced queues three time as long (maximum queue length increased from 21 to over 60). Therefore, AVG and TRATE must be well-estimated.

STRENGTHS AND WEAKNESSES OF THE ONRAMP MODEL

1. Strengths of the ONRAMP Model

The fundamental structure of the basic building block of ONRAMP is simple and coherent. ONRAMP is especially strong in the following areas:

a. The platoon part is very consistent and robust. With little change in parameters XMIN, XMOD, and XMAX, the model exhibits extreme flexibility and ability to reproduce *any* desired distribution shape. Testing the ONRAMP composite scheme on all ramps and across all time slices proved that it has *global* applicability.

b. ONRAMP requires a minimum number of parameter estimates (minimal calibration). The only parameters to estimate are the platooning ratio PRATE (ratio of arrivals < 4 seconds), the tail weight TRATE (ratio of extended pauses) and the mean value AVG for the exponential part. All other parameters could be set to default values unless certain special distribution shapes are desired.

c. ONRAMP could be used to model ramps with or without the following characteristics:

1. Platooned (clustered) and non-platooned vehicular arrivals.
2. Single or multiple ramp meters with static, dynamic, or random metering rates.
3. 1-car per green or 2-cars per green (or more cars per green).
4. Single or multiple primary queues (queues on ramp lanes).
5. Primary queue blockage and single or multiple secondary queues (queue overspills on surface streets).
6. M number of time slices, each is "SLICSZ" minutes long.
7. K number of top-of-the-minute "TOTM" system status checks.

d. All probabilistic components of ONRAMP block could be modified to deterministic ones if desired.

e. The model is relatively short and simple (ONRAMP computer section itself is about two pages long).

f. Unlike other traffic models that require intensive input data and origin-destination matrix specification, no exhaustive data, coding, or matrix formulation are required. No previous familiarity with probability theory is necessary. If every thing is set to default, only three parameters will be required: AVG (exponential average), PRATE (platooning pate) and TRATE (the tail rate) need to be specified. Even those may be guessed if necessary.

2. Limitations of the ONRAMP Model

Not much further work is required on ONRAMP to fine-tune its fundamental molecular unit or its arrivals composite distribution, which are consistent and coherent. However, the correlation between the tail

part and queueing appears to need more analysis. The weaknesses of the model are summarized below:

a. Estimating the AVG and TRATE parameters for the exponential and tail parts is done empirically with some uncertainty. A more concise way for estimating these parameters needs to be found. This may be done in future improvements on the ONRAMP.

b. Queueing dynamics are not as certainly or robustly modeled as vehicular arrivals. Queues seem to build up or dissipate too quickly. This is attributed to the distribution of tail (pause) arrivals. For example, if a queue length is 15 and if two long headways, say 60 seconds from the exponential distribution and 60 seconds from the tail distribution are generated in sequence, the queue will have only two arrivals during a 120 second period. If the service time is 10 seconds, 12 vehicles would be processed in the meantime and the queue length will be 15-12+2=5 vehicles. This produces very noticeable change in the QL plots. Due to the randomized arrival process, this kind of successive arrivals could occur, although with low probability. The model does not warren against nor preclude this event from happening.

c. ONRAMP is currently implemented on SLAM.II. SLAM.II is intended for very wide range of simulation applications from manufacturing systems, scheduling CPM networks, and automated warehouses to communications systems, banking services and sporting events. ONRAMP uses only very small portion of SLAM's diversified capabilities.

Although SLAM's randomization and entity processing capabilities were used very effectively and efficiently, ONRAMP development was slowed down by some different types of SLAM limitations. SLAM is

especially limited and cumbersome in conditional testing/branching. SLAM is also limited in performing extensive mathematical operations and number crunching, a common work in engineering applications. Only four arithmetic operations +, -, *, and / are currently allowed in SLAM.II Furthermore, SLAM.II does not allow the use of parenthesis in mathematical expressions. This meant that FORTRAN inserts were necessary to do what SLAM.II was incapable of doing, or what it does poorly. The author aimed at striking a balance between SLAM and FORTRAN inserts to optimize the model. However, some modelers and users consider FORTRAN inserts a source of weakness to this kind of simulation models because they require separate set of FORTRAN compilers and debuggers in the software system, and familiarity not only with SLAM, but also with the FORTRAN language.

---*---*---

CHAPTER VI

RESEARCH SUMMARY AND CONCLUDING REMARKS

CONCLUSIONS OF FREEWAY CONTROL VIA RAMP METERING

Freeway congestion is a condition which occurs when the number of vehicles entering the freeway exceeds the freeway capacity to accommodate them and move them at an acceptable level of service. One of the most effective measures of *direct* control of freeway congestion is *ramp metering*; a technique by which vehicle entry to the freeway is regulated by traffic signals (meters) at the entrance ramps. Ramp meters are run with calibrated influx rates to prevent highway saturation.

It has been demonstrated by literature reviews and field observations that in the field of freeway management, a well-designed metering scheme can help accomplish the following objectives:

a. Control freeway saturation.

b. Utilize available street capacity.

c. Disperse peak period traffic.

d. Redistribute traffic demand more uniformly throughout the highway corridor.

e. Regulate mainlane merging and improve continuity of flow on the freeway.

f. Reduce "total" trip time (by improving highway speeds).

g. Reduce the rate of freeway incidents (better safety).

h. Assign proper traffic priority for freeway and entering traffic.

The root cause of highway congestion is, on one hand, the rapid increase in car ownership and automobile use. On the other hand, the rate of new freeway construction has been declining for the past decade due to diminished funding and the growing anti-highway sentiment. Expansion of existing highway facilities in urban areas is constrained as well by numerous spatial and resource limitations. Therefore, any measure to curtail congestion will need to deal with these root causes of congestion, not only its symptoms. A well-planned metering strategy, could further contribute to the accomplishment of the following additional long-term goals:

i. Influence modal split and modal choice to increase car occupancy and use of mass transit.

j. Attain efficient use of the highway infrastructure.

k. Increase energy conservation and reduce pollution.

l. Identify some other needed system deficiencies.

It cannot be concluded that ramp metering is a flawless measure. For example, the repeated field observations at some sites in San Diego, CA have demonstrated how ramp metering, during peak-hour demand, tends to help the freeway flow while creating considerable traffic back-ups on local streets, thus transferring many congestion problems from the freeway to the nearby intersections.

Field observations, literature reviews, and the pilot study conducted in this project have identified the following problems. Notice that items "a" through "l" below have near one-to-one correspondence to the metering objectives "a" through "l" listed above.

a. Queue overspills on surface streets.

b. Congestion transfer from freeway to intersections.

c. Hardship on working people who are constrained in their trip time and route.

d. Meter rationing (for uniformity) is not always fairly done.

e. Metering does not always perform well. studies of off-peak-hour metering showed that it could lead to opposite results (more total delays). Metering may not be feasible at some locations because meters could hinder the process of merging when acceleration lanes are short. Also, signal surveillance and enforcement is difficult. If voluntary compliance is absent or if the meter violation rate is high, the premise of metering will be undermined.

f. Disproportionate delays between freeway and entrance ramps.

g. Increased street hazards due to queue overspills.

h. Denying freeway access will have socio-political implications.

i. Motorists are addicted to automobile use and there are opposing lobbying effort to perpetuate dependency on cars.

j, k, & l. Ramp metering imposes "individual" costs to gain "collective" benefits.

In summary, freeway control via ramp metering serves the *global* freeway interest well but that is often done at the cost of the *local* intersection interest. And the global freeway interest almost always prevails over the local one.

Interactions between freeway control via ramp metering and socio-economic trends that affect highway congestion such as urban and suburban development, employment concentrations, and demographics, and

the impacts of metering on the vitality of adjacent intersections, land use and development, public acceptance, enforcement difficulties, and the objectives and political power of organizations involved in decision making need to be viewed, understood and considered in order to devise better strategies for ramp metering. An outline of a systems multiple perspective approach has been proposed to help research in this regard.

ADVANTAGES OF SIMULATION MODELING

It was concluded that simulation modeling is an excellent systems analysis tool for freeway control via ramp metering. Simulation is a powerful technique that could be:

1. An alternative to "hard" analytical methods.

2. A multifaceted approach which can combine all the following in one package:

    a. Capacity analysis.

    b. Vehicle-driver "entity" processing.

    c. Discrete and continuous events handling.

    d. Stochastic representation.

    e. Queueing/shockwave analysis.

    f. Real-time/compressed time/expanded time simulations.

    g. If-else-then Testing.

    h. Decision making Processes.

    e. Graphics.

3. Applicable to: a. Systems engineering phase.

b. Operations research phase.

4. An alternative to "expensive/unsafe" field testing.

LIMITATIONS OF SIMULATION

Like any other systems analysis and operations research technique, simulation modeling has several limitations which are summarized below:

1. Simulation can be sometimes too complex.
2. It could be time-consuming (must be used only as "last resort").
3. Requires extensive/expensive input (i.e., data, system characterization, choice of variables, validation, etc.).
5. It can never duplicate human or social reality.
6. It could be deceptive (*capacity* to store/run models does not mean an *ability* to represent social complexity).
7. Large-scale Models are "guilty" of the seven sins [44]:
    a. Hypercomprehensiveness.
    b. Grossness.
    c. Hugeness.
    d. Wrongheadedness.
    e. Complicatedness.
    f. Mechanicalness.
    g. Expensiveness.
8. The most common modeling mistakes are to [61]:
    a. Measure the non-measurable.
    b. Disregard the non-measurable.
    c. Omit the "too difficult/messy".

QUALITIES OF GOOD MODELS

Simulation experts recommend having the following qualities in any simulation model in order for it to be a good model:

1. Simplicity, transparency, and clarity.
2. Balance between theory, objectivity, intuition, & applicability.
3. Key elements/variables which are well-identified, well-measured, and well-modeled.
4. A global approach to provide context for model development, validation, application, and deployment.
5. Probabilistic approach (if applicable) to count for uncertainty and provide adequate representativeness.
6. A mesoscopic stance which indicates combining microscopic system analysis and macroscopic model synthesis (whenever possible).
7. A holistic perspective which means attending the parts and their relationships while being mindful and appreciative of the whole.

RBM BASIC BUILDING BLOCK

Field observations and data analysis at the eight on-ramp sites in San Diego indicated that the basic building block of the on-ramp has the the following components:

1. **VEHICLE SOURCE**: Arrivals from each source were classified into three categories:
   a. Platooned (headways of 0.0 - 3.5 seconds).
   b. Semi-Platooned (headways of 3.5 - 15.0 seconds appx.).
   c. Dispersed (headways of 15.0 - 120.0 seconds appx.).

2. **QUEUE(s)**: Ramp queues were divided into two categories:

Primary: Contained, delineated, fully observable, easy to measure, and simple to model.

Secondary: Boundary-less, difficult to observe (mixed with street traffic), difficult to measure (no boundaries), and hard to model but may be modeled by making some simplifying assumptions.

3. **METERS** (& RMLL): Those are best represented as servers and service times. Service times are normally determined by freeway conditions, but based on observed data, they were considered approximately uniformly distributed.

ONRAMP BASIC BUILDING BLOCK

The basic building block of the ONRAMP model was conceptualized after the RBM basic building block. The following components were found best to devise and construct the ONRAMP basic building block:

1. **CREATE NODE** (to generate entities).
2. **PLATOONING BLOCK**: This block has three components:
    a. Triangular distribution (to model platooned arrivals).
    b. Exponential distribution (to model semi-platooned arrivals).
    c. Triangular (tail) distribution (to model dispersed arrivals).
3. **SQ NODE**: Capacity = Unlimited.
4. **PQ NODE**: Capacity = RBM observed capacity.
5. **SERVER**: With uniform distribution service time.

COMMENTS ON ONRAMP MODEL CALIBRATION, VALIDATION, & SENSITIVITY ANALYSIS

ONRAMP arrival processing was validated mathematically by proving that the cumulative distribution function (cdf) of the proposed composite function for inter-arrival times (time headways) was F(x)=1. The distribution of headways generated by ONRAMP was verified empirically by comparing it with that of the RBM. Methods of curve fitting and chi-square test were used to test these distributions at five ramps and eleven, 12-minute time slices (covering the period from 6:30 to 8:42).

ONRAMP Simulated queue dynamics (queue lengths QL and queue delays QD) were partly validated by four ways:

1. Visual inspection of observed RBM and simulated ONRAMP queues for same simulated periods.
2. direct statistical comparisons with the observed queue dynamics of the RBM.
3. Visual inspection of scatter plots of QD vs. QL for both observed RBM queues and ONRAMP simulated queues.
4. Regression analysis of QD on QL for both observed RBM queues and ONRAMP simulated queues.

ONRAMP STRENGTHS

Many of ONRAMP strengths come from the elaborate work which was invested during the pre-model development phase (i.e., the system perspective, the Holistic approach, the extensive data collection and analysis of observations to build the RBM). The assumptions which were

built into ONRAMP make it flexible and easy to apply. The strengths of ONRAMP are summarized below:

1. ONRAMP is short and simple. ONRAMP detailed diagram fits in one page and its SLAM.II source code is only two-page long.
2. The platooning block is consistent and robust.
3. It has extreme flexibility and global applicability (with some more model calibration, it may be applied to any on-ramp).
4. It needs minimum data input preparation (no matrices etc.).
5. It requires minimum number of parameter estimation (only PRATE, TRATE and AVG are currently needed).
6. It is applicable & expandable to any ramp configuration with the following characteristics:
   a. Platoons / no-platoons.
   b. Single / multiple / none ramp meters (static, dynamic, or random metering rates).
   c. 1-car/green or 2-cars/green (or more).
   d. Single / multiple primary Queues.
   e. Primary queue blockage.
   f. M number of time slices.
   g. K number of Top-Of-The-Minute "TOTM" system status checks.
7. Its probabilistic components are switchable to deterministic.

ONRAMP LIMITATIONS

Because of some shortfalls during model development and of the uncertainty in some assumptions, ONRAMP has some limitations. Some of the weaknesses were exhibited during the sensitivity analysis phase.

These limitations are listed below:

1. The proposed method of parameter estimation for AVG & TRATE is weak and needs further improvement.
2. ONRAMP is very sensitive to minor changes in AVG and TRATE. Further model calibration on queueing dynamics (queue formation/dissipation) is needed.
4. It needs more testing near edges, crash points, etc.
5. It utilizes only a small portion of SLAM.II capabilities.
6. It is implemented using SLAM.II and SLAM.II has some gross limitations such as:
   a. Cumbersome conditional testing/branching.
   b. Only +, -, *, and / operators are allowed.
   c. FORTRAN inserts were necessary (added complexity).

RESEARCH ACCOMPLISHMENTS & CONTRIBUTION(S)

The significance of this research stems from the fact that it has identified, analyzed, and proposed some techniques to help solve a growing problem (which ought to be addressed sooner rather than later). Highway congestion will only increase requiring more ramp metering. More metering will complicate existing problems and create new ones.

The ramp metering technique has been investigated in a multiple-perspective and comprehensive manner which included technical and non-technical elements. An outline of a systems approach to investigate the problem of freeway control via ramp metering has been proposed and partly tested and implemented in a pilot study. This approach properly places the study of ramp metering within a larger systems context.

The simulation modeling alternative has been thoroughly investigated to check its feasibility as a systems analysis and an operations research tool for the ramp metering application. All major traffic modeling approaches (microscopic, mesoscopic, and macroscopic) have been examined and analyzed. Existing highway simulation models were also reviewed and discussed.

The reference behavior mode which has been formulated for the metered segment in San Diego could be used to guide further simulation efforts for freeway ramp meters. The stochastic intermediate headway state of ramp arrivals has been examined and advanced by proposing a new composite headway distribution.

ONRAMP basic building block is needed in the area of traffic modeling. It is ready for further development, enhancement, and deployment. The potential applications of ONRAMP include highway models, stop-controlled intersection models, signalized intersections models, manufacturing models, parcel processing models and other entity-oriented modeling areas.

The prospective beneficiaries of the outcome of this project are the transportation industry, traffic engineering professionals, system analysts, and modeling and simulation practitioners.

---*---*---

REFERENCES

[1] Baerwald, Thomas J., "Land Use Changes in Suburban Clusters & Corridors." TRB, *Transportation Research Record 861*, 1982.

[2] Blumentritt, C.W., Pinnel, C., McCasland, W. R., Ross, D. W., & Glazer, J., "Guidelines for Selecting Ramp Control System." *Transportation Research Record 232*, TRB, 1981.

[3] Buffington, J. L., Ohul, M. K., & Memmott, Jeffery L., "Effects of Freeway Stage Construction on Nearby Land Uses and Relative User Costs." *Transportation Research Record 1046*, TRB, 1985.

[4] California Department of Transportation, *The Effects of Ramp Metering On City Streets*. Caltrans HQ, Sacramento, CA, 1979.

[5] California Department of Transportation, *Traffic Volumes*. Caltrans District 11, San Diego, CA, 1988.

[6] California Department of Transportation, *Long Range Operations Plan*, Caltrans District 11, San Diego, CA, 1988.

[7] California Department of Transportation, *Report To The Legislature -San Diego Traffic Control Center, Feasibility Study*. District 11, San Diego, CA, 1989.

[8] Corcoran, L. and Hickman, G., "Freeway Ramp Metering Effects in Denver." ITE 1989 Compendium of Technical Papers, ITE, 1989.

[9] Cotugno, Andrew C. (Metro Transportation Director, Metropolitan Service District, Portland, OR), Personal Interview, April 1989.

[10] Dotterrer, Steve (Architect, Chief Urban and Transportation Planner, City of Portland, OR), Personal Interview, May 1989.

[11] Englen, Rodney E. "Coordination of Transportation System Management and Land Use Management." *Transportation Research Record 93*, TRB 1982, pp 2-4,25, & 33-34.

[12] Federal Highway Administration, *Ramp Metering, Signal for Success* (videotape). U.S. Dept. of Transportation, Washington, D.C., 1989.

[13] Foss, Michael L. (Sergeant, Bureau of Police, City of Portland, OR), Personal Interview, May 1989.

[14] Highway Research Board, *1965 Highway Capacity Manual, SR 87*. National Academy of Sciences, Pub. 1328, Washington, D.C., 1982.

[15] Kell, James H. & Associates, *Before and After Study, Ramp Metering, Eastbound Long Island Expressway, Suffolk County.* Technical Memorandum, April 1989.

[16] Kostyniuk, Lidia P., et.al., *An Evaluation of the Detroit Freeway Operations (SCANDI) Project.* Michigan State University, June 1988.

[17] Linstone, Harold, *Multiple Perspective For decision Making.* Elsevier, New York, 1984.

[18] Marsden, Blair, *Ramp Meters and Travel Quality in Austin, Texas.* Texas DH&PT, May 1981.

[19] Mehyar, Omar & Kim, Henry, *Six-Year Flow Evaluation.* Washington State Department of Transportation, January 1989.

[20] Michael W. Bauer (Traffic Engineer, City of Portland, OR), Personal Interview, May 1989.

[21] Minnesota Department of Transportation, *Freeway Traffic Management Program: Status Report.* Minneapolis, MN, January 1989.

[22] Oregon Department of Transportation, *City-State Agreement I-5 Ramp Metering.* Portland, OR, 1980.

[23] Oregon Department of Transportation, *Ramp Metering Managing Traffic.* ODOT, Portland, OR, 1981.

[24] Oregon Department of Transportation, *First Two Weeks of Operation, I-5 North Freeway Metering.* Portland, OR, 1981.

[25] Oregon Department of Transportation, *Ramp Metering Technical Advisory Committee Meeting Proceedings.* Portland, OR, 1982,

[26] Oregon Department of Transportation, *I-5 North Freeway Ramp Metering, Portland, Oregon, Project Development Operation.* 1982.

[27] Potter, Stephen, "The Transportation Versus Land Use Dilemma." *Tranqportation Research Record 964*, TRB, 1984.

[28] Robinson, J. & Doctor, M., "Ramp Metering Status In North America." Presentation at the ITE Conference, San Diego, CA, September 1989.

[29] Schwab, Thomas (Transportation Analysis Manager, Oregon Department of Transportation), Personal Interview, May 1989.

[30] U.S. Department of Transportation, *Guidelines for Assessing Environmental Impacts of Public Mass Transportation Projects.* Report No. DOT-P-79-00-001, April 1979, (quoted in Meyer, M. *Urban Transportation Planning*, 1984).

[31] U.S. General Accounting Office, *Traffic Congestion Trends, Measures, and Effects*. Pub. GAO/PEMD-90-1. Washington, D.C., 1989.

[32] Winston, Clifford, "Efficient Transportation Infrastructure Policy." *Journal of Economic Perspectives*, V5, No. 1, Winter 1991.

[33] Aerde, M. V., Yagar, S., Ugge, A. and Case, E. R., "A Review of Candidate Freeway-Arterial Corridor Traffic Models." *Transportation Research Record 1132*, TRB, National Research Council, Washington, D.C., 1987.

[34] Babcock, P. S., Auslander, D. M., Tomizuka, M, and May, A. D., "The Role of Adaptive Discretization in a Freeway Simulation Model." *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984.

[35] Benekohal, R. F. and Treiterer, Joseph, "CARSIM: Car-Following Model for Simulation of Traffic in Normal and Stop-and-Go Conditions." *Transportation Research Record 1194,* TRB, National Research Council, Washington, D.C., 1988.

[36] Chang, G., Mahmassani, H., and Herman, R., "Macroparticle Traffic Simulation Model (MPSM) to Investigate Peak-Period Commuter Decision Dynamics." *Transportation Research Record 1005,* TRB, National Research Council, Washington, D.C., 1985.

[37] Cohen, S. and Clark, J., "Analysis of Freeway Reconstruction Alternatives Using Traffic Simulation." *Transportation Research Record1132*, TRB, National Research Council, Washington, D.C., 1987.

[38] Davila, M. C. and Lieberman, E. B., "Hybrid Macroscopic-Microscopic Traffic Simulation Model (LEVEL I)." *Transportation Research Record 772,* TRB, National Research Council, Washington, D.C., 1980.

[39] Derzko, N. A., Ugge, A. J., and Case, E. R., "Evaluation of Dynamic Freeway Flow Model By Using Field Data." *Transportation Research Record 905,* TRB, National Research Council, Washington, D.C., 1983.

[40] Drake, T. S., Schofer, J. L., and May, A. D., *A Statistical Analysis of Speed-Density Hypothesis*. Expressway Surveillance Project, Chicago, Report 16, 1965.

[41] Fox, Phyllis and Lehman, Frederick, "A Digital Simulation of Car Following and Overtaking." *Highway Research Record 199*, HRB, 1976.

[42] Greenshields, B. D., "A Study in Highway Capacity." *HRB Proceedings*, Vol. 14, 1934.

[43] Greenberg, H., "An Analysis of Traffic Flow." *Operations Research*, Vol. 7, 1959.

[44] Lee, Douglass B. Jr., "Requiem for Large-Scale Models." *AIP (American Institute of Planners) Journal*, May 1973.

[45] Lendaris, G. G., "On Systemness and the Problem Solver: Tutorial Comments." *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-16, Number 4, July/August 1986.

[46] Lieberman, E. B., *Macroscopic Simulation for Urban Traffic Management, Vol. 1: Executive Summary*. KLD Associates, Huntington Station, NY, January 1982.

[47] Lighthill, M. T., and Witham, G. B., "A Theory of Traffic Flow on Long Crowded Roads." *Royal Society Proceedings*, London, Series A, Vol. 229, 1955.

[48] Mannering, F. L., *Generation and Assessment of Incident Management Strategies, Summary Report*. Washington State Transportation Center, University of Washington, 1989.

[49] Mikhalkin, B. et al., *Changes Needed in Simulation Models to Test Freeway Bottleneck Solutions*. JFT Associates, Los Angeles, CA, November 1983.

[50] May, Adolf and Imada, Tstoumu, *FREQ10PE: A Freeway Corridor Simulation & Ramp Metering Optimization Model*. Institute of Transportation Studies, Berkeley, CA, 1985.

[51] May, Adolf D., "Freeway Simulation Models Revisited." *Transportation Research Record 1132*, TRB, National Research Council, Washington, D.C., 1987.

[52] May, Adolf D., *Traffic Flow Fundamentals*. Prentice-Hall, Englewood Cliffs, NJ, 1990.

[53] Payne, H. J., "FREFLO: A Macroscopic Simulation Model of Freeway Traffic." *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., 1979.

[54] Pipes, L. A., "An Operational Analysis of Traffic Dynamics." *Journal of Applied Physics*, Vol. 24, 1953.

[55] Pritsker, A. Alan B., *Introduction to Simulation and SLAM II*. Wiley, New York, NY, 1986.

[56] Rathi, A. K. and Nemeth, Z. A., "FREESIM: A Microscopic Simulation Model of Freeway Lane Closures." *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986.

[57] Rathi, A. K., Lieberman, E. B., and Yedlin, M., "Enhanced FREFLO: "Modeling of Congested Environments." *Transportation Research Record 1112,* TRB, National Research Council, Washington, D.C., 1987.

[58] Reuschel, A., "Fahrzeugbewegungen in der Kolonne." *Oesterreichishes Ingenieurarchiv*, Vol. 95, West Germany, 1950.

[59] Ross, Paul., "Some Properties of Macroscopic Traffic Models." *Transportation Research Record 1194,* TRB, National Research Council, Washington, D.C., 1986.

[60] Shannon, Robert, *Systems Simulation the Art and the Science.* Printice-Hall, Englewood, NJ, 1975.

[61] Smith, A. (pseudonym), *Supermoney.* Random House, (quoted in Linstone, 1984, reference 17) New York, NY, 1972.

[62] Waurick, P., *Evaluation of the INTRAS Microscopic Freeway Simulation Model.* Ministry of Transportation and Communications, Downsview, Ontario, Canada, September 1985.

[63] Williams, J. C., Mahmassani, H. S., and Herman, R., "Urban Traffic Network Flow Models." *Transportation Research Record 1112,* TRB, National Research Council, Washington, D.C., 1987.

[64] Zeigler, Bernard P., *Multifacetted Modeling and Discrete Event Simulation.* Academic Press, New York, 1984.

[65] Billheiemer, J., Kaylor, K. and Shade, C., *Use of Videotape in HOV Lane Surveillance and Enforcement.* US Department of Transportation, DOT-T-91-02, March 1990.

[66] Daou, A., "The Distribution of Headways in a Platoon." *Operations Research.* Vol. 12, No. 2, 1964.

[67] Dawson, R. F. and Chimini, A., "The Hyperlang Probability Distribution --A generalized Traffic Headway Model." *Highway Research Board Record 230*, HRB, Washington, D.C., 1968.

[68] Gerlough, Daniel L. Barnes, Frank C., and Schuhl, Andre., *Poisson and Other Distributions in Traffic.* Eno Foundation, Saugatuck, Conn., 1971.

[69] Greenberg, I., "The LogNormal Distribution of Headways." *Australian Road Research*, Vol. 2, No. 7, March 1966.

[70] Tolle, J. E., "The Lognormal Headway Distribution Model." *Traffic Engineering and Control*, Vol. 13, No. 1, 1971.

---*---*---

APPENDIX

SLAM.II SOURCE CODE (NETWORK PROGRAM)
& FORTRAN "USERF" INSERT

A.1
SLAM.II NETWORK PROGRAM

```
GEN, ALKADRI, SYSTEMS MODELING,6/10/91,1,Y,,,,,132;
LIMITS,3,10,20000;                      RESERVE 5 FILES,
                  ;                     10 ATTRIBUTES/ENTITY,
                  ;                     & SPACE FOR 20000 ENTITIES
EQUIVALENCE/ATRIB(5),HEADWY/
              XX(1),AVG/
              XX(2),ST/
              XX(3),SD/
              XX(4),TNE/
              XX(5),STL/
              XX(6),STH/
              XX(7),ESIZE/
              XX(8),TRATE/
              XX(9),PSIZE/
             XX(10),TLSTEP/
             XX(11),XLOW/
             XX(12),XMID/
             XX(13),XHIGH/
             XX(14),XTOP/
             XX(15),XMIN/
             XX(16),XMOD/
             XX(17),XMAX/
             XX(20),XCALL/
             XX(21),PSTHDY/
             XX(22),PLONG/
             XX(23),PRATE/
             XX(24),CUMHDY/
             XX(25),VEHCUM/
             XX(26),XLMT/
             XX(27),ECOUNT/
             XX(28),TLNEXT/
             XX(29),PCOUNT/
             XX(30),CONHDY/
             XX(31),TOTM/
             XX(32),SLICID/
             XX(33),SLICSZ/
             XX(34),TSCUM/
             XX(35),TOTMQ1/
             XX(36),TOTMQ2/
             XX(38),RAMPID/
             XX(87),TOTMNW/
             XX(37),CALINC/
             XX(88),AFLAG/
;
```

```
;INITIALIZE ALL GLOBAL SYSTEM VARIABLES/VECTORS:
INT,RAMPID=3,AVG=6.72,ST=10,SD=.5,TNE=616,STL=9.5,STH=10.5;
INT,ESIZE=6,XLMT=3.5,TRATE=.15,XLOW=10,XMID=10,XHIGH=90,XTOP=95;
INT,PSIZE=4,XMIN=-.5,XMOD=2.0,XMAX=4.5,PLONG=5,PRATE=.5;
INT,PSTHDY=0,VEHCUM=0,CUMHDY=0,CONHDY=0,ECOUNT=0,PCOUNT=0,TLNEXT=10;
INT,TOTM=0,TOTMNW=0,AFLAG=1,SLICID=1,SLICSZ=720,TSCUM=0,CALINC=0;
;
; SLAM.II Network Statement              Comments
;=========================-------------================================
;
;-----------------------BEGIN SYSTEM NETWORK--------------------------
NETWORK;
;--------------------UPSTREAM SUBSYSTEM ACTIVITIES---------------------
;
ORIG   CRE,0,0,,620,1;                  GENERATE TNE ENTITIES/SLICE
STRT   GOON,1;
        ACT,,CUMHDY.GE.CONHDY,CHCK;
        ACT,,PCOUNT.LE.PSIZE,YPTN;      ELSE, GENERATE TRIAN'LAR HEADWYS
        ACT,,ECOUNT.LE.ESIZE,NPTN;      GENERATE ESIZE EXPONEN'L HEADWYS
        ACT,,,RSET;                     ELSE RESET COUNTERS
CHCK   GOON,1;
        ACT,,CALINC.GE.4,SSLC;          CHECK CALLING INCREMENT
        ACT,,,SLOK;                     TIME SLICE IS OK
SSLC   ASS,SLICID=SLICID+1;             ADVANCE TIME SLICE BY 1
        ASS,CALINC=0;                   RESET CALL INCREMENT
SLOK   ASS,CALINC=CALINC+1;             UPDATE CALLING INCREMENT
        ASS,XCALL=USERF(1);             ASSIGN MEANS & STD DEV'S
        ASS,CONHDY=CONHDY+180;          ADVANCE BY 12 MINUTES (720 SEC)
        ASS,XLOW=AVG+10;                SET TAIL LOW END
        ASS,XMID=XLOW;                  SET TAIL MODE
        ASS,XHIGH=XLOW+UNF(60,90);      SET TAIL HIGH END
        ASS,XTOP=XHIGH+AVG;             SET MAX SUCC. ENTRIES
       GOON,1;
        ACT,,,STRT;
RSET   ASS,ECOUNT=0,PCOUNT=0;           RESET COUNTERS
       GOON,1;
        ACT,,,STRT;                     GO GENERATE NEW BATCHES
NPTN   ASS,XCALL=UNF(0,1,6);            GET RV BETWEEN 0 AND 1
       GOON,1;
        ACT,,XCALL.GT.TRATE,EXPA;       EXPON HEADWAY IF > TRATE
        ACT,,XCALL.LE.TRATE,TAIL;       OR TAIL HEADWAYS IF < TRATE
EXPA   ASS,HEADWY=USERF(6),1;           SET EXPONEN'LY DISTR'D HEADWAYS
        ASS,ECOUNT=ECOUNT+1;            UPDATE EXPONEN'L COUNTER
       GOON;
        ACT,,,CKSM;                     CHECK SUM HEADWY(I)+HEADWY(I-1)
TAIL   ASS,HEADWY=USERF(5);             SET TAIL TRIANGULAR HEADWAYS
        ASS,ECOUNT=ECOUNT+1;            UPDATE EXPONEN'L COUNTER
       GOON,1;
        ACT,,,CKSM;                     CHECK SUM HEADWY(I)+HEADWY(I-1)
YPTN   ASS,HEADWY=USERF(7);             SET PLATOONED HEADWAYS
        ASS,PCOUNT=PCOUNT+1;            UPDATE PLATOONED COUNTER
```

```
CKSM    ASS,VEHCUM=VEHCUM+1;              UPDATE CUMULATIVE VEHICLES
         ASS,XCALL=USERF(3);              ROUND HEADWY & UPDATE HISTOGRAM
         ASS,PSTHDY=HEADWY;               SAVE HEADWY FOR COMPARISON
         ASS,XCALL=USERF(8);              SHIFT BACK TO REAL-WORLD TIME
         ASS,CUMHDY=CUMHDY+HEADWY;        UPDATE CUMULATIVE TIME
STSG    QUE(3);                           QUEUE AT UPSTREAM STREET SIGNAL
         ACT(1)/3,HEADWY;                 RELEASE ENTITIES AT HEADWY PACE
;
;-------------------ONRAMP SUBSYSTEM & ACTIVITIES----------------------
;
RMID    ASS,ATR(1)=TNOW;                  MARK ARRIVAL TIME
         ASS,ATR(2)=RAMPID;               ASSIGN RAMP ID
         ASS,ATR(3)=SLICID;               MARK SLICE ID
         ASS,ATR(4)=TOTM;                 MARK TOTM ID
         ASS,ATR(7)=NNQ(1)+NNACT(1);      SAVE PRIMARY QUEUE LENGTH
         ASS,ATR(8)=NNQ(2);               SAVE SECONDARY QUEUE LENGTH
        GOON,1;
         ACT,,AFLAG.EQ.0,SECQ;            BYPASS MESSENGER ASSIGNMENT
         ACT,,AFLAG.EQ.1;MSGR             FLAG ENTITY AS MESSENGER
MSGR    ASS,ATR(6)=1;                     MARK ENTITY AS A MESSENGER
         ASS,AFLAG=0;                     SET OFF TOTM CHECKING
SECQ    QUE(2);                           SECONDARY ON-RAMP QUEUE
         ACT(1)/2;                        ZERO SERVICE TIME
PRMQ    QUE(1),2,49,BLOCK;                PRIMARY ON-RAMP QUEUE
         ACT(1)/1,USERF(9);
         ASS,ATR(9)=TNOW-ATR(1);          COMPUTE QUEUE DELAY
        GOON,1;                           ROUTE TO APPROPRIATE TIME SLICE
         ACT,,SLICID.EQ.1, EX01;          TIME SLICE 1
         ACT,,SLICID.EQ.2, EX02;                     2
         ACT,,SLICID.EQ.3, EX03;                     3
         ACT,,SLICID.EQ.4, EX04;                     4
         ACT,,SLICID.EQ.5, EX05;                     5
         ACT,,SLICID.EQ.6, EX06;                     6
         ACT,,SLICID.EQ.7, EX07;                     7
         ACT,,SLICID.EQ.8, EX08;                     8
         ACT,,SLICID.EQ.9, EX09;                     9
         ACT,,SLICID.EQ.10,EX10;                    10
         ACT,,SLICID.EQ.11,EX11;                    11
;
;------------------------END OF ONRAMP SUBSYSTEM-----------------------
```

```
;--------------COLLECT INFORMATION AND TERMINATE ENTITIES--------------
;
EX01   COL,HEADWY,SLICE  1 TBC; COLLECT INFO ON INTER-ARRIVALS
       COL,NNQ(1),QLP;                    INFO ON PRIMARY QUEUE LENGTHS
       COL,NNQ(2),QLS;                    INFO ON SECONDARY QUEUE LENGTHS
       COL,INT(1),TIME IN SYS;            INFO ON TOTAL TIME IN SYSTEM
       GOON,1;                            INFO FROM MESSENGER ENTITIES
        ACT,,ATR(6).EQ.1,QDND;            INFO ON TOTM QUEUE DELAYS
        ACT,,ATR(6).EQ.0,DONE;            TERMINATE IF HEADWY <> 1
EX02   COL,HEADWY,SLICE  2 TBC;           Repeat for all time slices
       COL,NNQ(1),QLP;                    .
       COL,NNQ(2),QLS;                    .
       COL,INT(1),TIME IN SYS;            .
       GOON,1;
        ACT,,ATR(6).EQ.1,QDND;
        ACT,,ATR(6).EQ.0,DONE;
EX03   COL,HEADWY,SLICE  3 TBC;
       COL,NNQ(1),QLP;
       COL,NNQ(2),QLS;
       COL,INT(1),TIME IN SYS;
       GOON,1;
        ACT,,ATR(6).EQ.1,QDND;
        ACT,,ATR(6).EQ.0,DONE;
EX04   COL,HEADWY,SLICE  4 TBC;
       COL,NNQ(1),QLP;
       COL,NNQ(2),QLS;
       COL,INT(1),TIME IN SYS;
       GOON,1;
        ACT,,ATR(6).EQ.1,QDND;
        ACT,,ATR(6).EQ.0,DONE;
EX05   COL,HEADWY,SLICE  5 TBC;
       COL,NNQ(1),QLP;
       COL,NNQ(2),QLS;
       COL,INT(1),TIME IN SYS;
       GOON,1;
        ACT,,ATR(6).EQ.1,QDND;
        ACT,,ATR(6).EQ.0,DONE;
EX06   COL,HEADWY,SLICE  6 TBC;
       COL,NNQ(1),QLP;
       COL,NNQ(2),QLS;
       COL,INT(1),TIME IN SYS;
       GOON,1;
        ACT,,ATR(6).EQ.1,QDND;
        ACT,,ATR(6).EQ.0,DONE;
EX07   COL,HEADWY,SLICE  7 TBC;
       COL,NNQ(1),QLP;
       COL,NNQ(2),QLS;
       COL,INT(1),TIME IN SYS;
       GOON,1;
        ACT,,ATR(6).EQ.1,QDND;
        ACT,,ATR(6).EQ.0,DONE;
```

```
EX08    COL,HEADWY,SLICE  8 TBC;
        COL,NNQ(1),QLP;
        COL,NNQ(2),QLS;
        COL,INT(1),TIME IN SYS;
        GOON,1;
         ACT,,ATR(6).EQ.1,QDND;
         ACT,,ATR(6).EQ.0,DONE;
EX09    COL,HEADWY,SLICE  9 TBC;
        COL,NNQ(1),QLP;
        COL,NNQ(2),QLS;
        COL,INT(1),TIME IN SYS;
        GOON,1;
         ACT,,ATR(6).EQ.1,QDND;
         ACT,,ATR(6).EQ.0,DONE;
EX10    COL,HEADWY,SLICE 10 TBC;
        COL,NNQ(1),QLP;
        COL,NNQ(2),QLS;
        COL,INT(1),TIME IN SYS;
        GOON,1;
         ACT,,ATR(6).EQ.1,QDND;
         ACT,,ATR(6).EQ.0,DONE;
EX11    COL,HEADWY,SLICE 11 TBC;
        COL,NNQ(1),QLP;
        COL,NNQ(2),QLS;
        COL,INT(1),TIME IN SYS;
        GOON,1;
         ACT,,ATR(6).EQ.1,QDND;
         ACT,,ATR(6).EQ.0,DONE;
QDND    ASS,XCALL=USERF(10);
DONE    TER;                                DEPART THE CAR
;
;------END OF COLLECTION OF INFORMATION & TERMINATION OF ENTITIES-------
```

```
;------------------SUBSYSTEM TO PRINT HEADWAY ARRIVALS-----------------
;
      CRE,7980,7980,,1;                  PRINT HEADWAY ARRIVALS
 QQQQ ASS,XCALL=USERF(4);
      TER;
;
;-----------------SUBSYSTEM TO REINITIALIZE TIME SLICE-----------------
;
      CRE,0,.05,,1;                      GENERATE 1 ENTITY
       ASS,SLICID=0;                     RESET SLICE ID
      TER;
;
;--------------------SUBSYSTEM TO ADVANCE TIME SLICE--------------------
;
      CRE,SLICSZ,0.1,,11;                CREATE 11 TIME SLICES/720 SEC
                                          EACH
       ASS,SLICID=SLICID+1;              ADVANCE BY 1 TIME SLICE
       ASS,XCALL=USERF(1);               ASSIGN MEANS & STD DEV'S
       ASS,TSCUM=TSCUM+SLICSZ;           ADVANCE BY 12 MINUTES (720 SEC)
      TER;
;------------------------SUBSYSTEM TO ADVANCE TOTM----------------------
;
      CRE,60,0.1,,133;                   GENERATE 133 MINUTE INTERVALS
       ASS,TOTM=TOTM+USERF(2);           ADVANCE ONE MINUTE AT A TIME
      TER;
     END;
;
;-----------------------------END OF NETWORK-----------------------------
;
INI,0,7980;                              START AT TIME 0 / END AT 7980
                                                  (6:30 - 8:42)
FIN;                                     FINISH SIMULATION
;
;---------------------------END OF SIMULATION----------------------------
```

A.2
FORTRAN INSERT: USER FUNCTION "USERF"

```
C-----------------------------------------------------------------------
      FUNCTION USERF(IFN)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      COMMON/OURS/ ARRAY(100)
      COMMON/SCOM1/ATRIB(100),DD(100),DDL(100),DTNOW,II,MFA,
     >             MSTOP,NCLNR,NCRDR,NPRNT,NNRUN,NNSET,NTAPE,
     >             SS(100),SSL(100),TNEXT,TNOW,XX(100)
      EQUIVALENCE (XX(1),AVG),(XX(2),ST),(XX(3),SD),(XX(4),TSN),
     >    (XX(5),STL),(XX(6),STH),(XX(7),ESIZE),(XX(8),TRATE),
     >    (XX(9),PSIZE),(XX(10),TLSTEP),(XX(11),XLOW),(XX(12),XMID),
     >    (XX(13),XHIGH),(XX(14),XTOP),(XX(15),XMIN),
     >    (XX(16),XMOD),(XX(17),XMAX),(XX(22),PLONG),(XX(23),PRATE),
     >    (XX(25),VEHCUM),(XX(26),XLMT),(XX(27),ECOUNT),
     >    (XX(28),TLNEXT),(XX(29),PCOUNT),(XX(31),TOTM),
     >    (XX(32),SLICID),(XX(33),SLICSZ),(XX(34),TSCUM),
     >    (XX(35),TOTMQ1),(XX(36),TOTMQ2),(XX(30),CONHDY),
     >    (XX(87),TOTMNW),(XX(88),AFLAG)
C
      GOTO(1,2,3,4,5,6,7,8,9,10),IFN
C--------1---------2---------3---------4---------5---------6---------7--
 1    IF    (SLICID.EQ.1)THEN
C  SET MEAN FOR EXPONENTIAL ARRIVALS
       AVG=12.86
C  READ MEAN OF RAMP METER SERVICE TIME (FROM DATA)
       ST=13.0
C  READ STANDARD DEVIATION OF RAMP METER SERVICE TIME (FROM DATA)
       SD=0.00
C  SET TOTAL NUMBER OF ENTITIES/SLICE
       TSN=28.
C  SET EXPONENTIAL PARCEL SIZE
       PLONG=5.
C  SET PLATOON SIZE
       PRATE=.4
C  SET PLATOON LOW, MODE, AND MAX
       XMIN=-0.5
       XMOD= 2.0
       XMAX= 4.5
C  SET TAIL RATE (PERCENT OF TAIL/EXPONENTIAL)
       TRATE=0.3
C  SET TAIL END
        XHIGH=81.0
```

```
ELSEIF(SLICID.EQ.2)THEN
 AVG=12.27
 ST=13.33
 SD=0.49
 TSN=55.
 PLONG=5.
 PRATE=.35
 XMIN= 0.0
 XMOD= 2.0
 XMAX= 5.0
 TRATE=0.4
 XHIGH=76.0
ELSEIF(SLICID.EQ.3)THEN
 AVG=12.82
 ST=13.42
 SD=0.52
 TSN=61.
 PLONG=5.
 PRATE=.45
 XMIN=-1.0
 XMOD= 1.0
 XMAX= 4.0
 TRATE=0.3
 XHIGH=70.0
ELSEIF(SLICID.EQ.4)THEN
 AVG=10.5
 ST=7.25
 SD=0.45
 TSN=58.0
 PLONG=5.
 PRATE=.45
 XMIN=-0.5
 XMOD= 2.0
 XMAX= 4.5
 TRATE=0.3
 XHIGH=68.0
ELSEIF(SLICID.EQ.5)THEN
 AVG=13.81
 ST=13.42
 SD=0.52
 TSN=44.
 PLONG=5.
 PRATE=.4
 XMIN=-0.5
 XMOD= 2.0
 XMAX= 5.0
 TRATE=0.3
 XHIGH=65.0
```

```
ELSEIF(SLICID.EQ.6)THEN
 AVG=12.50
 ST=12.08
 SD=0.29
 TSN=62.
 PLONG=5.
 PRATE=.5
 XMIN=-0.5
 XMOD= 2.0
 XMAX= 3.5
 TRATE=0.3
 XHIGH=66.0
ELSEIF(SLICID.EQ.7)THEN
 AVG=13.5
 ST=12.0
 SD=0.25
 TSN=84.
 PLONG=5
 PRATE=.42
 XMIN=-0.5
 XMOD= 2.0
 XMAX= 4.5
 TRATE=0.2
 XHIGH=57.0
ELSEIF(SLICID.EQ.8)THEN
 AVG=15.0
 ST=11.58
 SD=0.90
 TSN=70.
 PLONG=5.
 PRATE=.5
 XMIN=-0.25
 XMOD= 3.0
 XMAX= 3.5
 TRATE=0.35
 XHIGH=106.0
ELSEIF(SLICID.EQ.9)THEN
 AVG=15.04
 ST=10.42
 SD=0.67
 TSN=45.
 PLONG=5.
 PRATE=.5
 XMIN=-0.5
 XMOD= 1.0
 XMAX= 3.5
 TRATE=0.4
 XHIGH=100.0
```

```
      ELSEIF(SLICID.EQ.10)THEN
       AVG=8.6
       ST=9.92
       SD=1.31
       TSN=78.
       PLONG=5.
       PRATE=.6
       XMIN=-1.0
       XMOD= 1.0
       XMAX= 4.65
       TRATE=0.3
       XHIGH=60.0
      ELSEIF(SLICID.EQ.11)THEN
       AVG=9.23
       ST=0.0
       SD=0.0
       TSN=22.
       PLONG=5.
       PRATE=.45
       XMIN=-0.5
       XMOD= 1.0
       XMAX= 3.5
       TRATE=0.2
       XHIGH=42.0
      ELSEIF(SLICID.EQ.12)THEN
C THIS IS FOR THE ENTIRE PERIOD 6:30-8:42
       AVG=8.15
       ST=12.25
       SD=1.29
       TSN=616.
       PLONG=5.
       PRATE=.48
       XMIN=-0.5
       XMOD= 2.0
       XMAX= 4.5
       TRATE=0.4
       XHIGH=106.0
      ENDIF
C THIS SECTION IS TO OVERRIDE THE ABOVE ASSIGNMENTS IF NECESSARY
C TEMPORARY
       AVG=6.5
C      ST=12.25
C      SD=1.29
C      TSN=616
       PLONG=5
       TRATE=.2
       XMIN=-0.5
       XMOD= 2.0
       XMAX= 4.5
       PSIZE=NINT( UNFRM( (PLONG-1),(PLONG+1),4))
       ESIZE=NINT( PSIZE/PRATE-PSIZE )
```

```
C  COMPUTE TAIL STEP (POINTER TO NEXT TAIL OCCURRENCE)
        TLSTEP=1/( (TRATE*ESIZE)/(ESIZE+PSIZE) )
C  SET LOWER VALUE FOR METER SERVICE TIME (FOR THE MODEL)
        STL=ST-SD
C  SET UPPER VALUE FOR METER SERVICE TIME (FOR THE MODEL)
        STH=ST+SD
C
C  WRITE INTERMEDIARY RESULTS IF NEEDED
        WRITE(8,98)SLICID,CONHDY,VEHCUM,ESIZE,AVG,PSIZE,TLSTEP,TLNEXT
  98    FORMAT('SL=',F4.1,' TC=',F6.1,' VC=',F5.1,' E=',F5.1,
     >         ' AVG=',F5.2,' P=',F5.1,' STP=',F6.2,' NXT=',F6.2)
C
        RETURN
C
C-------------------------------------------------------------------------
C
C SET TIME FOR TOTM, TURN FLAGING OPTION ON (TO FLAG A MESSENGER ENTITY)
C & PRINT TOP-OF-THE-MINUTE STATUS FOR SLICID, AVG, & QUEUE LENGTHS
 2      USERF=1.0
        TOTMNW=TNOW
        TOTMQ1=FLOAT(NNQ(1))
        TOTMQ2=FLOAT(NNQ(2))
        AFLAG=1
        IF(TOTM.EQ.0) WRITE(8,100)
        IF(TOTM.NE.0) THEN
          WRITE(8,200)SLICID,TOTM,AVG,NNQ(1),NNQ(2)
        END IF
 100    FORMAT(//22X,'MEAN TIME BET'/1X,'SLICE   ',3X,'  TOTM  ',4X,
     >         'ARRIVALS',4X,'CARS IN',4X,'CARS IN'/1X,'NUMBER',
     >        5X,'(MINUTES)',3X,'(SECONDS)',3X,'PRIM QUE',3X,
     >         'SECND QUE'/57('-'))
 200    FORMAT(3(F6.2,6X),2(I4,6X))
        RETURN
C
C-------------------------------------------------------------------------
C
C THIS SECTION IS TO ROUND HEADWAYS UP/DOWN TO NEAREST WHOLE NUMBER
C  & UPDATE ARRAY.
 3      USERF=1.0
        IF(ATRIB(5).LE.30.)THEN
          IL=NINT(ATRIB(5))
          ATRIB(5)=FLOAT(IL)
          ARRAY(IL)=ARRAY(IL)+1
        ENDIF
        RETURN
C
C-------------------------------------------------------------------------
```

```
C THIS SECTION WILL WRITE OUT "ARRAY" TO CHECK HEADWAY DISTRIBUTION
 4      DO 405 I=0,30
          WRITE(7,410)I,ARRAY(I)
 405    CONTINUE
 410    FORMAT(2X,I4,2X,F6.2)
        RETURN
C
C-------------------------------------------------------------------------
C
C GENERATE TRIANGULAR TAIL ARRIVALS
 5      USERF=TRIAG(XLOW,XMID,XHIGH,1)
C       ECOUNT=ECOUNT+1
        TLNEXT=TLNEXT+TLSTEP
C       WRITE(8,505)VEHCUM,ECOUNT,USERF,TLSTEP,TLNEXT
505     FORMAT('VEHCUM=',F5.1,' ECOUNT=',F5.1,' TAIL=',F5.1,
     >         ' TLSTEP',F5.1,' TLNEXT=',F5.1)
        RETURN
C
C-------------------------------------------------------------------------
C
C GENERATE EXPONENTIAL ARRIVALS
 6      USERF=EXPON(AVG,1)
        IF(USERF.LT.XLMT)GOTO 6
C       WRITE(8,605)VEHCUM,ECOUNT,USERF
605     FORMAT('VEHCUM=',F5.1,' ECOUNT=',F5.1,' EXPONL=',F5.1)
C       ECOUNT=ECOUNT+1
        RETURN
C
C-------------------------------------------------------------------------
C
C GENERATE PLATOONED ARRIVALS
 7      USERF=TRIAG(XMIN,XMOD,XMAX,2)
C       WRITE(8,705)VEHCUM,PCOUNT,USERF
705     FORMAT('VEHCUM=',F5.1,' PAOUNT=',F5.1,' PLATON=',F5.1)
        IF(USERF.LT..45)USERF=1.0
C       PCOUNT=PCOUNT+1
        RETURN
C
C-------------------------------------------------------------------------
C
C SHIFT HEADWAYS TOWARD REAL-WORLD VALUES
 8      IF(ATRIB(5).LT.XMOD)THEN
          ATRIB(5)=ATRIB(5)+.667
C         WRITE(7,*)ATRIB(5)
        ELSEIF(ATRIB(5).LT.XLMT)THEN
          ATRIB(5)=ATRIB(5)+.333
        ELSE
          ATRIB(5)=ATRIB(5)+.5
        END IF
        RETURN
C
```

```
C-------------------------------------------------------------------------
C
C SET SERVICE TIME FOR RAMP METER
 9      USERF=UNFRM(STL,STH,3)
C       IF(ST.GT.0)USERF=USERF-USERF/4.
C       WRITE(7,*)USERF
        RETURN
C
C-------------------------------------------------------------------------
C
C CONVERT CUMULATIVE TIME TO HH:MM FORMAT & PRINT MESSENGER ENTITY
C INFORMATION ON ITS TIME SLICE, TOTM, NNQ(1), NNQ(2), AND TIME IN SYS
10      DUMMY=23400.+ATRIB(1)
        HOURS=DUMMY/3600.
        IHOUR=AINT(HOURS)
        IMINT=AINT((HOURS-IHOUR)*60)
        HOURS=FLOAT(IHOUR)+IMINT/100.
        IF(TOTM.EQ.1)WRITE(7,1005)
        WRITE(7,1010)ATRIB(3),ATRIB(4),HOURS,ATRIB(7),ATRIB(8),
     >              ATRIB(9)
1005    FORMAT(//1X,'TIME',9X,'CHECK'/1X,'SLICE',2X,'TOTM',3X,'TIME',4X,
     >              'PQL',3X,'SQL',5X,'RAMP DELAYS'/56('-'))
1010    FORMAT(2(F5.1,2X),F5.2,3(F6.1,2X))
        RETURN
C
C-------------------------------------------------------------------------
C
        END
```