TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA/TX-94/1232-23	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ADVANCED FREEWAY SYSTEM RAMP METERING STRATEGIES FOR TEXAS		5. Report Date October 1993
		6. Performing Organization Code
7. Author(s) Carroll J. Messer		8. Performing Organization Report No. Research Report 1232-23
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135		10. Work Unit No.
		11. Contract or Grant No. Study No. 0-1232
12. Sponsoring Agency Name and Address Texas Department of Transportation Office of Research and Technology Transfer P.O. Box 5051 Austin, Texas 78763		13. Type of Report and Period Covered Interim: September 1, 1991 - October 31, 1993
		14. Sponsoring Agency Code

15. Supplementary Notes

Research Performed in Cooperation with Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.
Research Study Title: Urban Highway Operations Research and Implementation Program

16. Abstract

Freeway ramp metering systems have been used to improve urban freeway flow. However, control strategies must be properly adjusted to account for ramp queues overflowing onto surface streets and provide equitable on-ramp control during various operating periods. An improved solution can be obtained by optimizing this problem simultaneously for a group of time slices. This study identifies and examines a microcomputer-based optimization scheme that can assist in developing efficient freeway control strategies for on-line freeway surveillance and control.

A multi-level freeway control structure is employed for which ramp metering control algorithms are developed for each level of control. Flow-based and lane occupancy-based system algorithms are presented. Detailed data file requirements are provided for each control level. A micro-computer prototype, or laboratory test version of the system level, will be described in a companion project report.

17. Key Words Freeway Corridor, Traffic Management, Ramp Metering, Algorithm Development, Operational Strategy	18. Distribution Statement No Restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 104	22. Price

ADVANCED FREEWAY SYSTEM RAMP METERING STRATEGIES FOR TEXAS

by

Carroll J. Messer, Ph.D., P.E.

Research Engineer
and
Professor of Civil Engineering

Texas Transportation Institute
The Texas A&M University System
College Station, Texas 77843-3135

Research Report Number 1232-23

Research Study Number 0-1232

Sponsored by the
Texas Department of Transportation
In Cooperation with
U.S. Department of Transportation
Federal Highway Administration

October 1993

IMPLEMENTATION STATEMENT

The results of this research provide several items that TxDOT's Freeway Traffic Management section of D-18TM may use for immediate implementation in ramp metering projects around the state. An overall hierarchical design concept for ramp metering is provided in which most data structures, control variables and operational algorithms are identified. A complete mathematical formulation of the systems ramp metering problem is specified.

Support software is being developed in a subsequent research task. A less complex multilevel ramp metering system based on the previous hierarchical design concept which uses lane occupancy rather than freeway volumes is also provided. Additional traffic detectors and some modifications to existing traffic detector designs would be needed to obtain the full benefits of this simple but elegant design concept. Some modification to the current TxDOT ramp meter specification would be needed to take advantage of the full features and capabilities of these proposed ramp metering strategies.

DISCLAIMER

The contents of this report reflect the views of the author who is responsible for the opinions, finding, and conclusions presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration, or the Texas Department of Transportation. This report does not constitute a standard, specification, or permit regulation. Additionally, this report is not intended for construction, bidding, or permit purposes. Carroll J. Messer, P.E. #31409, was the engineer in charge of this work.

ACKNOWLEDGMENT

The research reported herein was performed as a part of Study 0-1232, entitled "Urban Highway Operations Research and Implementation Program" conducted by the Texas Transportation Institute (TTI) and sponsored by the Texas Department of Transportation (TxDOT) in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

Many people contributed significantly to this work. Among these include Dr. Thomas Urbanik II, who served as the overall project 1232 principal investigator for Texas Transportation Institute. Dr. Edmond C.P. Chang provided computer automation management services, and Dr. Nadeem A. Chaudhary provided critical operations research technology to the linear programming model. At TxDOT, Mr. Gary Trietsch (D-18) served as technical coordinator, and Ms. Carol Rawson of TxDOT (D-18TM) served as technical advisor on this phase of the research effort.

Thanks are also extended to the numerous other researchers and support staff of the Texas Transportation Institute who worked on this project. Several deserve special acknowledgment, including Dr. Youngchan Kim, Dr. Eumni Park, Ali Haider Husaini, Su-Hwa Wong, and Dayakar Prabhakar. In addition, special thanks are due to Ms. Lisa Badillo, Elizabeth Escamilla, and Mr. George Wier for their collective word processing skills. A special note of gratitude is also due to Mr. George Wier for his splendid graphical assistance.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF FIGURES

LIST OF TABLES

SUMMARY

A comprehensive research program is underway in Texas to develop advanced methods to improve traffic operations of which this research is a small part. This research report describes advanced ramp metering strategies that have been developed specifically for Texas urban freeways using Texas Department of Transportation's (TxDOT) specification for ramp meters together with Division 18's (D-18TM) Freeway Traffic Management system design concept. Freeway ramp metering systems have been used in Texas as early as 1965 to improve urban freeway flow. Such control strategies were operated on an isolated basis and used limited computational capabilities. However, control strategies should be properly adjusted to account for ramp queue overflow onto surface streets and provide equitable on-ramp control during various operating periods. An improved solution can be obtained by optimizing this problem simultaneously for a group of time slices. This study identifies and examines a microcomputer-based optimization scheme that will assist engineers in developing freeway control strategies that will enhance the on-line freeway surveillance and control system being developed by D-18TM.

A multi-level freeway control structure is employed for which ramp metering control algorithms are developed for each level of control. Flow-based and lane occupancy-based system algorithms are presented to assist in the practical implementation of these systems. Detailed data file requirements are provided for each control level. Systems-based algorithms using linear programming can develop optimal ramp metering rates and directly estimate the operational performance for the freeway and for all entrance ramps. A micro-computer prototype of the system level will be described in a subsequent project report.

A simplified systems-based ramp metering strategy is also recommended that uses most of the advanced traffic flow theory embedded in the more complex system, but is based on lane occupancy measures as used by the current TxDOT ramp meter specification. This system may be utilized as an interim measure until the advanced Freeway Traffic Management System being developed by D-18TM is installed and becomes operational.

1.0 INTRODUCTION TO RAMP METERING

THE URBAN CONGESTION PROBLEM

A recently completed study of the traffic growth in the six largest cities in Texas found that traffic demand on Texas urban freeways is continuing to increase at about 3.2 percent per year (*1*). Even with all the new freeway construction recently completed in Texas, urban traffic congestion persists and will soon return to severe levels during peak period conditions as the economy continues to grow (*2*). The former study (*1*) also showed that the duration of the congestion periods (rush hours) is continuing to lengthen at about the same growth rate as the traffic demand.

On congested urban freeways, average operating speeds are frequently less than 30 mph in stop-and-go traffic, resulting in excessive fuel consumption and emissions into the atmosphere, endangering our general health and public safety. Traffic engineers have identified two types of freeway congestion: recurrent and non-recurrent congestion. The occurrence of either type of congestion arises when traffic demand exceeds capacity.

Freeway Traffic Management (FTM) systems have been proposed and implemented in some cities around the country as one traffic management strategy for reducing traffic congestion without adding extra physical capacity to the existing facility. Adding physical capacity to an existing urban freeway is not only extremely costly but may have serious negative side effects on the local social and economic environment. As the following literature will show, well designed and maintained ramp metering systems have proven to be an effective FTM tactical component in reducing freeway congestion in large urban areas. Subsequent chapters will present current ramp metering systems and proposed new systems, whose features provide both more advanced control theory and graphical user interface (GUI) design features than do present systems.

Causes of Recurrent Congestion

Mainline Freeway. Routine recurrent freeway congestion, as described above, is caused by persistently growing (say at 3% per year) and unregulated freeway traffic demand exceeding the physical geometric capacity of the freeway mainlines along the roadway, usually at an entrance ramp merge point or downstream from it at the next geometric bottleneck. Variability in the traffic demands, either daily, hourly or minutely, can overload the freeway system to varying degrees. Most of this freeway congestion is recurrent, frequently starting and ending at generally predictable times (like during the morning and afternoon peak hours). If the mainline freeway demand exceeds the downstream capacity, as shown in Figure 1, then congestion will immediately form, last longer than the overload period, and end only after the storage queue has been dissipated. When the unregulated mainline freeway demand already exceeds downstream capacity, simple ramp metering alone cannot eliminate freeway congestion, and more advanced strategies will be needed.

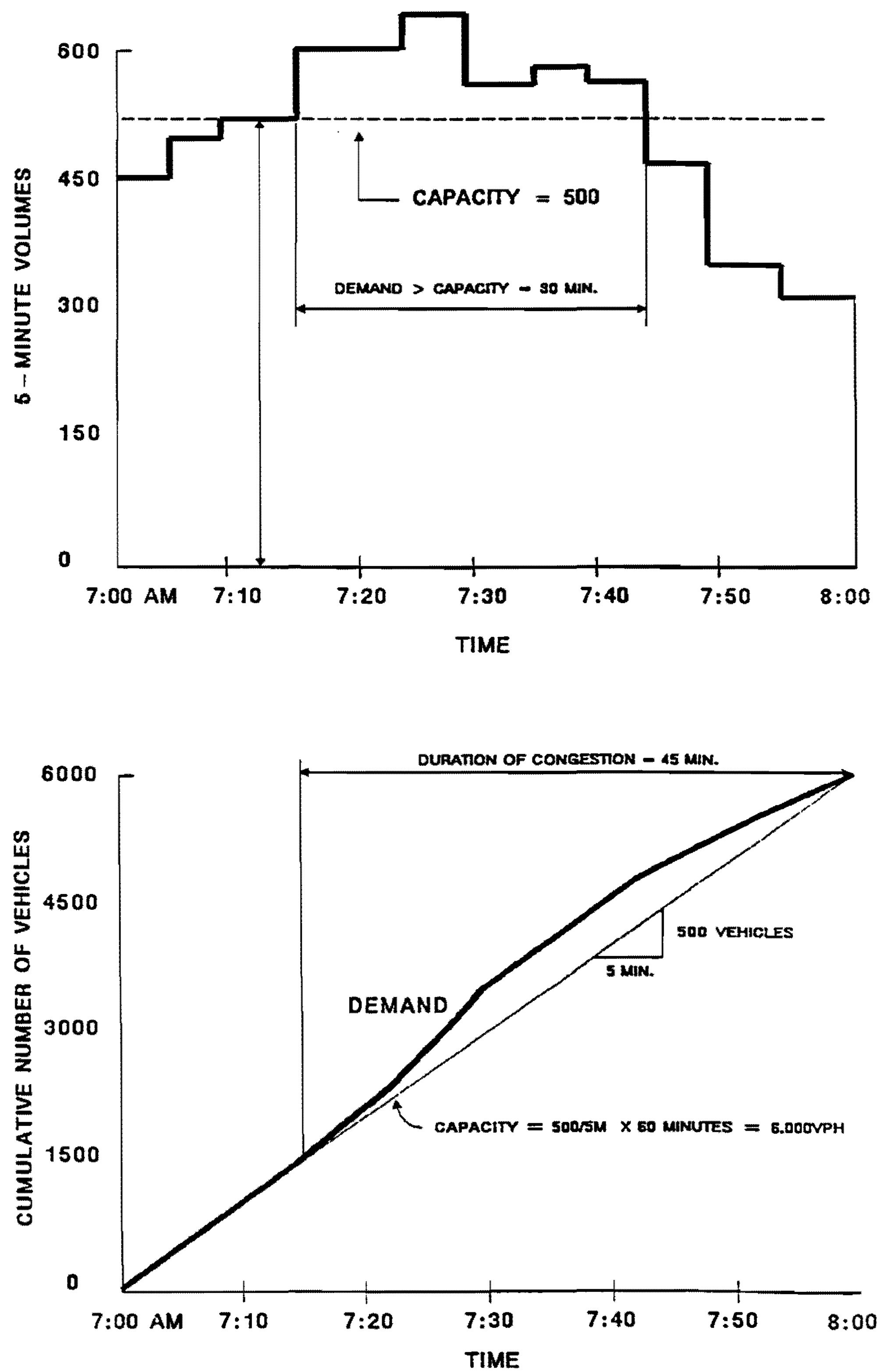


Figure 1. Congestion Generated by Freeway Mainline Demand Exceeding Capacity.

Entrance Ramps. The freeway traffic demand in Figure 1, when measured just downstream of an entrance ramp, is the sum of the upstream freeway demand and the on-ramp demand, as shown in Figure 2. While the mainline freeway demand is typically many times larger than the ramp demand and alone may exceed downstream capacity, both flows can be important contributors to the onset of freeway congestion. The ramp demand takes on a higher priority when congestion will only form if the added unregulated ramp demand causes the downstream freeway demand to exceed capacity, even for a short period of time.

Several time-based characteristics of unregulated entrance ramp demand are known to exist and contribute to demand overload with subsequent congestion. The most common effect is related to the "average" demand level at the entrance ramp during the time slice (interval) of the study. As shown in Figure 2, average ramp volumes may vary between the time slices and these average volumes may, when added to the approaching mainline flow, exceed the downstream merge capacity or mainline bottleneck capacity of the freeway. When the total freeway demand exceeds mainline capacity for an extended period of time (say five minutes), one can be assured that freeway congestion will form with all of its subsequent operational problems and potential future detrimental impacts (pollution, accidents, further loss in roadway capacity, etc.). The undesirable outcome of long-term traffic demand exceeding capacity is a major operational problem for traffic management.

The second type of unregulated ramp demand flow pattern that can routinely cause congestion problems is short-term variability in flow. We begin by recognizing that rural traffic flow is intrinsically random, or Poisson, with a variance in flow rate equal to the average flow. Urban freeway mainline flow also tends to be random, but possibly becoming somewhat constrained (less random) at higher volumes approaching capacity (*3*). However, traffic studies of the output from signalized urban diamond interchanges feeding entrance ramp traffic to the freeway show that these traffic flow rates are highly variable, can even be more variable than random for half-minute intervals (*4*), and routinely produce pulsed flow rates within the platoon flow to the entrance ramps that are 2-3 times the average ramp volume for a 1/2 minute or more. As urban freeway volumes continue to grow in Texas and again reach near capacity levels, these platoons of oncoming ramp traffic may no longer be able to find an acceptable gap and safely merge onto the freeway without breaking down the freeway flow. The basic problem is that the short-term ramp demand exceeds mainline freeway capacity. Once the mainline traffic slows or stops to allow ramp vehicles to enter the freeway under high density conditions, shockwaves form in the freeway flow and move upstream which can lead to rear-end accidents and major traffic congestion.

Another more recent and disturbing outcome of entrance ramp merging overload, with its resulting mainline congestion, is the loss of freeway right-of-way priority. Once freeway speeds drop below ramp speeds at the merge, i.e., approach zero during stop-and-go traffic, freeway mainline vehicles typically begin to yield the right-of-way to ramp vehicles, sometimes in an alternating manner, as long as both streams are queued at the merge point (*5*). The upshot of this loss of freeway priority is that the freeway loses its capacity to serve mainline freeway traffic, even though the downstream volumes may only be slightly less (*6*).

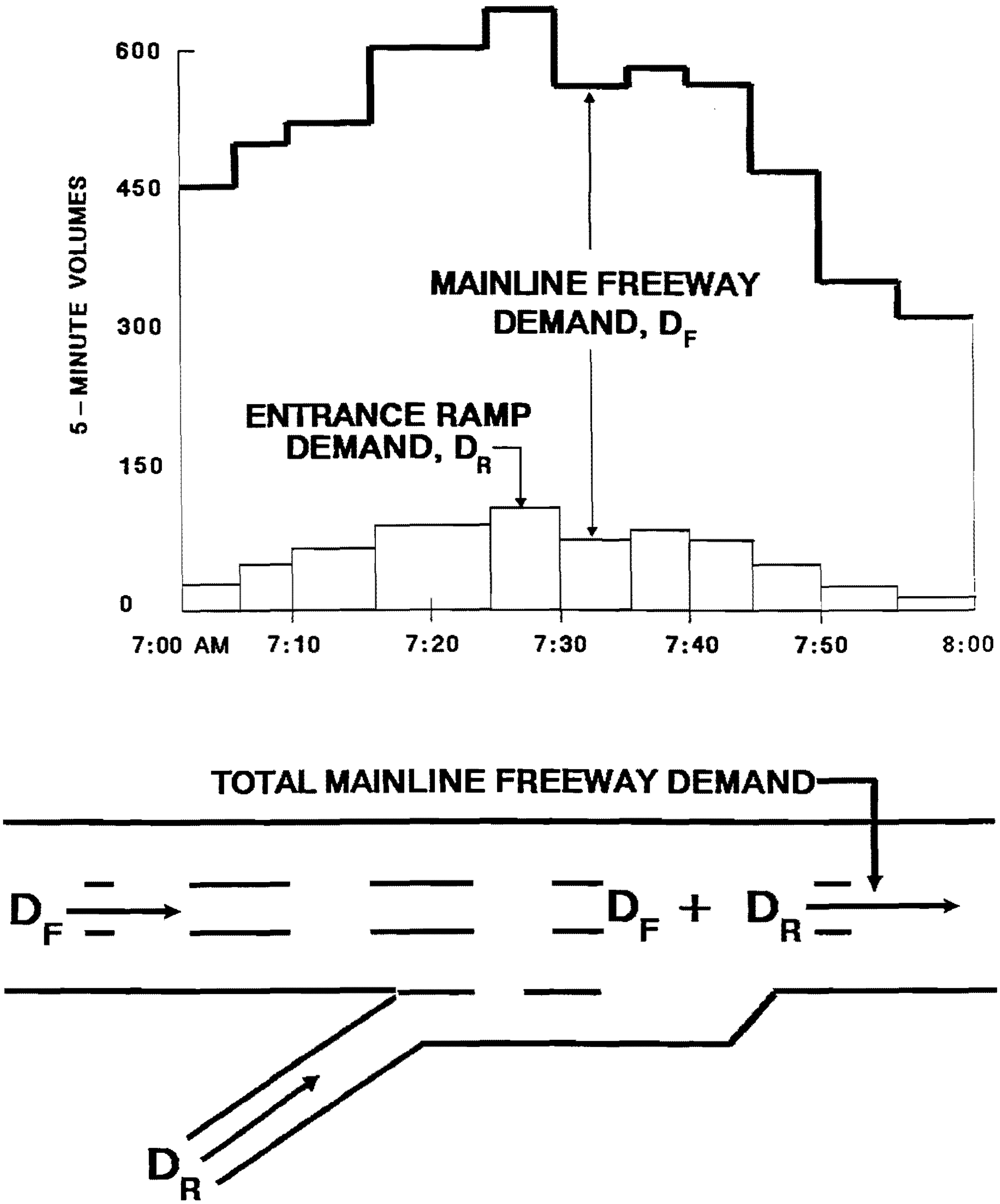


Figure 2. Total Freeway Mainline Demand Downstream of Entrance Ramp.

The operational functionality of the freeway is diminished as a consequence of unregulated ramp operations during high volume conditions. In essence, once congestion begins on the freeway, ramp vehicles operationally have the same priority as freeway vehicles. Each additional ramp entry further reduces freeway throughput and related mainline capacity, thereby delaying numerous other vehicles already using the freeway.

The above examples illustrate that recurrent freeway congestion caused by unregulated entrance ramp demand feeds on itself, significantly reducing freeway capacity and its functionality. Moreover, recurrent congestion can cause nonrecurrent congestion to form and last much longer than normal if traffic accidents occur at unexpected shockwaves or within turbulent stop-and-go traffic. Furthermore, it is known that traffic accident rates rise as the density of flow increases, as in congested traffic conditions. A minor rear-end accident, blocking one freeway mainlane for only a few minutes, may cause traffic congestion lasting for an hour or more (*7*).

Exit Ramps. Normally, freeway metering systems try to maximize the throughput of the system essentially by maximizing the output flow from the freeway. Freeway traffic management strategies should always try to keep the exit ramps clear of mainline traffic congestion. Queueing on the freeway mainlanes simply blocks possible output flow to the exit ramps. Once the exit ramps become restricted, strategies that meter, close, or otherwise limit the flow to exit ramps may be considered, particularly during major incident conditions or special events. One case may be where severe congestion is arising at a downstream diamond interchange, perhaps caused by an accident blocking the approach lanes. Another case may be where major trip diversion would occur to another freeway having adequate capacity if the exit ramp were closed, thereby reducing the demand to the congested section.

Causes of Non-recurrent Congestion

Freeway capacity is not a fixed value, as noted above. Accidents caused by poor traffic operations, unexpected shockwaves, and driver errors in general, greatly reduce the average throughput capacity of the freeway mainlanes. Strategies that reduce incidents should be implemented where possible. Table 1 (*8*) shows that a 1-lane blockage of the freeway mainlanes by a minor accident (or stalled vehicle) reduces capacity to about 49% of its normal value, even though the physical reduction is only 33%, because of the "rubber necking" phenomenon. An accident that blocks 2 lanes reduces flow to about 21% of normal (compared to a 67% reduction in freeway lanes). Thus, freeway incidents create a major reduction in capacity that is larger than the physical reduction of the facility. Traffic congestion may also indirectly contribute to additional incidents, such as vehicles overheating or running out of gas due to unexpected major delays.

Freeway capacity may change for reasons other than when accidents or stalled vehicle induced incidents have occurred. Research has found that the capacity of the freeway mainlanes may be reduced by 15-20% during light rain showers (*9*). Capacity of a section is also a function of the characteristics of the current traffic stream--its traffic mix, driver population,

Table 1. Summary of Observed Flow Data At Freeway Incident Sites

Condition	Sample Size (no. of min)	Average Flow (veh/min)	Standard Deviation (veh/hour)	Average Flow Rate (veh/hour)	Percent of Normal (%)
Normal flow	312	92.6	6.3	5,560	100
Noninjury accident					
1 lane blocked					
Outside lane	46	48.1	11.7	2,880	52
Center lane	42	42.7	11.4	2,560	46
Median lane	79	46.1	8.8	2,770	50
Combined	167	45.8	10.5	2,750	49
Stalled vehicle					
1 lane blocked	43	47.9	8.6	2,880	52
Accident					
2 lanes blocked	53	19.1	7.3	1,150	21
Accident					
on shoulder	254	67.1	13.1	4,030	72

Source: Reference 8.

and speed variance (*10*). These variations in capacity cannot be predicted with precision, and any unusual variability will only reduce the quality of traffic flow. Operational performance will be degraded unless these "environmental" variations in capacity can be detected and adaptive control strategies implemented in a timely manner.

TRAFFIC MONITORING

Traffic monitoring is the eyes and ears of an advanced freeway traffic management (FTM) system. Real-time monitoring of freeway traffic flow is needed to determine the status of the freeway flow and to provide better estimates of the operational capacity of each freeway section. The need for freeway traffic management systems that provide real-time traffic monitoring and ramp metering systems should be evident to improve operations and safety for large urban areas. FTM systems should provide total monitoring stations, those where all the basic traffic flow variables can be obtained for all lanes, at regular intervals of 0.5 to 1.0 miles (0.8 to 1.6 km). As indicated in Figure 3, design tradeoffs have to be made between the cost (spacing between detectors or the number of detectors per mile) of the traffic monitoring system as compared to its speed for detecting incidents (*11*).

RAMP METERING SYSTEMS

Ramp metering, considered in the context of traffic management systems, offers several operational features for improving freeway flow, traffic safety and air quality by the optimal regulation of input flow to the freeway. Ramp meters are traffic signals placed on freeway entrance ramps that, by judiciously cycling the signal, regulates the ramp flow to the freeway in an objective manner. In the "metering" mode (*12*), ramp meters operate to discharge traffic at a measured rate based on real-time conditions, thereby protecting the delicate demand-capacity balance at the ramp merge or downstream bottleneck. As long as mainline traffic demand does not exceed capacity, throughput is maximized, speeds remain more uniform, and congestion related accidents are reduced.

Ramp meters also regulate the ramp traffic in order to break up platoons of vehicles that have been released from nearby signalized intersections. The mainline, even when traffic flow nears capacity, can usually accommodate merging vehicles one or two at a time. On the other hand, when platoons of vehicles attempt to force their way into the freeway traffic, this action creates turbulence that can cause the mainline flow to break down. Reduced turbulence in the merge zones also leads to reduced sideswipe and rear-end accidents that are associated with unrestricted ramp access during high volume conditions.

History

Ramp metering is not a new traffic management concept. Various forms of ramp control were used experimentally in Detroit in the early 1960's. In Chicago, ramp meters have been

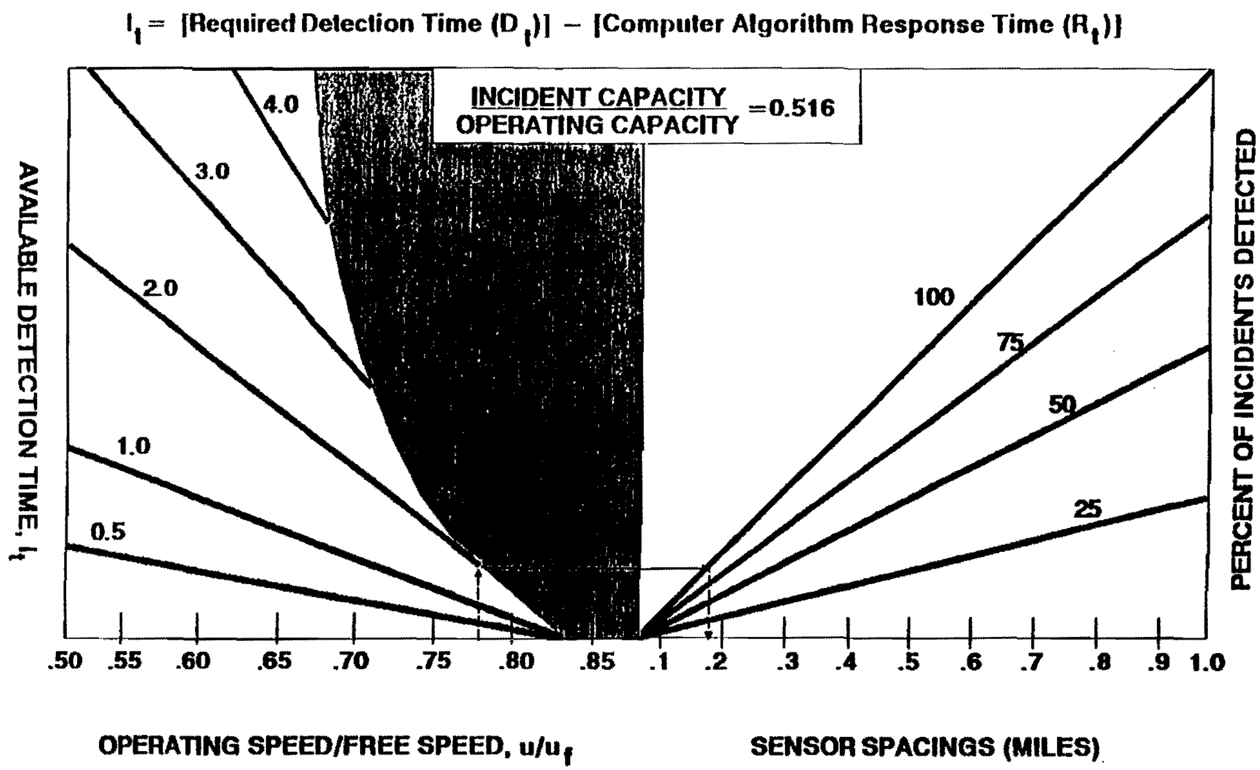


Figure 3. Sensor Spacings for an Incident Duration of Two Minutes or More.

in operation on the Eisenhower Expressway since 1963 (12). Eight ramp meters were installed on the Gulf Freeway in Houston in 1965 (13, 14) and operated successfully until freeway reconstruction caused their removal in 1975. Over 30 ramp meters were operated successfully on the North Central Expressway in Dallas from 1971 until major freeway reconstruction forced most of them to be removed in 1990. In Los Angeles, ramp metering began in 1968. The system has been expanded continually until there are now over 900 meters in operation in metropolitan L.A., making it the largest system in the country. Ramp meters are currently operating in 20 metropolitan areas in North America (15). These metering systems vary from fixed time operation at a single ramp to the responsive control of every ramp along many miles of a freeway. One measure of the effectiveness of ramp metering is the fact that nearly every existing system has been or is proposed to be expanded.

Case Studies

Robinson (15) presents a representative sample of ramp metering applications in several cities and describes the benefits that have been reported. A summary of his findings, updated to 1992 conditions by related TRB reference materials (16), follows.

Austin, Texas. The successful use of ramp meters does not require large systems. In Austin, the Texas Department of Transportation (TxDOT) implemented ramp meters at 3 ramps along a 2.6 mile segment of northbound I-35 for operation during the A.M. peak period. The section of freeway had two bottleneck locations that were reducing the quality of travel. One bottleneck was a reduction from 3 to 2 lanes, and the other was a high volume entrance ramp just downstream of the lane reduction. Metering resulted in an increased vehicle throughput of 7.9% and an increased average peak period mainline speed of 60% through the section. The meters were removed after the reconstruction of I-35 eliminated the lane drop and added sufficient capacity to this freeway section.

Detroit, Michigan. Ramp metering is an important aspect of the Michigan DOT (MDOT) Surveillance, Control, and Driver Information (SCANDI) System in Detroit. The SCANDI metering operation began in November 1982 with six ramps on the eastbound Ford Freeway (I-94). Nineteen more ramps were added on I-94 in January 1984, three in November 1985, and 50 additional ramp metering locations are planned in Detroit.

An evaluation performed for MDOT determined that ramp metering increased speeds on I-94 by about 8%. At the same time, the typical peak hour volume on the three eastbound lanes increased to 6400 vehicles per hour (vph) from an average of 5600 vph before metering. By reducing merging interference, ramp metering has allowed traffic volumes to approach theoretical capacity. In addition, the total number of accidents was down 71%. The evaluation also showed significant additional benefits could be achieved by metering the three freeway-to-freeway connectors on the section of I-94. As noted, freeway traffic management that combines the ramp metering element has been very successful in Detroit.

Minneapolis/St.Paul. The Twin Cities ramp metering system is composed of several systems and sub-systems that have been implemented over a 20-year period by the Minnesota DOT (MnDOT). The first two fixed-time meters were installed in 1970 on a five-mile section of southbound I-35E north of downtown St. Paul. In November 1971, these were upgraded to operate on an isolated traffic responsive basis and 4 additional meters were activated. This section of I-35E has been evaluated periodically since the meters were first installed. The most recent study shows that after 14 years of operation, average peak hour speeds remain 16% higher (from 37 to 43 mph) than before metering while peak period volume increased 25% over the same time period. The average number of peak period accidents decreased 24% and the peak period accident rate decreased 38%.

In 1974, a freeway management project was initiated on a 17-mile section of I-35W from downtown Minneapolis to the southern suburbs. In addition to 39 ramp meters, the system included closed-circuit television, variable message signs and 380 vehicle detectors. Computer control was centralized at modern traffic management center (TMC) located near downtown. This project also included extensive freeway flyer bus service and 11 ramp meter bypass ramps for HOVs. An evaluation of this project after 10 years of service showed that average peak period freeway speeds increased from 34 to 46 mph (35%). Over the same 10-year span, average peak period throughput increased by 32%, the average number of peak period accidents declined 27%, and the peak period accident rate declined 38%.

Additional ramp metering projects were implemented in 1980 and 1985 and more projects are now in the design or construction phase. The MnDOT TMC currently operates 308 ramp meters in the Twin Cities metropolitan area with 120 meters to be added in 1993 (16). The long range plan is to extend the traffic management program over the entire Twin Cities freeway network as traffic conditions warrant. The success of the system has shown that the staged implementation of a comprehensive freeway management system on a segment-by-segment, freeway-by-freeway basis, is possible over a long period of time.

Portland, Oregon. The first ramp meters in the Northwest U.S. were installed along a six-mile section of I-5 in Portland in January 1981. The meters are operated by the Oregon DOT. This freeway is the major north/south link through the metropolitan area and is an important commuter route. The initial system consisted of 16 metered ramps between downtown Portland and the Washington state line. Nine of the ramp meters operated in the northbound direction during the P.M. peak and seven controlled southbound entrances during the A.M. peak. The meters operated in local or isolated pretimed mode.

Prior to ramp metering, it was common along this section of I-5 for platoons of vehicles to merge onto the freeway and aggravate the slowdown of congested traffic. The northbound P.M. peak hour average speed was 16 mph. Fourteen months after installation, the average speed for the same time period was 41 mph. Before metering, conditions in the southbound A.M. peak were much less troublesome and, hence, improvements were smaller. The average speed increased from 40 to 43 mph, resulting in only a slight reduction in southbound travel time. Overall, there was a 43% reduction in peak period traffic accidents along the freeway during the control periods. The success of the initial installation led to expansion of the system.

Ramp meters were extended along the I-5 corridor and also a new group of meters were installed on I-84.

Seattle, Washington. In September 1981, Washington DOT implemented metering on I-5 north of downtown Seattle as part of its FLOW system. By 1989, the system was controlling 17 southbound ramps during the A.M. period and five northbound ramps during the P.M. period. Between 1981 and 1987, mainline volumes increased over 86% northbound and 62% southbound. Before metering, the travel time on a specific 6.9 mile course on I-5 took 22 minutes. After metering was installed, the same course took 11.5 minutes even with the higher volume. Over the same time period, the accident rate decreased by 39%.

Operational experience has shown that the ramp metering system installed should not only adequately solve the operational problem, but it should also provide the operational features and be well maintained so that the user public will support the system over an extended period of time. Simplicity of design may be cheaper in the short run, but the public may not like its simple operation and begin a trend of disrespecting the ramp meters until their credibility and efficiency are lost.

Types of Ramp Metering Systems

The types of ramp metering systems implemented reflect the traffic control needs and technology of the times. Most systems were installed before the advent of rugged microprocessor-based computing, although some Type 179 systems have been installed in California and other states. A brief summary of these systems is presented below. Chapter 2 will address this subject in greater detail. Several references (*17*, *18*) provide further details on the subject. In retrospect, theoretically more advanced ramp metering systems were tested by the Texas Transportation Institute (TTI) in 1969 (*19*).

Local Pretimed Control. The simplest form of ramp metering uses the isolated or local pretimed control mode of operation. Ramp metering volumes are established initially based on highway capacity manual methods of analysis, which are then fine-tuned based on local field observations. Pretimed signal timing plans are established based on observed mainline volumes, merging volumes, and existing ramp volumes. To maintain demand-capacity balance on the freeway, ramp merging volumes may need to be reduced. Traffic detectors (sensors) are used only to drive the on/off status of ramp metering signals and are not used to measure or estimate freeway demands or traffic conditions.

Local Traffic Responsive Control. The next higher level of control establishes traffic responsive metering rates based on measured freeway traffic conditions upstream of the ramp. The local traffic responsive approach utilizes detectors and a micro-processor to determine the mainline flow in the immediate vicinity of the ramp and ramp demand to select an appropriate metering rate. Traffic responsive metering can be expected to produce results that are on the order of 5 to 10 percent better than those of pretimed metering (*15*).

Adaptive Local Traffic Responsive Ramp Control. This ramp control mode provides traffic volume and occupancy data from both upstream and downstream detector stations to the ramp metering controller. The data coming from the two adjacent traffic sensor stations can be used in a variety of ways at the local controller. The primary way is to first determine if the downstream section has stable, non-congested flow. If so, then the upstream detectors operate much as in local traffic responsive control. If not, then the downstream flow is congested and more restricted metering rates are implemented.

System Ramp Control. The application of ramp metering to a series of entrance ramps, where the operation of all ramps in the "system" is taken into account, is known as system ramp control. The primary objective of system ramp metering is to prevent freeway congestion. The next highest objective is to respond to unexpected congestion in a systems manner. Ramp control is based on overall system capacity considerations rather than just on the capacity at each ramp. If congestion is to be prevented along the entire freeway, the concept of system ramp control must be used (*20*). System control has been either pretimed system metering (including ramp closure) or traffic-responsive system metering. Another form, gap acceptance merge control, developed at TTI in the late 1960's (*21*), has not been implemented in recent years in the United States and is not considered a viable option at this time. Brief descriptions of the two forms of system ramp control follow.

Pretimed System Control. The metering rate for each ramp in the freeway system is determined in accordance with available capacity constraints at the other ramps as well as its own local available capacity constraint. These metering rates, which are computed from historical data pertaining to each control interval, require the following information:

1. Mainline and entrance ramp demands,
2. Freeway capacities immediately downstream of each entrance ramp, and
3. Description of the traffic pattern within the freeway section to be controlled.

This information provides the basis for establishing the available capacity constraints of the entrance ramps and their interdependencies. Note that this mode of operation does not have the intrinsic capability for responding to incident induced congestion.

Traffic Responsive System Control. Integrated traffic-responsive metering is the application of traffic-responsive metering to a series (group) of entrance ramps where metering rates at each ramp are selected in accordance with system, as well as local, available capacity constraints. During each control interval, real-time measurements are taken of traffic variables (usually volume, occupancy, and/or speed). These data are then used to define the available capacity conditions at each entrance ramp. Then, on the basis of these measurements, both an independent and an integrated metering rate are calculated for each entrance ramp. Of these two metering rates, the one that is the more restrictive is selected to be used during the next successive control interval. Usually, rather than calculating metering rates in real time, a set of metering rates is precomputed for the range of expected traffic conditions from which metering rates are selected in real time.

2.0 EXISTING RAMP METERING STRATEGIES

OVERVIEW

This chapter provides a synthesis of the design and operations of the freeway ramp metering systems identified in Chapter 1. Strategies for calculating metering rates are provided for each system. Coverage begins with the simplest system and progresses to the more complex and costly systems. This synthesis of the literature (17, 18, 20) provides a framework to describe a more advanced ramp metering strategy in the next chapter. Some new technology has also been added to the basic systems to improve their performance.

LOCAL PRETIMED CONTROL

The simplest form of ramp metering is local pretimed control. Ramp metering volumes are established initially based on Highway Capacity Manual (22) methods of analysis which are then fine-tuned based on local field observations. Figure 4 presents the basic concept of local pretimed metering. Pretimed signal timing plans are established based on manually observed mainline volumes, merging volumes, and existing ramp volumes. No traffic detectors are used to provide real-time traffic counts. To maintain the demand-capacity balance on the freeway, ramp merging volumes may need to be reduced. The duration of the fixed flow periods may be shorter than one hour to provide better metering if the flow profile patterns are stable and rather predictable. Note in Figure 4 that current ramp demand is measured by the demand detector used to activate the ramp signal control function of "call for green" (the demand or check-in detector), possibly a "call for red" (a passage or check-out detector), and a "hold the red" (a merge detector not shown). In the most basic ramp metering systems, some of these detector operations may be deleted.

Local pretimed metering rates may be changed every 5 - 15 minutes or so by time-based coordinators (TBC), depending on the stability of the traffic patterns. The metering rate MR(k), for time slice k of duration T, may be selected as the smaller solution from the following three cases. Case 1 - for control based on the outside merging lane capacity, then

$$MR(k) \times ERM + F1VOL(k) = MCAP \qquad (1)$$

where ERM is the ramp metering equivalent (like a passenger car equivalent of trucks) that accounts for both merge speeds being less than freeway speeds and the cyclic dead time in ramp metering operations, F1VOL(k) is the estimated freeway volume in the right-hand merging lane (lane 1) in time slice k, as shown in Figure 4, and MCAP is the merge point capacity (here about 2,200 vph for equivalent unregulated ramp operations). Case 2 - for control based on the total flow across all lanes of the freeway, as also shown in Figure 4,

$$MR(k) + FVOL(k) = FCAP \qquad (2)$$

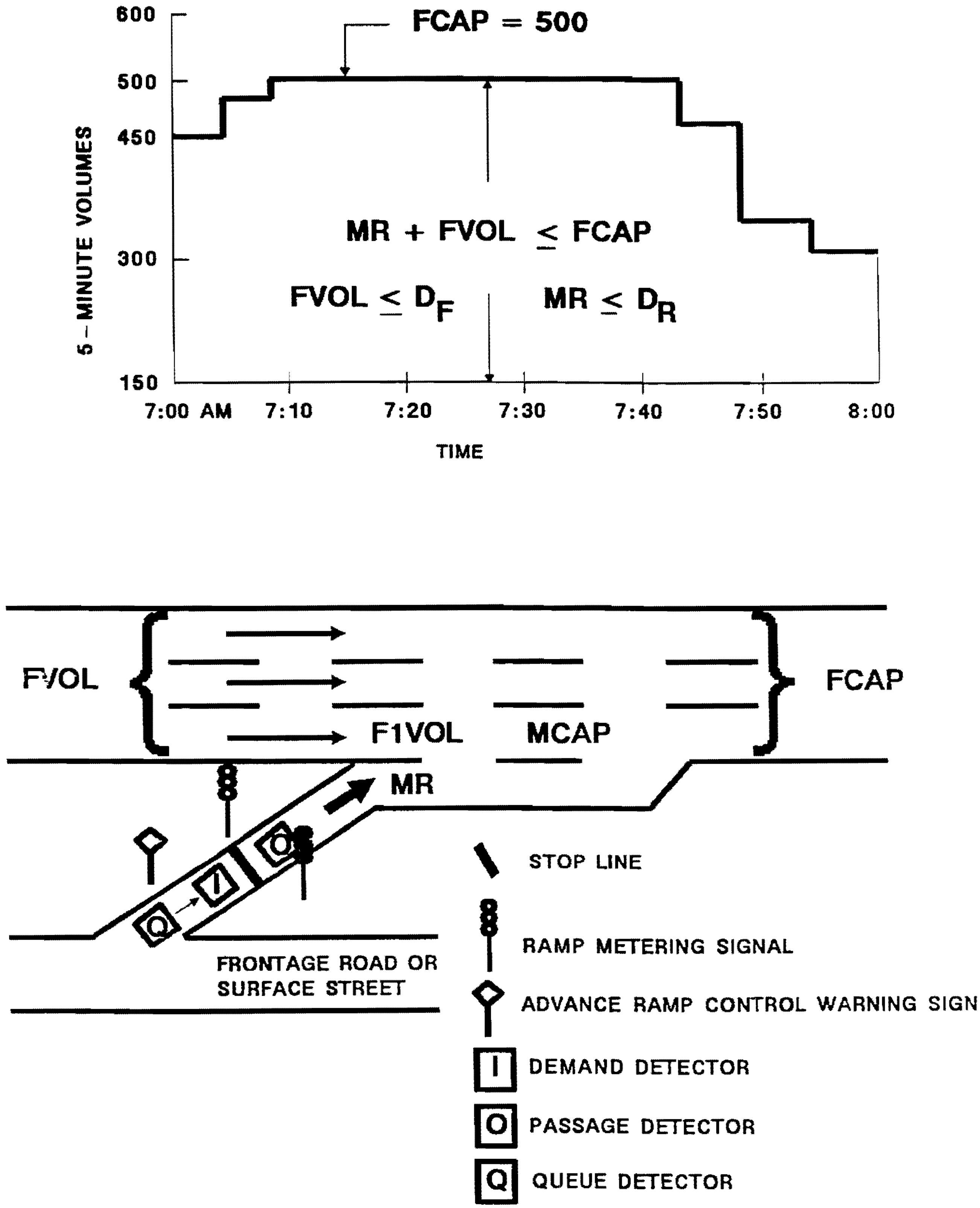


Figure 4. Local Pretimed Ramp Metering.

where FVOL(k) is the mainline freeway arrival volume expected for time slice k, and FCAP is the estimated mainline freeway bottleneck capacity downstream of the entrance ramp. Earlier ramp metering research by Messer suggests that ERM = 1.83 (23). Case 3 - the metering rate cannot exceed some maximum value, nominally a value of about 800 vph. If the meter operates with one vehicle per cycle, then the maximum number of cycles per hour that can be generated depends on the minimum cycle time of C = G + Y + R such that

$$\text{Maximum Metering Rate, } MR(k) = 3600/\ C_m, \quad C_m = 4.5 \text{ sec} \qquad (3)$$

The previous three equations form the core of all demand/capacity ramp metering strategies, either as explicitly enumerated in Equations 1, 2 and 3, or as implied by indirect measurements. The analyst should determine which of the three metering equations controls the local metering process. To aid in this determination, the critical fraction of mainline traffic in the merging lane can be determined from Equation 4 as:

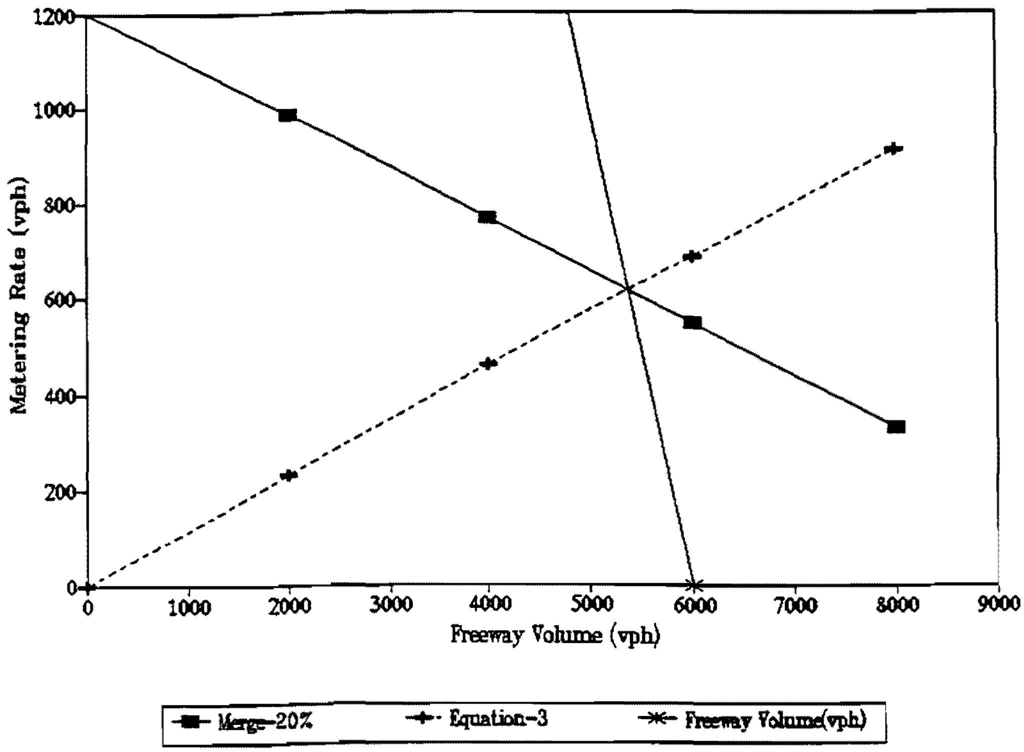
$$CF1 = CF0 - (ERM - CF0)\ FRF \qquad (4)$$

where

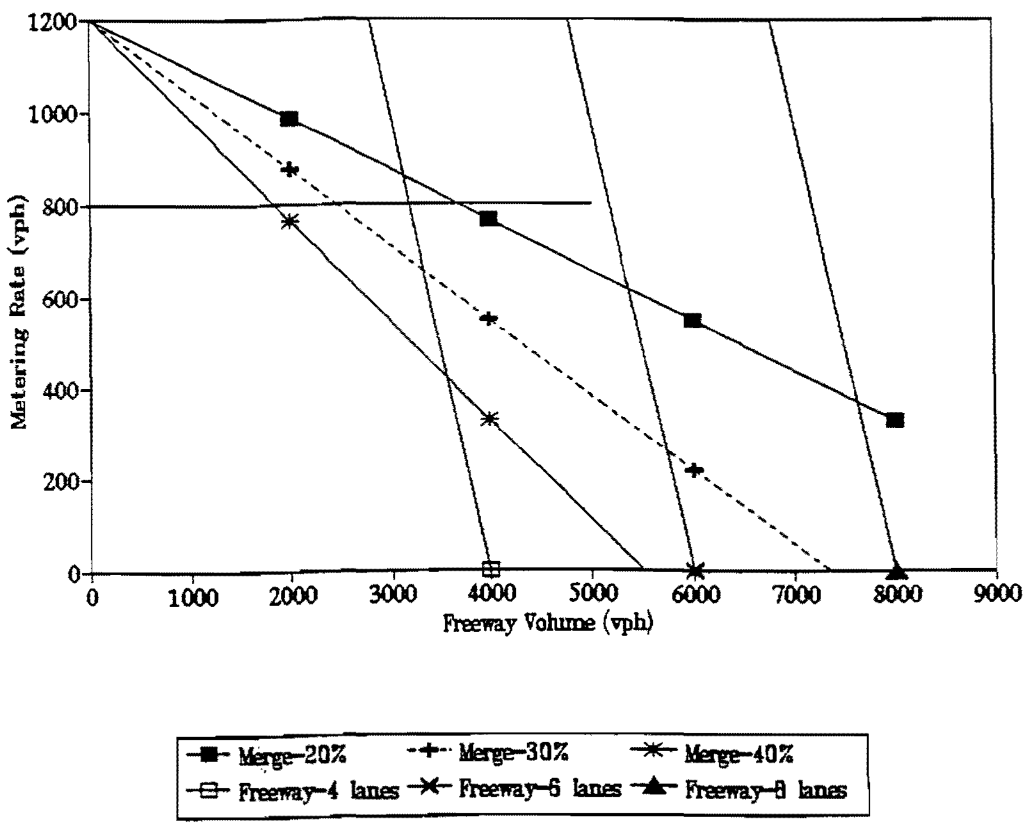
CF1	=	critical fraction of mainline freeway traffic in the merging lane,
CF0	=	base fraction of mainline traffic that depends on the ratio of the freeway merge point capacity, estimated as FCAPM/n, to the downstream bottleneck capacity, FCAP, or CF0 = FCAPM/(n FCAP),
ERM	=	ramp metering mainline capacity equivalent, (about 1.83), and
FRF	=	flow ratio of ramp merge demand volume to mainline freeway volume.

If the actual fraction of mainline traffic is greater than CF1, then ramp merging Case 1 applies (Equation 1); otherwise, Case 2 controls the metering process (Equation 2). This concept is further illustrated in Figure 5. Equation 1 above is plotted in Figure 5 (a) for an assumed 20% of the mainline approach volume being in the merging lane, or F1VOL(k) = 0.20 x FVOL(k) for a merging capacity MCAP of 2,200 vph. Other similar plots of Equation 1 arise as the percentage of mainline traffic varies in the merging lane, as noted in Figure 5 (b). Equation 2 is also plotted in Figure 5 (a) for a six-lane freeway (n = 3 directional lanes) having a downstream bottleneck capacity FCAP of 6,000 vph.

Equation 4 provides the intercept point of Equations 1 and 2, as shown in Figure 5 (a) given only one simplifying assumption, that the ramp merge point capacity is FCAPM/n, where FCAPM is the total capacity of the "n" freeway lanes approaching the merge with the ramp metering equivalency ERM appropriately selected. This merging capacity assumption is consistent with the 1985 Highway Capacity Manual (22). For 20% of the mainline traffic in the merging lane, Figure 5 (a) implies that the bottleneck controls for low metering rates and that the merging capacity controls at higher rates. Figure 5 (b) examines a wider range of merging lane distributions (20, 30 and 40%) for 4-, 6- and 8-lane freeways. This figure shows the importance of lane distribution in establishing metering rates for all freeways.



(a)



(b)

Figure 5. Examination of Lane Distribution on Capacity of Local Ramp Metering.

The prior descriptions of local merging operations can be collectively used to estimate other useful critical flow parameters for the metered ramp when the lane distribution is known. These parameters help identify when traffic conditions are such that the merge capacity controls (Equation 1), rather than the downstream bottleneck (Equation 2), again as illustrated in Figure 5. In addition, these parameters provide insight into the general operational conditions that influence basic demand-capacity metering. Reworking the previous formulations, the critical mainline freeway arrival flow is given by

$$FVOL_{crit} = \frac{(f \times ERM - 1/n)\ FCAPM}{ERM - PF1} \qquad (5)$$

where the resulting volume-to-capacity ratio, XFM, for the freeway arrivals would be

$$XFM_{crit} = \frac{(f \times ERM - 1/n)}{ERM - PF1} \qquad (6)$$

In addition, the critical metering demand flow rate for these conditions is

$$MR_{crit} = \frac{(1/n - f \times PF1)}{ERM - PF1}\ FCAPM \qquad (7)$$

where the new variables are defined as

n = number of freeway mainlanes at the merge point,
f = ratio of FCAP/FCAPM, downstream/upstream freeway capacity, and
PF1 = percentage (fraction) of total mainline arrival flow in merging lane

To illustrate using the same data as for Figure 5-a, we find that $FVOL_{crit}$ = 5,386 vph with XFM_{crit} = 0.842, and that MR_{crit} = 614 vph. PF1, the fraction of total mainline approach flow was assumed to be 0.20 as in Figure 5-a. Thus, for the merging constraint to be active, two conditions must occur simultaneously: (1) the metering rate must exceed 614 vph and (2) the volume-to-capacity ratio must be less than 0.842 (freeway flow less than 5,386 vph).

Assuming equilibrium lane distribution of the mainline freeway flow at the merge point results in the following fraction of through traffic in the merge lane (for PF1)

$$PFE = 1/n \left[1 - \frac{(n-1) \times EMo \times MR}{XFM \times FVOL} \right] \qquad (8)$$

where EMo is the base merging flow equivalency compared to the normal mainline through traffic. Values of EMo range from 1.1 to 1.5, depending on the number of freeway lanes. Recommended values for EMo are 1.1, 1.3, 1.4, 1.5 for n = 2, 3, 4, 5, respectively.

LOCAL TRAFFIC RESPONSIVE CONTROL

The next higher level of control establishes traffic responsive ramp metering rates based on locally measured freeway traffic conditions. The local traffic responsive approach utilizes traffic detectors and a microprocessor to determine the mainline flow upstream of the ramp but in the immediate vicinity of the merge, the average ramp demand, and the ramp queueing status from which to select an appropriate metering rate. Figure 6 presents a typical detection plan for local responsive control. A freeway mainline sampling detector (perhaps two) has been added to the previous plan for local pretimed control (Figure 4).

While it may appear at first glance that traffic responsive control simply measures traffic flow rates in real time for the volumes described above for pretimed control, this traditionally has not been the case even though the concept is implied. The problem with using flow rates directly is twofold. Firstly, many detectors must be used and be highly maintained to accurately count traffic across all lanes in real time. This has been a major issue for many previous ramp metering systems. Secondly, during high-volume traffic, measuring traffic volume alone does not reliably define the state of the system; that is, whether the freeway flow is stable and uncongested, or is unstable and congested. This is, of course, due to the classic parabolic shape of the speed (or density) and volume curve. Many years of ramp metering in Chicago have shown that freeway lane occupancy (a good surrogate measure of traffic density) is the preferred single measure of traffic conditions.

There are two modes of local traffic responsive ramp metering that can be implemented, basically the same two as with pretimed operations. Since traffic volumes are not traditionally used, however, calibrated operations with indirect measures are required. One mode controls demand to the merge point, the other controls demand based on total freeway capacity. It is unlikely that both control objectives can be achieved simultaneously with only one control algorithm. The operator should select the preferred metering mode for each ramp. Figure 7 illustrates the classic metering relationship used for almost all local traffic responsive ramp metering systems. The metering rates shown in Figure 7 for local responsive control can be derived from traffic flow theory and the demand/capacity equations given previously for pretimed metering. Field adjustments are made to fine tune the derived ramp metering rates to local geometric and traffic conditions.

Traffic responsive ramp metering of the merge point reflects the fundamental demand/capacity management objective of merging operation. This strategy seeks to keep the sum of the freeway lane one volume arriving at the merge plus the ramp merging volume, as measured in real-time "t", from exceeding the merge point capacity. To achieve this control objective, the metering rate should not exceed MR(t) as calculated from

$$MR(t) \times ERM + F1VOL(t) = MCAP \qquad (9)$$

where ERM is the ramp metering equivalency factor (about 1.83), F1VOL(t) is the freeway traffic flow in the merging lane in real-time t, and MCAP is the estimated ramp merging

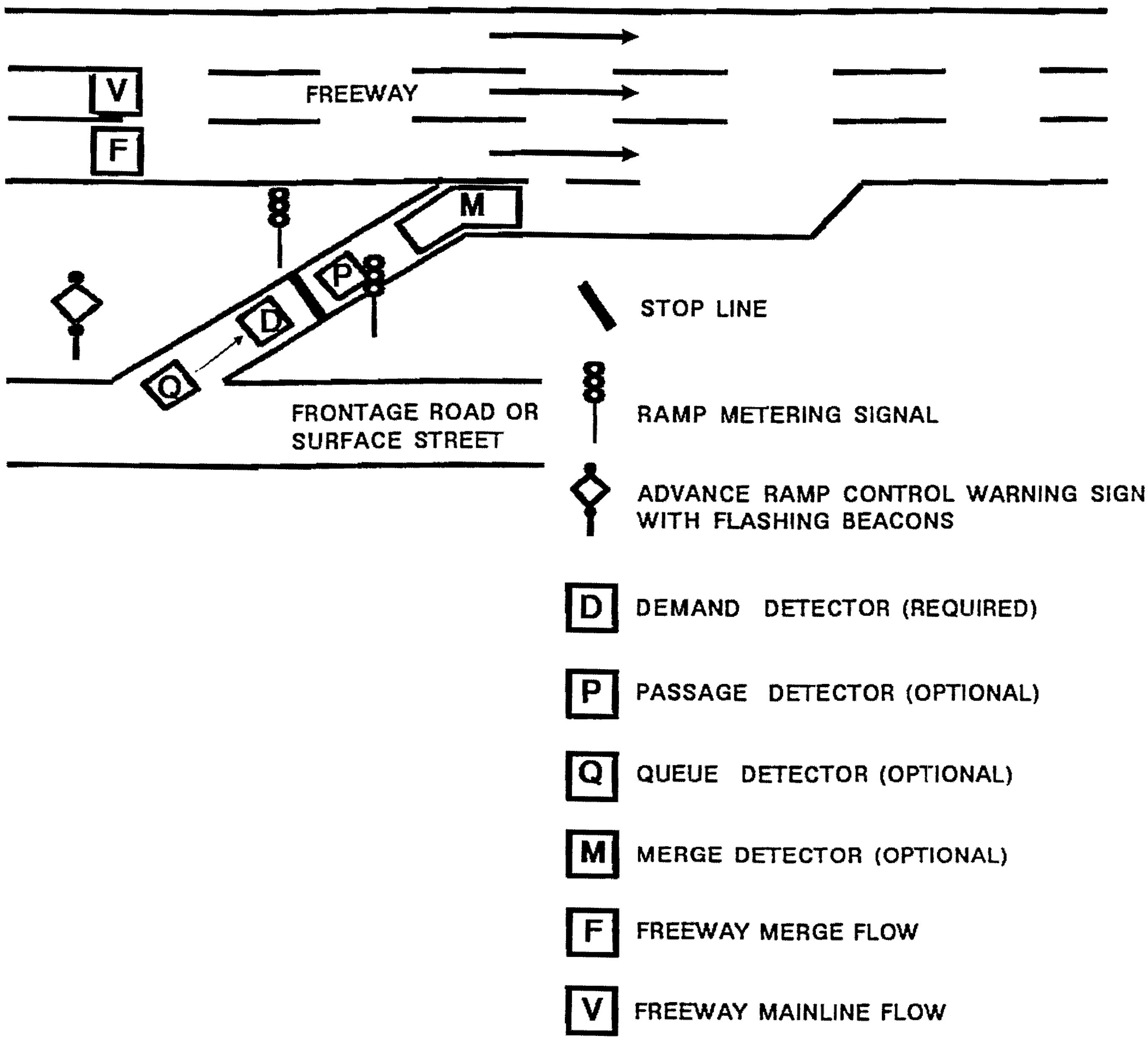


Figure 6. Local Traffic Responsive Ramp Metering.

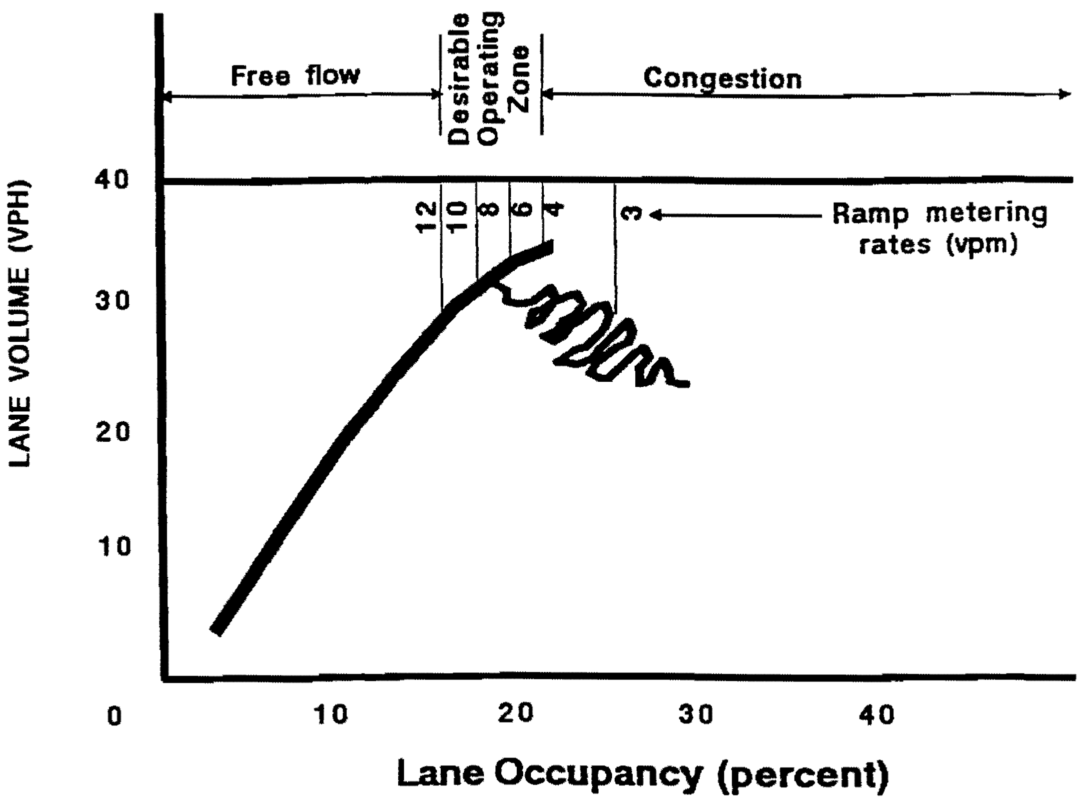


Figure 7. Typical Local Traffic Responsive Ramp Metering Strategy.

capacity (about 2,200 vph). Functionally relating lane volume F1VOL(t) to lane density D1(t) and assuming the generalized traffic flow shape parameters m and l as given by May (24) yields the traffic responsive metering rate MR(t) related solely to lane density of

$$MR(t) \times ERM = MCAP - K1 \times D1(t|m,l) \tag{10}$$

Since traffic density $D1(t|m,l)$ is directly and linearly related to measured lane occupancy such that $D1(t) = Kd \times O1(t)$, the ramp metering rate MR(t) can be expressed as

$$MR(t) \times ERM = MCAP - K1 \times Kd\ O1(t|m,l) \tag{11}$$

The calibration coefficient, Kd, depends on the length of the detector used and the average length of vehicles being detected in the merging lane. That is, Kd = 52.8/s where s is the sum of the average length of the vehicles in the traffic stream plus the effective length of the detector being used to sense the traffic flow. Using typical values for average vehicle length of 15 feet (4.6 m) and detector length of 6 feet (1.8 m), then s = 21 feet (15 + 6), or (6.4 m = 4.6 m + 1.8 m) . The resulting Kd value would be 2.51 (52.8/21) for converting lane occupancy O1 (in %) in the merging lane to lane density.

Letting Kd equal 2.51, then the metering rate that would satisfy the lane 1 merge point capacity constraint for a given measured occupancy in the merging lane would be

$$MR(t) \times ERM = MCAP - 2.51\ K1\ O1(t|m,l) \tag{12}$$

as suggested by Figure 7. Clearly, practical ramp metering, even at the local level, should be considered both (1) a theoretical modeling process using field calibration of parameters and (2) an empirical field implementation task to assess the parameters yet guided by the theoretical concepts, due to the large number of parameters involved in the process.

Local traffic responsive ramp metering may also have as its control objective to maintain total freeway demand less than the "across all lanes" capacity downstream of the entrance ramp. In this case, we have a similar equation to that of merge point metering of

$$MR(t) = FCAP - 2.51\ Kf\ Of(t|m,l) \tag{13}$$

where $Of(t|m,l)$ is the average "across all freeway lanes" lane occupancy. The calibration factor, Kf, for density depends on the number of freeway mainlanes and on the traffic flow parameters. The metering rate could be calculated from Equation 8 (also as suggested in Figure 7) as long as the traffic density and lane occupancy do not exceed the freeway capacity thresholds. FCAP, the downstream "bottleneck," is normally a little less than the mainline capacity at the merge, FCAPM, which is about n (the number of directional freeway mainlanes) times MCAP. FCAP should always reflect the best estimate of the current downstream freeway bottleneck capacity, whose value should include the effects of traffic mix, incidents and weather.

An analysis of local traffic responsive ramp metering using lane occupancy as the point indication of the "state of the system" reveals the following aspects. Firstly, lane occupancy is a good single point indicator, if not the best, because it uniquely characterizes all possible states of traffic flow: stable, unstable, and forced flow. Occupancy is also strongly correlated to traffic density, the primary level of service measure.

Secondly, local responsive control usually can provide only one metering strategy. This strategy can be based either on single-lane merging capacity, or on total mainline demand/capacity balance, but not both with a single metering algorithm. This limitation may not have been recognized and some may not have determined which metering option is being used. Such a design limitation can negatively impact system performance.

Thirdly, when traffic demand is rising but has not exceeded downstream capacity, upstream measurement of demand is desired so that a feed-forward control system is provided. That is, present conditions depend only on past conditions upstream of the control element that occurred in an earlier time period. But if downstream congestion forms for any reason, future traffic conditions at the ramp are best estimated from feed-back systems such that downstream (not upstream) conditions are the best indicator of future traffic conditions at the subject ramp. Downstream monitoring of traffic flows across all lanes could also provide updated estimates of traffic capacity as traffic mix and environmental factors change during the control period.

Thus, the benefit of providing a "system design" that provides a broader view of freeway ramp metering is apparent. However, there are several types of "system designs." The next level of control provides for the horizontal integration (i.e., data flowing between adjacent ramps) of data needed for adaptive local traffic responsive ramp metering.

ADAPTIVE LOCAL TRAFFIC RESPONSIVE RAMP METERING

This ramp control mode provides traffic volume and occupancy data from both upstream and downstream detector stations to the ramp metering controller. Figure 8 illustrates the design of this type of control. The data coming from the two separate traffic sensor stations can be used in a variety of ways at the local controller.

The principal adaptive control use would be for monitoring the status of the downstream occupancy to determine if traffic congestion is backing upstream along the freeway into the control area of the ramp. Seattle has found that if the lane occupancy exceeds 18% on their system, congestion has reached that loop detector station (*15*). When congestion is detected, the metering rate is no longer determined by the upstream detectors but by the downstream detectors. This shift of monitoring station for control is a rather powerful adaptive control feature. A second use of the downstream detectors could be to monitor the maximum flows (given the lane occupancy) to better estimate the present traffic mix and freeway capacity at that location.

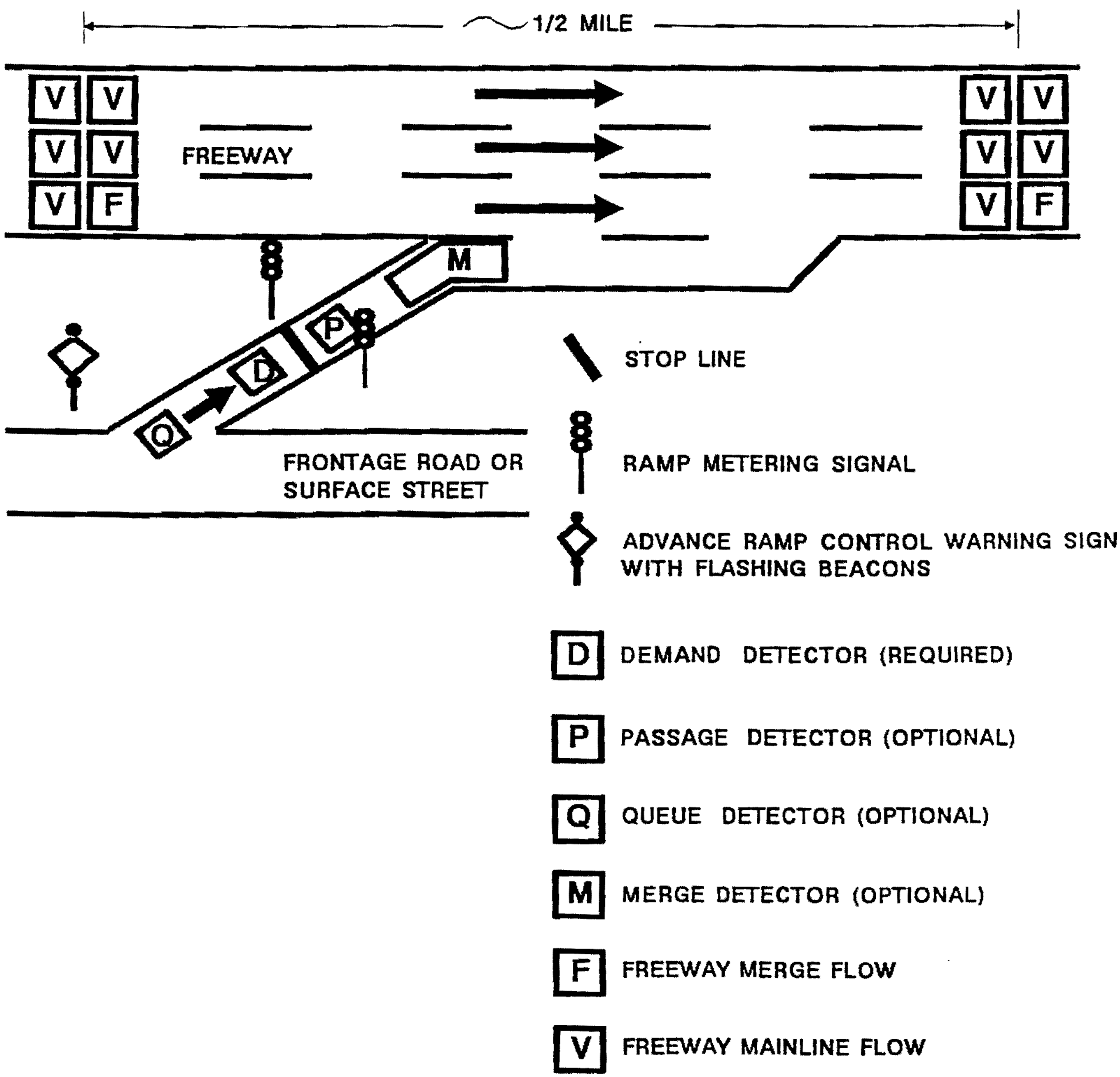


Figure 8. Adaptive Local Traffic Responsive Ramp Metering.

Other system concepts, such as horizontally passing traffic flow data between entrance ramp controllers, have been proposed or used. TxDOT has restricted ramp metering rates at upstream meters once the presence of downstream congestion was identified (25). Washington DOT uses downstream congestion indicators to reduce upstream metering rates (16). Denver implements a "system coordination plan" based on the status of downstream metering rates that are checked every 20 seconds (26). None of these methods for achieving spatial (system) ramp metering coordination are totally satisfying from a control systems viewpoint because of their inherent empirical nature. A representative example of the operations of these metering algorithms follows (18).

Assume a control system has across-all-lanes detector count stations at regular intervals and speed detectors at critical bottleneck locations. Any given entrance ramp has count stations both upstream and downstream of the ramp, together with a speed measurement at a critical bottleneck farther downstream. The strategy compares each of these three parameter measurements against a table of threshold volumes, and preassigned metering rates would be obtained for each of these measurements. One strategy might pick the smallest of the stored metering rates as the implemented metering rate. For example, assume a 3-lane freeway section with a 1-minute upstream flow rate of 95 vehicles, a downstream flow of 85 vehicles, and a critical downstream bottleneck speed of 25 mph. (Occupancy could replace speed in this algorithm.) The associated parameters are given in Table 2, and the metering rates for upstream, downstream, and bottleneck locations, respectively, are 3, 6, and 9 vehicles per minute (vpm). Selecting the minimum metering rate of 3 vpm would dictate that the entrance ramp meter 3 vpm over the next minute, at which time the parameter movements would again be evaluated for the following minute's metering rate, and so on.

Table 2. Ramp Metering Strategy Parameters.

Upstream		Downstream		Bottleneck	
1 Min. Vol.	Metering Rate vpm	1 Min. Vol.	Metering Rate vpm	Speed mph	Metering Rate vpm
91-100	3	91-100	3	41-50	15
81-90	6	81-90	6	31-40	12
71-80	9	71-80	9	21-30	9
61-70	12	61-70	12	11-20	6
<61	15	<61	15	0-10	3

SYSTEM RAMP CONTROL

System ramp metering refers to the application of ramp control to a series of entrance ramps where the interdependency of ramp operations is analytically taken into account. The primary objective of system ramp metering is to prevent freeway congestion from forming on the mainlanes. The control of each ramp is based on the overall capacity considerations for the whole system rather than on just the capacity at each individual ramp. If congestion is to be prevented along the entire freeway, the concept of system ramp control must be used (*20*). System control has two forms of operation: Pretimed metering (including ramp closure) and traffic-responsive metering. A detailed description of these forms of system ramp control follows.

Pretimed Systems Metering

The metering rate for each ramp in pretimed systems is determined in accordance with available capacity constraints at the other ramps as well as its own local available capacity constraint. These metering rates, which are computed from historical data pertaining to each control interval, require the following information:

- Mainline and entrance ramp demands,
- Freeway capacities immediately downstream of each entrance ramp, and
- Description of the traffic pattern within the freeway section to be controlled.

This data provides the basis for establishing the demand on the system together with the available capacity constraints of the entrance ramps and their interdependencies.

Given the required data, the fundamental procedure for computing the pretimed metering rates involves five steps (*17*):

1. Start with the entrance ramp which is farthest upstream.
2. Determine the total demand (upstream mainline demand plus ramp demand) for the freeway section immediately downstream of the ramp.
3. Compare total demand to capacity of downstream section, and proceed as follows:
   a. If the total demand is less than the capacity, metering is not required at the ramp by this capacity constraint. Therefore, skip step 4 and go immediately to step 5.
   b. If the total demand is greater than the capacity, metering is required at this ramp by this capacity constraint. Therefore, proceed to step 4.
4. Compare the upstream mainline demand to the capacity of the downstream section, and proceed as follows:
   a. If the upstream mainline demand is less than the capacity, then the allowable entrance ramp volume (or metering rate) is set equal to the difference between the capacity and the upstream mainline demand.
   b. If the upstream mainline demand is greater than or equal to the capacity, then the

allowable entrance ramp volume would be zero, and the ramp should be closed. If the upstream mainline demand is greater than the capacity, the volumes permitted to enter at ramps upstream must be reduced accordingly. The total reduction in the allowable entrance ramp volumes upstream would be equal to the difference between the upstream mainline demand and capacity, adjusted to account for that portion of the traffic entering upstream that exits before it reaches the downstream entrance ramp being closed.

5. Select the next entrance ramp downstream, and go back to step 2. This procedure is illustrated by the following three examples (18).

Example No. 1 -- Pretimed metering rates are to be calculated for the example shown in Figure 9 for integrated systems control such that the average traffic demand does not exceed any bottleneck capacity along the freeway system. In reviewing this example, the following points should be noted:

Since only entrance ramp control is being considered and not mainline control, the allowable mainline volume entering the system is set equal to the mainline demand D. Using the notation given in Figure 9, the demand S_j at a section j is computed by the following equation:

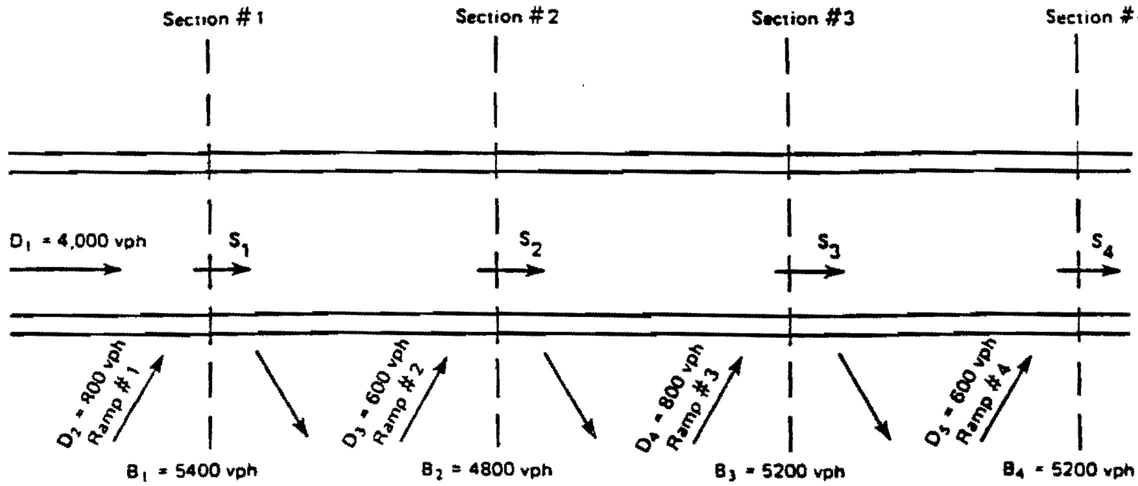
$$S_j = \sum_{i=1}^{N} A_{ij} X_i + A_{j+1,j} \quad D_{j+1} \qquad (14)$$

where

S_j = calculated pretimed demand at section j,
X_i = allowable volume at input i over all inputs N feeding j,
D_i = demand at input i, and
A_{ij} = fraction of vehicles at input i which pass through section j.

As it happens in this example, the metering rate computed for each entrance ramp is determined solely by the capacity constraint at the section immediately downstream of the ramp and is not influenced by the capacity constraints at the other ramps. This case is basically equivalent to local ramp metering operations.

Example No. 2 -- The data used in this example is given in Figure 10. This data is similar to the previous case except that the mainline demand has increased from 4,000 vph to 4,600 vph. In this example, the metering rates at ramps 2, 3, and 4 are set based solely on their respective downstream capacity constraints as before. However, the metering rate at ramp 1 is established based on the capacity constraint at ramp 2 rather than ramp 1, as described in the following paragraphs.



X_i = allowable volume at input i

D_i = demand at input i

B_j = capacity at section j

A_{ij} = decimal fraction of vehicles entering at input i which pass through section j

S_j = demand at section j

A_{ij} Values

i \ j	1	2	3	4
1	1.00	0.95	0.90	0.85
2	1.00	0.75	0.70	0.60
3	–	1.00	0.90	0.85
4	–	–	1.00	0.90
5	–	–	–	1.00

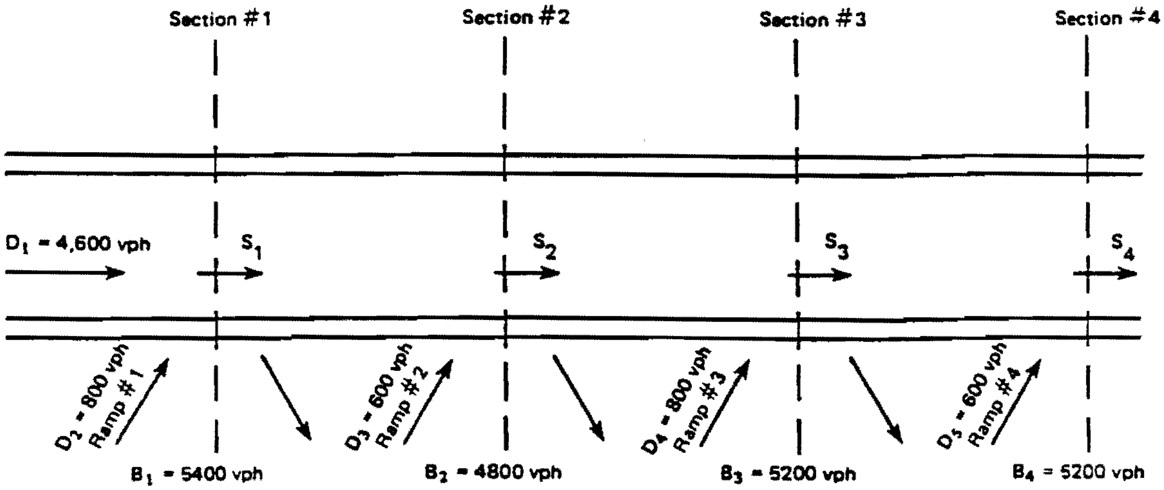
<u>Compute X_i's starting at section #1</u>

- Set $X_1 = D_1 = \underline{4000 \text{ vph}}$
- $S_1 = A_{11}X_1 + A_{21}D_2 = (1.00)(4000)+(1.00)(800) = 4800 \text{ vph} < B_1 = 5400 \text{ vph}; \therefore \underline{X_2 = 800 \text{ vph}}$
- $S_2 = A_{12}X_1 + A_{22}X_2 + A_{32}D_3 = (0.95)(4000)+(0.75)(800)+(1.00)(600) = 5000 \text{ vph} > B_2 = 4800 \text{ vph}; \therefore \underline{X_3 = 400 \text{ vph}}$
- $S_3 = A_{13}X_1 + A_{23}X_2 + A_{33}X_3 + A_{43}D_4 = (0.90)(4000)+(0.70)(800)+(0.90)(400) + (1.00)(800) = 5320 \text{ vph} > B_3 = 5200 \text{ vph}; \therefore \underline{X_4 = 680 \text{ vph}}$
- $S_4 = A_{14}X_1 + A_{24}X_2 + A_{34}X_3 + A_{44}X_4 + A_{54}D_5 = (0.85)(4000) + (0.60)(800) + (0.85)(400) + (0.90)(680) + (1.00)(600) = 5432 \text{ vph} > B_4 = 5200 \text{ vph}; \therefore \underline{X_5 = 368 \text{ vph}}$

<u>Conclusion:</u>

- Ramp #1: No control needed.
- Ramp #2 : Meter at a rate of 400 vph.
- Ramp #3: Meter at a rate of 680 vph.
- Ramp #4: Meter at a rate of 368 vph.

Figure 9. Example No. 1 - Calculation of Pretimed Metering Rates.



X_i = allowable volume at input i

D_i = demand at input i

B_j = capacity at section j

A_{ij} = decimal fraction of vehicles entering at input i which pass through section j

S_j = demand at section j

A_{ij} Values

i \ j	1	2	3	4
1	1.00	0.95	0.90	0.85
2	1.00	0.75	0.70	0.60
3	–	1.00	0.90	0.85
4	–	–	1.00	0.90
5	–	–	–	1.00

Compute X_i's starting at section #1

- Set $X_1 = D_1 = \underline{4600 \text{ vph}}$
- $S_1 = A_{11}X_1 + A_{21}D_2 = (1.00)(4600)+(1.00)(800) = 5400 \text{ vph} = B_1 = 5400 \text{ vph}; \therefore \underline{X_2 = 800 \text{ vph}}$
- $S_2 = A_{12}X_1 + A_{22}X_2 + A_{32}D_3 = (0.95)(4600)+(0.75)(800)+(1.00)(600) = 5570 \text{ vph} > B_2 = 4800 \text{ vph};$

  $\therefore \underline{X_3 = 0}$ and volume entering upstream must be reduced by 170 vph;
  $\therefore X_2 = 800 - 170/A_{22} = 800 - 170/0.75 = \underline{573 \text{ vph}}$
- $S_3 = A_{13}X_1 + A_{23}X_2 + A_{33}X_3 + A_{43}D_4 = (0.90)(4600)+(0.70)(573)+(0.90)(0)$
  $+ (1.00)(800) = 5341 \text{ vph} > B_3 = 5200 \text{ vph}; \therefore \underline{X_4 = 659 \text{ vph}}$
- $S_4 = A_{14}X_1 + A_{24}X_2 + A_{34}X_3 + A_{44}X_4 + A_{54}D_5 = (0.85)(4600) + (0.60)(573) + (0.85)(0) + (0.90)(659)$
  $+ (1.00)(600) = 6447 \text{ vph} > B_4 = 5200 \text{ vph}; \therefore \underline{X_5 = 353 \text{ vph}}$

Conclusion:

- Ramp #1: Meter at a rate of 573 vph.
- Ramp #2: Close.
- Ramp #3: Meter at a rate of 659 vph.
- Ramp #4: Meter at a rate of 353 vph.

Figure 10. Example No. 2 - Calculation of Pretimed Metering Rates.

The demand S_2 at section 2 is 5,570 vph, which is 770 vph greater than the capacity B_2 at section 2 (4,800 vph). If ramp 2 is closed, the demand at section 2 is reduced to 4,970 vph, a volume which also exceeds the capacity B_2. Therefore, it is necessary to reduce the allowable volume X_2 entering at ramp 1 (input 2). The allowable volume X_2 must be reduced enough to lower the demand S_2 by 170 vph. The amount of the reduction is equal to the 170 vph divided by the decimal fraction A_{22} of the vehicles entering at ramp 1 and passing through section 2 (170 vph/0.75 = 227 vph.) Therefore, the allowable volume X_2 at ramp 1 would be 573 vph instead of 800 vph.

In the procedure outlined above, excess freeway demand S_j - B_j at any section j is removed by reducing the allowable volume on the entrance ramp immediately upstream. If, instead, the allowable volumes on the entrance ramps farther upstream were reduced, a larger number of vehicles would have to be removed from these ramps in order to reduce the demand S_j sufficiently at section j. This is necessary so that some of the vehicles that enter these ramps will exit the freeway before they reach section j.

Example No. 3 -- Consider again the situation presented earlier in Figure 9. If the excess freeway demand (200 vph) were removed by reducing the allowable volume X_2 at ramp 1, the volume at ramp 1 would have to be reduced by 267 vph. Consequently, the allowable entrance ramp volumes would be the following: Ramp 1: 533 vph, Ramp 2: 600 vph, Ramp 3: 687 vph, Ramp 4: 352 vph, such that the total input = 2,172 vph.

The total input of 2,172 vph, however, is less than that of 2,248 vph, the volume which is obtained if ramp 2 is metered as in Example No. 1. Thus, the fundamental approach described will result in the optimal utilization of the freeway. It maximizes the sum of the allowable entrance ramp volumes, a procedure that corresponds to maximizing system output for steady-state, uncongested flow conditions. Total travel time in the system is also minimized (*27*).

Linear Programming for Systems Control. The fundamental process described for Examples 1 and 2 was first formulated as a linear programming (LP) model by Wattleworth in 1963 (*12*). This basic LP model can be used to compute optimal allowable entrance ramp volumes according to the model. In terms of the notation defined above, this basic linear programming model would be:

Step 1. Objective Function: Generally, maximize the total input flow rate per unit time.

$$\text{Maximize } P = \sum_{i=1}^{N} X_i \quad \text{where N is the number of inputs.}$$

Step 2. Constraints: Subject to the following constraints:

Mainline demand/capacity:

$$\sum_{i=1}^{N} A_{ij} X_i \leq B_j \quad j = 1,2, \ldots, L$$

Mainline demand must be served:

$$X_1 = D_1$$

Metered ramp volume can not exceed ramp demand:

$$X_i \leq D_i,\ i = 2,3, \ldots, N$$

Metered ramp volume should not be less than tolerable rate:

$$X_i \geq \min X_i \quad i = 2,3, \ldots, N$$

Use of this linear programming model yields allowable entrance ramp volumes that equal or exceed those manual procedures previously described.

<u>Integrated Systems Traffic Responsive Control</u>

This ramp metering strategy is the application of traffic-responsive metering to a series of entrance ramps where the metering rates at each ramp are selected in accordance with system, as well as local, available capacity constraints.

During each control interval, real-time measurements are taken of the relevant traffic variables, (usually volume, occupancy, and speed). The data are used to define the available capacity conditions at each entrance ramp. Based on these measurements, local and integrated metering rates are calculated for each entrance ramp. Of these two metering rates, the one that is the more restrictive is selected to be used during the next successive control interval.

The methods used to calculate the independent and integrated traffic-responsive metering rates are basically the same as those used to compute independent and integrated real time metering rates. Usually, rather than calculating metering rates in real time., a set of metering rates is precomputed for the range of expected available capacity conditions from which metering rates are selected in real time. The linear programming model may be used to calculate predetermined sets of integrated traffic-responsive metering rates. Simpler algorithms are normally used. Also, the metering rates are subject to the overrides for the merge detector, queue detector, and maximum red time used in traffic responsive metering.

Operational Considerations. Operational experience has shown that there are some practical considerations when determining ramp metering rates. These are based on both driver behavior and system limitations.

Low metering rates of less than 180 to 240 vph (cycle times of 20 to 15 seconds, respectively) are not considered feasible because drivers being metered would have to wait too long at the ramp signal. At longer cycle times, operational experience has shown that drivers often become frustrated and will run the signal. Thus, if a metering rate of less than 180-240 vph is calculated, consideration should be given either to closing the ramp or to metering it at a higher rate. The preferred choice depends on the availability of quality alternative routes for the diverted traffic and on equity issues. The LP model described above will not develop a metering rate less than that allowed, if the minimum metering rate is mathematically feasible based on freeway capacity considerations.

Variability in metering rates is also a practical problem with very dynamic systems. The pace of metering between time periods should not be reduced significantly unless operational problems on the freeway mainlanes are readily visible to the ramp traffic. A reasonably predictable metering pace at a ramp is desirable for a given day and time of day. Many motorist have selected the ramp after estimating their ramp delay. Actual ramp delays probably should not exceed expected ramp delays by more than 1 minute between days. Similarly, ramp service times probably should not change by more than 25% between control periods for the same day. Predictability of delays is important to urban motorists in optimizing their trip schedules to arrive at their destinations on time. A predictable trip travel time lets the motorist leave at the latest possible moment.

Maximum metering rates also exist that depend on the type of metering system in operation. Maximum theoretical metering rates of about 800 vph exist for single lane metering as produced by the minimum red signal interval permitted, given fixed intervals for the green interval (say 1.5 seconds) and yellow signal (say 2.0 seconds). Given these timings, a 1.0 second minimum red interval would produce a 4.5 second cycle, or 800 cycles per hour. With typical rush hour mainline volumes, this metering rate is a highly theoretical maximum value for several reasons. Firstly, there simply may not be adequate capacity to merge this high entry flow rate. Secondly, there will be times when the merge area is blocked and the meter will be momentarily stopped when using traffic responsive control. These lost times can not be made up at maximum metering levels. So practically speaking, maximum metering rates may not exceed perhaps 700 vph even though the (theoretical) "metering rate" is set at 800 vph (per lane). To achieve higher rates would probably require either bulk metering, allowing more than one vehicle to go per lane per cycle, or multiple lane metering with either simultaneous or staggered departures.

Should a metering rate be calculated that is near zero, some authors have suggested that the ramp should be closed. This is one option, but it raises the social question of fairness, or "equity" of access to the freeway system by users. This result may be due to the limitations in the ramp control system area of coverage, the control software, and/or the overall excess demand on the system.

Some of these limitations can be treated using more sophisticated interactive freeway traffic management software described in Chapter 3. It is noted that the equity issue noted above was observed by TTI staff on the Gulf Freeway Project in Houston in 1968 (*14,23*), and ramp metering design features to ameliorate the problem have been included in our proposed advanced freeway traffic management system.

Metering rates at each entrance ramp should be evaluated with regard to the adequacy of available storage at the ramp and the potential congestion that might be created on the adjoining surface street (or frontage road) system. If storage is not sufficient, it may be necessary either to close the ramp or to increase the metering rate of the subject ramp while reducing others. This queue management strategy is not available explicitly on any operational ramp metering system today; however, FREQ10 (28) provides some capabilities in this area for engineering analysis purposes. Model 2 of TTI's proposed advanced FTM presented in Chapter 3 explicitly provides this queue management option, with the ability to visually and interactively manage the queueing situation.

The Denver System. An example of algorithmic-based, horizontally integrated ramp metering systems is represented by the present system in Denver, Colorado (26) shown in Figure 11. The primary integration feature provides transfer of control from local control to central control as traffic congestion is detected on the freeway or at the ramp meters.

The local ramp controllers are programmed to operate only during weekdays for one or both of the peak periods. During metering periods, each ramp meter initially selects one of six metering rates on the basis of local traffic conditions. Mainline primary and secondary detectors are used to determine the volume, occupancies, and speeds in each lane. The secondary detectors function as a backup for the primary detectors. The ramp presence and passage detectors inform the controller when a vehicle is waiting to be metered and when it has passed the metering signal. A 2-sec delay is programmed into the controller to prevent a vehicle from receiving an immediate green indication on arrival. Rate selections are made every 20 sec according to the information provided by the mainline detectors. An exponential smoothing function is included in the measurement algorithms to prevent rapid switching between flow rates.

Queue detectors are installed near the entrance of the ramp to sense when vehicles are backing toward the cross street. When the occupancies of the queue detectors exceed an established threshold, the controller overrides the normally selected metering rate. The controller then initiates less restrictive rates until the backup is reduced to an acceptable level. A HOV detector located in a parallel ramp, when actuated, extends the red signal time by a preset amount to help avoid conflicts in the merge area.

Every 20 sec the central computer collects detector and metering rate data from each ramp. If a ramp is in the most restricted rate (freeway congested) or in queue override (ramp congested), the ramp is defined as critical. A "system coordination plan" is then put into effect. This plan calculates the travel time between ramps and, after that time, forces a more restrictive metering rate on the next upstream ramp. If the ramp remains in a critical condition during the next 20-second sampling period, the rates of the next two upstream ramps are forced to the next more restrictive rate. The system continues to add upstream ramps to the coordination plan during each sampling cycle until all ramps in the freeway group are under central control. If more than one entrance ramp becomes critical, multiple plans are put into effect.

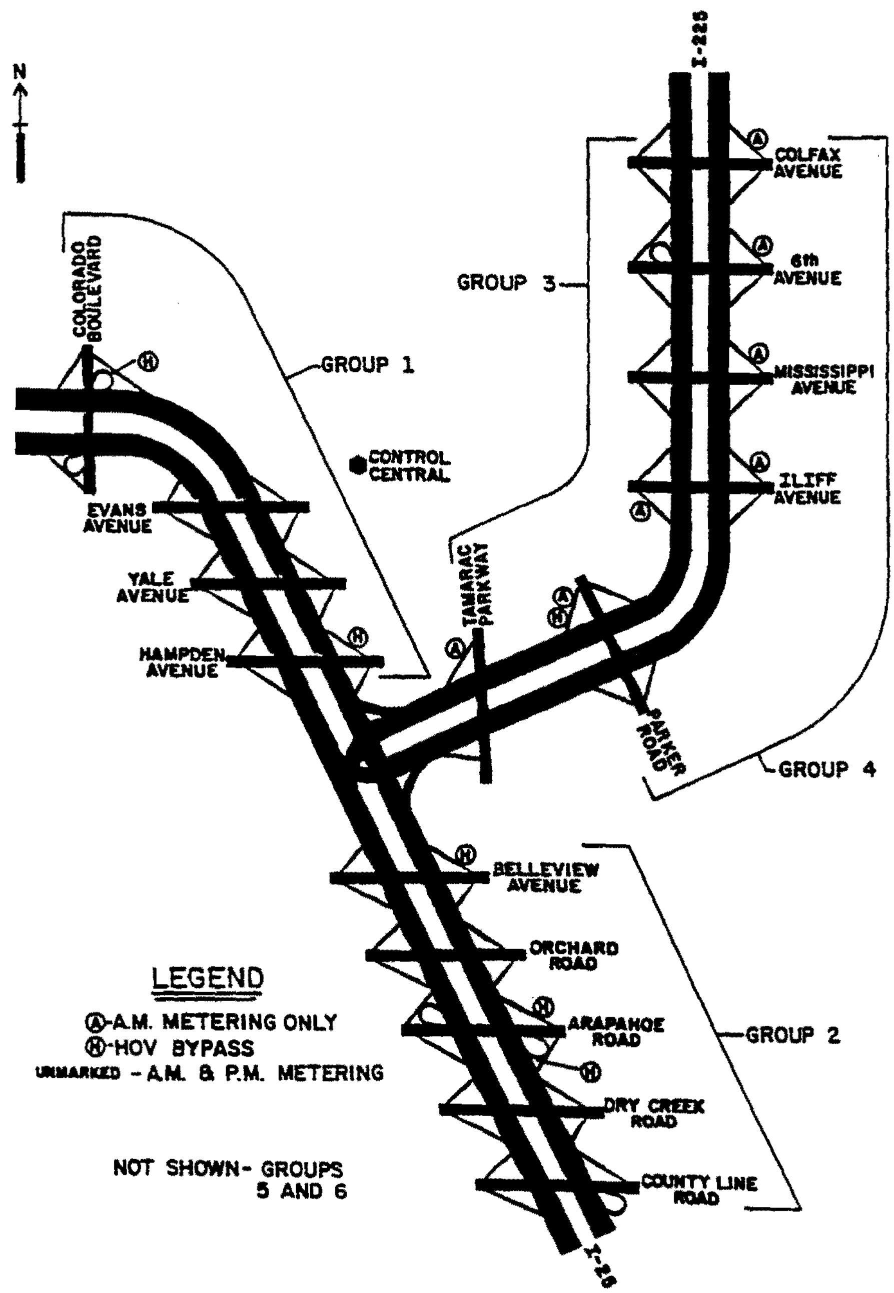


Figure 11. The Denver Area Ramp Metering System.

If and when the last ramp in a group is put under the system coordination plan, the central computer begins implementing coordination in the first ramp of the next upstream group that feeds the critical ramp, as identified in Figure 11. The system coordination plan continues until the ramps return to a noncritical condition. The plan returns the local controllers to normal operation one rate at a time at the 20-sec sampling intervals.

The central computer also monitors the traffic just before and after the peak periods to determine if metering should occur earlier or later than the core metering times. If conditions are favorable for ramp control, the central computer allows the local controllers to begin metering up to 20 minutes before and to remain on 30 minutes after the mandatory metering times.

3.0 ADVANCED SYSTEM METERING ALGORITHMS

BACKGROUND

As suggested in Chapter 2, freeway system's demand/capacity assessments can neatly be transformed into realistic linear programming formulations. Wattleworth and Berry initiated the first studies to use linear programming techniques to determine the optimum steady state ramp flow rates in Chicago in the early 1960's (12). The basic control objective was to maximize system input thereby minimizing system travel time. Messer improved Wattleworth's model for multilevel freeway control systems (23). Yuan and Kreer proposed an optimal control algorithm having queue-balancing capabilities (29). Chen and Cruz further proposed an on-ramp control model to maximize traffic flow rates on freeway mainlines (30). Papageorgiou suggested that dynamic programming be used for time-of-day ramp metering control (31).

The four optimization models identified above can all be categorized as sequential solution approaches. In the sequential approach, the optimization problem is solved separately for each time slice, and the total control plan is the sum of the individual plans. Most of these models were designed for steady state freeway traffic operating conditions.

DYNAMIC MODELS

The models presented in this chapter improve both the equality and efficiency of freeway ramp control by solving system problems for a group of time slices, simultaneously, through a "Dynamic Model" (32, 33). Important freeway geometric and operational factors impacting the performance of freeway ramp control in the models include the following:

- Freeway mainline input volume,
- On-ramp entrance volume,
- Off-ramp exit volume,
- Freeway link capacity,
- Freeway bottleneck capacity,
- On-ramp capacity,
- Probability of vehicles in merging lane, and
- Percent of on-ramp queue diversion.

These system factors can be used as input to determine optimal ramp metering plans. Optimal system performance evaluations can include system travel time, ramp queue, and ramp delay time. System travel time describes the total travel time by all vehicles traveling on the freeway during the control period. Ramp delay measures the average time any user must wait at an entrance ramp, in contrast to ramp queue used by some engineers to evaluate the effectiveness of freeway ramp control systems.

MODEL FORMULATION

The freeway control problem is to determine the optimal input flow for the given traffic demands, ramp geometrics, and freeway capacities. The system must maximize input flow and balance ramp queues by fully utilizing freeway and ramp storage capacities. The control objective is to maximize the total freeway input flow over the control period (the sum of the sequential time slices) subject to the following constraints:

1. Ensure that freeway demand does not exceed capacity anywhere along the freeway,
2. Impose merging capacity on freeway entrance ramp merging,
3. Satisfy the freeway mainline input flow,
4. Ensure ramp queue is within ramp storage capacity,
5. Ensure ramp flow rate is above a lower limit, and
6. Balance ramp queue demands at the entrance ramps.

The first constraint ensures that demand for service along the freeway does not exceed segment (link) capacity. Bottleneck capacity values are chosen to provide the minimum acceptable service quality. The second constraint imposes a limit on entrance ramp capacity, which is dependent on ramp type and flow in the merging lane. The third constraint, which specifies the freeway input flow rate where the control section begins, represents a physical constraint whereby vehicles entering this point must be served. The fourth and fifth constraints ensure that optimum entrance ramp flow is within reasonable storage capacity. The last constraint will provide user equity and balance considerations among different entrance ramps over the entire control period.

To investigate the various freeway system control functions and resulting benefits, four different linear programming (LP) system model formulations were examined:

1. Base Model, or Model 0
2. Dynamic Model 1 with ramp queue diversion of demand,
3. Dynamic Model 2 with maximum queue constraint, and
4. Dynamic Model 3 or model without ramp queue diversion.

Base Model

The Base Model uses the sequential solution approach as developed by Messer of TTI in 1969 (*23*). It maximizes freeway system input flow for individual time slices sequentially subject to: freeway mainline capacity, or level of service constraints, and ramp metering and merging capacity limits. This is done to make sure that minimum and maximum flow boundary limits are satisfied. On-ramp demand is updated using the prior queue carryover. Ramp queues remaining at one time slice simply become a part of the demand in the next time slice. This model cannot determine the optimal queue carry-over between time slices, which would fully utilize the available ramp entrance capacity.

Dynamic Models

The "Dynamic Model" solves ramp metering system optimization problems while also accounting for all ramp spillovers of demand between all time slices. This model balances input and output flows and provides a formal optimization mechanism. Beginning with the basic problem formulation, three different advanced models were proposed and tested on a segment of the Southwest Freeway in Houston, as shown in Figure 12.

$$\text{Maximize: } P = \sum_{k=1}^{W} \sum_{i=1}^{N} X_i(k) \quad \text{vehicles per time, vpt} \tag{15}$$

for all freeway system inputs, i=1 ... N, and time slices, k=1 ... W, at H time slices per time.

Subject to the following constraints:

$$\sum_{i=1}^{N} A_{ij}(k)\, X_i(k) \leq B_j(k) \quad \text{for all freeway sections } j=1 \ldots L \tag{16}$$

$$p_m \sum_{i=1}^{N-1} A_{im}(k)\, X_i(k) + E_m\, X_m(k) \leq C_m \qquad \text{for all m, k} \tag{17}$$

$$R_m(k) = T_m(k) - X_m(k) \qquad \text{for all m, k} \tag{18}$$

$$Q_m(k) = (1/H)\,(1-d_m)\, R_m(k) \qquad \text{for all m, k} \tag{19}$$

$$T_m(k+1) = D_m(k+1) + H\, Q_m(k) \qquad \text{for all m, k} \tag{20}$$

$$Q_m(k) \leq U_{m,max} \qquad \text{for some m} \tag{21}$$

$$X_{m,min} \leq X_m(k) \leq T_m(k) \text{ and } X_{m,max} \qquad \text{for all m, k} \tag{22}$$

where

- i,m - input i; metered ramps (m) start at i = 2 (i = 1, 2, ..., N)
- j - freeway section (j = 1, 2, ..., L)
- k - time slice (k = 1, 2, ..., W)
- $X_i(k)$ - input flow at input i in time slice k
- $A_{ij}(k)$ - proportion of vehicles entering at input i which pass through freeway section j in time slice k
- $B_j(k)$ - capacity of freeway section j in time slice k
- $D_i(k)$ - base demand at input i in time slice k
- $T_m(k)$ - total ramp demand $D_m(k) + Q_m(k-1)$ at ramp m in time slice k

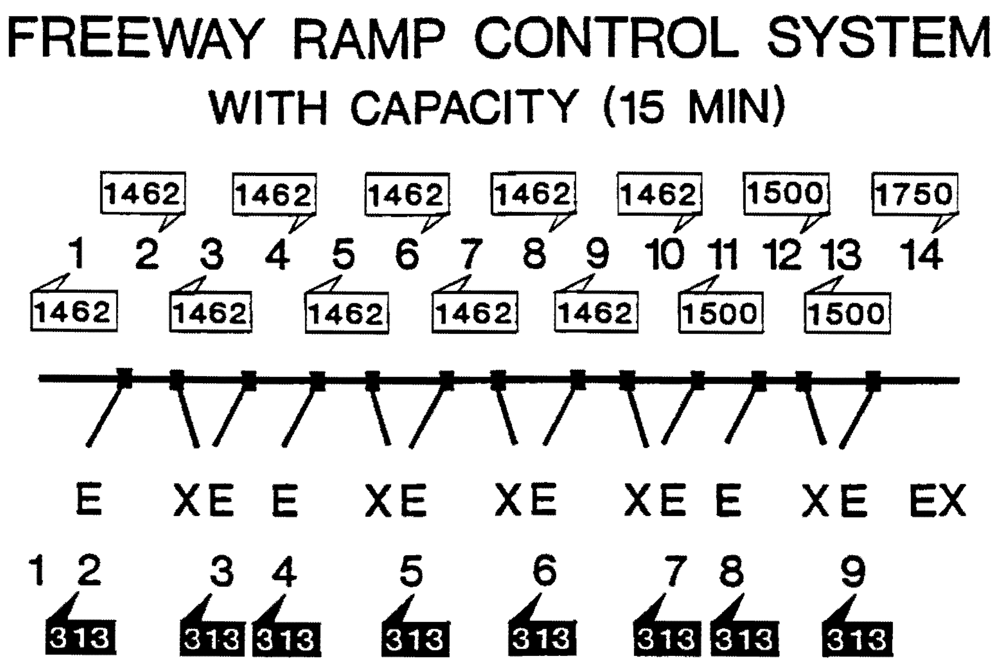


Figure 12. Southwest Freeway Test Problem.

$X_{m,min}$	-	minimum metering rate for metered ramp m
$X_{m,max}$	-	maximum metering rate for metered ramp m
U_m	-	queue storage capacity at ramp m
E_m, C_m	-	equivalency, capacity, based on the type of on-ramp m
p_m	-	proportion of mainline vehicles in merge lane at ramp m
$Q_m(k)$	-	queue length at on-ramp m at the end of time slice k the number of vehicles transferred to time slice k+1
d_m	-	fraction of potentially delayed vehicles diverting from metered ramp m
H	-	number of time slices per unit time t, H = 1, 2, 3, ...

Model 1

Model 1 will maximize system input identical to the basic model. It is assumed that no vehicles waiting at on-ramps are diverted and are subject to ramp merging conditions, i.e., Equations 13-16 above are not active. From the Base Model formulation, the number of trips input to the entrance ramp and output at the exit ramp are initially assumed for calculating the origin-destination matrix. The major drawback of Model 1 is its inability to handle ramp queue overflow when traffic demands exceed on-ramp storage capacity during heavily congested conditions.

Model 2

To provide improved system evaluation, the queue length constraints were added in Model 2, i.e., Equation 20 above is active. This will generally prevent queue overflow back onto freeway interchanges and surface streets.

Model 3

The primary control objective of Model 3 is the same as Model 2 except for the added consideration of ramp vehicle diversion of queues. When long queues are formed at on-ramps, some vehicles will join the back of the ramp queue while other vehicles will divert to other downstream facilities. To simplify the ramp control problem, it is assumed that the total number of vehicles that will divert from the specific freeway on-ramp per unit time is proportional to the potential on-ramp queue length. In addition, the greater the queueing delay, the higher the diversion factor, $D_{m,}$ should be estimated.

The estimated diversion factors and demand volumes should coincide with the actual before/after ramp operating conditions. Differences in diversion activity would usually be due to different ramp metering strategies and to changes that might occur in the traffic operating conditions within the nearby freeway corridor. Care should be taken to distinguish modest changes that might occur in ramp demand or diversion during nominal ramp metering from the dramatic changes in traffic patterns that might occur in ramp demand and diversion when unmetered ramps are initially metered.

MODEL ANALYSIS

Three forms of the Dynamic Model were tested for freeway on-ramp control. The optimal solutions of the dynamic models were compared to the FREQ Model (28). FREQ10 is a state-of-the-art traffic model that analyzes freeway control strategies. FREQ10 uses the sequential solution approach wherein the solution is achieved for each time slice independently and the remaining queues are then forwarded to the next time slice.

Test Case

A section of US 59 "Southwest Freeway," located in the southwest Houston area with 8 on-ramps and 5 off-ramps was selected, as shown in Figure 12. The Dynamic Model formulations for freeway on-ramp control was tested using Southwest Freeway historical volume count data.

The traffic volumes from a 1-hour control period or 4 consecutive 15-minute time slices was used in the study analysis. Different traffic volumes are used as input for each time slice.

Experimental Design

In this comparison study, six freeway ramp control plans were prepared in order to examine the effects of different origin-destination estimations, queue management functions, and possible on-ramp traffic diversion.

Plan 1

Plan 1 uses the same formulation as FREQ10. The major differences are two-fold. First, FREQ10 uses the sequential solution approach. By solving the optimization problem for individual time slices sequentially, FREQ10 uses a different O-D matrix in each analysis period.

Plan 2

Plan 2 was generated with Model 2 using the dynamic approach. An O-D synthesis algorithm developed by Messer of the Texas Transportation Institute was used in developing this plan (23). Model 2 can be considered as a special case of Model 3. When no ramp diversion activity occurs in Model 2, the diversion rate equals zero, i.e., $d = 0.0$.

Plan 3

Plan 3 is Model 2 using the same O-D matrices as in FREQ10.

Plan 4

Plan 4 is a variation of Model 2 with the additional entrance ramp queue constraints. Queue length constraints were engaged for critical on-ramps to provide a variable control strategy at different time periods.

Plan 5

Plan 5 is Model 3 with the diversion factor d = 0.2. That is, a d factor of 0.2 represents that a total of 20 percent of queued vehicles will be diverted.

Plan 6

Plan 6 is Model 3 with d = 0.5. That is, a d factor of 0.5 represents that a total of 50 percent of queued vehicles will be diverted. Each ramp has a minimum and maximum metering rate of 180 vph and 900 vph, respectively.

RESULTS

Table 3 presents the synthetic Origin-Destination (O-D) matrices from Time Slice 1 as estimated from both the Basic Model and FREQ10. Compared to the Basic Model and some field observations, the FREQ10 model tends to overestimate the fractions of the O-D distribution for short trips and underestimate them for long trips.

Table 4 summarizes the optimal solutions produced by FREQ10 (Plan 1) and Model 2 (Plans 2, 3, and 4). As indicated, Plan 2 produces much less input and longer queues than FREQ10 as generated using Model 2 since these two models use different O-D matrices. Model 2 uses one additional constraint to account for the ramp overflow queue carry-over as the traffic demand to be considered in the dynamic approach, while FREQ10 uses the sequential approach used by FREQ. Plan 3 was generated using the modified Model 2 with the same constraints and O-D matrices as the FREQ model. The only difference between FREQ and Model 3 is that FREQ uses the sequential solution approach and Model 3 uses the simultaneous solution approach. It is obvious that the simultaneous modelling approach is superior to the sequential approach. Comparing Plan 1 to Plan 3, the benefits from using the simultaneous solution approach can be demonstrated. As expected, the simultaneous solution approach provided 12% more ramp input volume and a 20% queue length increase. Plan 4 was developed using the modified Model 2 with additional queue length constraints. As shown in Table 4, Plan 4 produced slightly decreased input and increased queues compared to Plan 3, with better performance than FREQ10 (Plan 1) because Plan 4 significantly increased input and decreased queues.

Figure 13 illustrates the queue profiles as produced by both the sequential approach (Plan 1) and dynamic approaches (Plan 3 and Plan 4) at the end of Time Slice 4.

Table 3. Synthetic O-D Matrices from the Base Model and FREQ10 Model.

	To Off-Ramp					
From On-Ramp	1	2	3	4	5	6
1	1.42/.241 *	.089/.105	.038/.038	.038/.038	.022/.020	.670/.558
2	.073/.241	.076/.103	.036/.039	.040/.039	0.024/.022	.751/.556
3		.045/.135	.032/.050	.040/.050	0.26/.028	.858/.738
4		.028/.138	.028/.051	.038/.049	.026/.027	.881/.734
5			.014/.060	.031/.056	.025/.032	.931/.852
6				.008/.061	.019/.035	.973/.904
7					.013/.033	.987/.967
8					.007/.032	.993/.968
9						1/1

* Note - Each cell represents the values of the Base Model / FREQ10.

Table 4. Input and Queue Spillover of FREQ10 and Dynamic Models.

Total On-Ramp Input Volume		% Difference	Total On-Ramp Queues	% Difference
FREQ10 MODEL				
Plan 1	4,277	baseline	1,323	baseline
DYNAMIC MODELS				
Plan 2	3,487	-18.5%	2,352	+77.8%
Plan 3	4,805	+12.3%	1,035	-21.8%
Plan 4	4,791	+12.0%	1,049	-20.7%

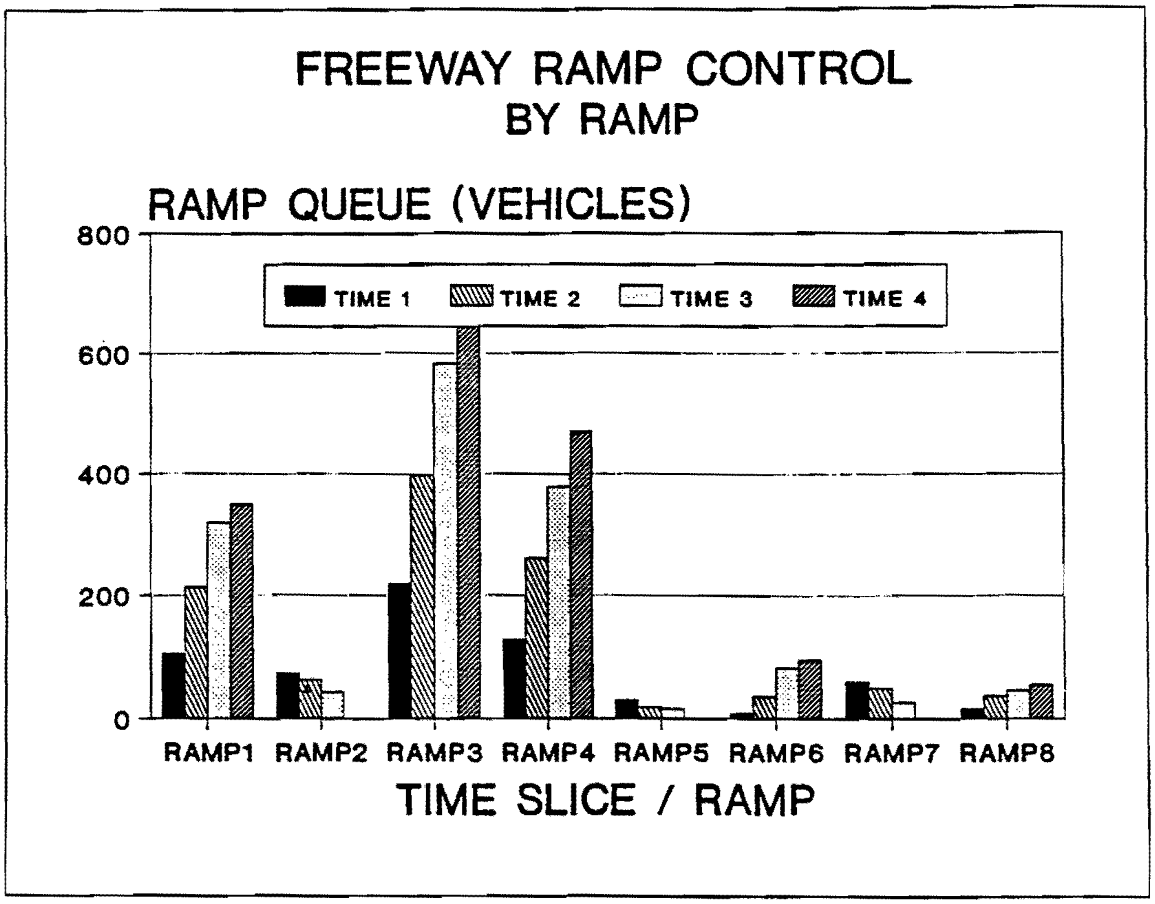


Figure 13. Examples of Queue Lengths at Entrance Ramps by Time Slice.

Table 5 summarizes optimal solutions produced by Model 2 (Plan 2) and Model 3 (Plan 5 and Plan 6). Model 2 assumes no vehicle diversion from ramp queues and Model 3 allows vehicle diversion. If d = 0.0, Model 3 becomes Model 2. Figure 14 illustrates the queue profiles of dynamic approach with various diversion percentages (Plans 2, 5, and 6) at the end of Time Slice 4. As indicated, a large difference is observed among the distribution of ramp queue lengths even though the same system inputs were provided. Furthermore, a larger diversion d factor value results in smaller queue lengths. When the d factor becomes larger, residual queue length decreases. Realistically, the diversion percentages depend on both traffic flow levels and the availability of alternative routes for each entrance ramp throughout the day.

Table 5. Dynamic Models with Different Ramp Diversion Percentages.

Total On-Ramp Input Volume		% Difference	Total On-Ramp Queues	% Difference
DYNAMIC MODELS				
d = .0	3,487	baseline	2,352	baseline
d = .2	3,486	0 %	1,314	-44.1 %
d = .5	3,483	-.1 %	459	-80.5 %

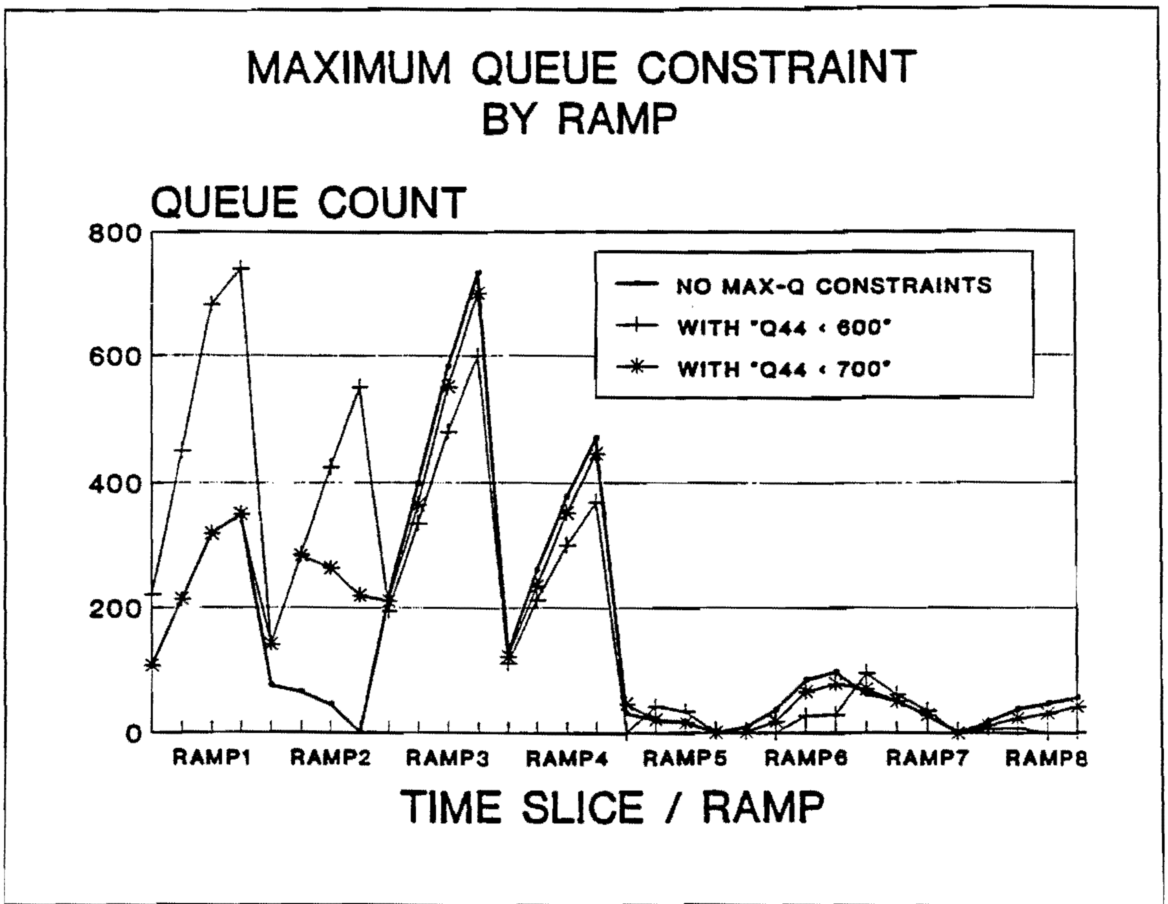


Figure 14. Examples of Queue Profiles as Related to Ramp Storage.

4.0 RECOMMENDED SYSTEM RAMP METERING ALGORITHMS

INTRODUCTION

The real-time ramp metering algorithms recommended for implementation in Texas can be best explained in the context of TxDOT's Freeway Traffic Management (FTM) system under development. The FTM system is one of four systems that have been designed to serve TxDOT's overall traffic management needs. The other three types of traffic management systems that are being developed include: (1) Arterial Traffic Management, (2) Signal Coordination and (3) High Occupancy Vehicle Lane systems.

Each of TxDOT's traffic systems will consist of field devices, Local Control Units (LCU), System Control Units (SCU), a Manager, and possibly a central control center.

1. Field Devices:
   - Located within TxDOT's right-of-way (on the shoulder or over the roadway)
   - Used for surveillance, communication and control techniques

2. Local Control Unit (LCU):
   - Located in field cabinets
   - Provides data collection, device control, and pattern implementation

3. System Control Unit (SCU):
   - Located on the traffic project (alongside the roadway, etc.), usually in a small building referred to as the Satellite building or Satellite location
   - Provides system processing and timing, and communications coordination

4. Manager:
   - Usually located at either a central control facility, or an operator or engineer's office
   - For a manned Satellite operation, located at a Satellite building
   - Serves as the operator interface
   - Provides overall system analysis, graphics, and maintenance reports

5. Control Center:
   - May consist of a single manager located at a district office or a large sophisticated system housed in a separate building
   - May house TxDOT personnel or multi-agencies
   - May expand to include high resolution monitors, overhead projectors, operator workstations, TV switching systems and graphic displays

Application software for each traffic system will be different, but the same basic LCU, SCU, and Manager hardware will be used on all systems. Figure 15 gives FTM system details.

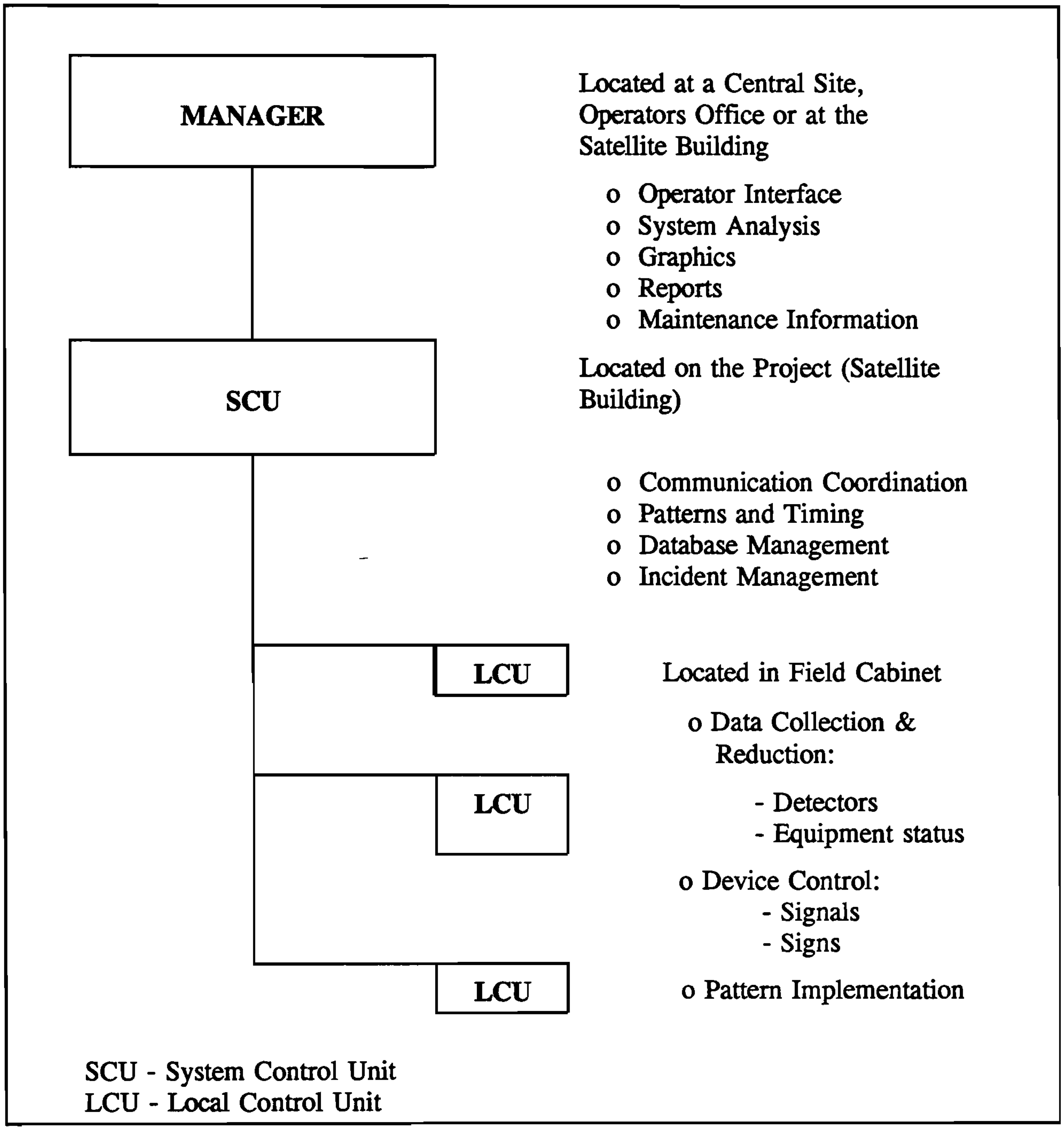


Figure 15. FTM Traffic System Components.

SYSTEM REAL-TIME CONTROL MODEL

The recommended algorithms for real-time ramp metering will be described in TTI's vision of an advanced multilevel control structure ideally suited to TxDOT's hierarchical distributed processing system control framework shown in Figure 15. The envisioned multilevel framework is based on earlier work by TTI(*12*, *23*) on ramp metering in Houston and Dallas, and is considered current with the latest control systems technology (*34*, *35*).

Multilevel Control Structure

Figure 16 presents a block diagram of the multilevel ramp metering control system recommended for system-based ramp metering in Texas. The recommended algorithms and their required data requirements will be presented by each level of hierarchy given in Figure 16. The recommended algorithms for providing real-time adaptive system control are shown as the shading region in Figure 16. This set of algorithms is collectively called "RAMBO" for Real-time Adaptive Metering Bottleneck Optimization.

The hierarchical system design provides several features which are particularly advantageous to working with large, complex, dynamic systems having a myriad of input and control variables together with a large spatial coverage. The large area combined with a rather slow moving flow of traffic (certainly not traveling at the speed of light as in distributed telecommunication systems) permits the freeway control system design to be "loosely coupled" and control response times to be modest yet "real-time" with respect to the speed of the process. Figure 16 implies that data interchange occurs between control levels. Higher levels have slower data exchange rates and are the supervisors of the lower levels.

The hierarchical system provides the mechanism to focus on the basic problem first at Level 0. This point also covers the smallest area but monitors/controls at the fastest rate. Real-time in Level 0 is very fast, often fractions of a second. As control moves up the control levels, the traffic control area increases rapidly together with a corresponding reduction in the speed of response needed for "real-time" control. At Level 4, a full freeway corridor is being considered and the response time may be from say six (6) minutes to daily activities. Level 5 may provide inputs that come from monthly Traffic Management Team meetings that have decided that some ramps need to be reorganized because of construction in the corridor. The selection of these response times have to balance the efficiency gained by providing more rapid response as compared to the workload and cost placed on the data links and processors. Level 5 would also deal with resolving these types of high-level issues.

Database Design

The database requirements for each control level will be described at the end of each section. These data are all those needed to operate RAMBO, the systems-based ramp metering controller shown in Figure 17. These include fixed and parametric data, together with short-term (TFS), intermediate-term (TFI), and long-term variable data (TFL).

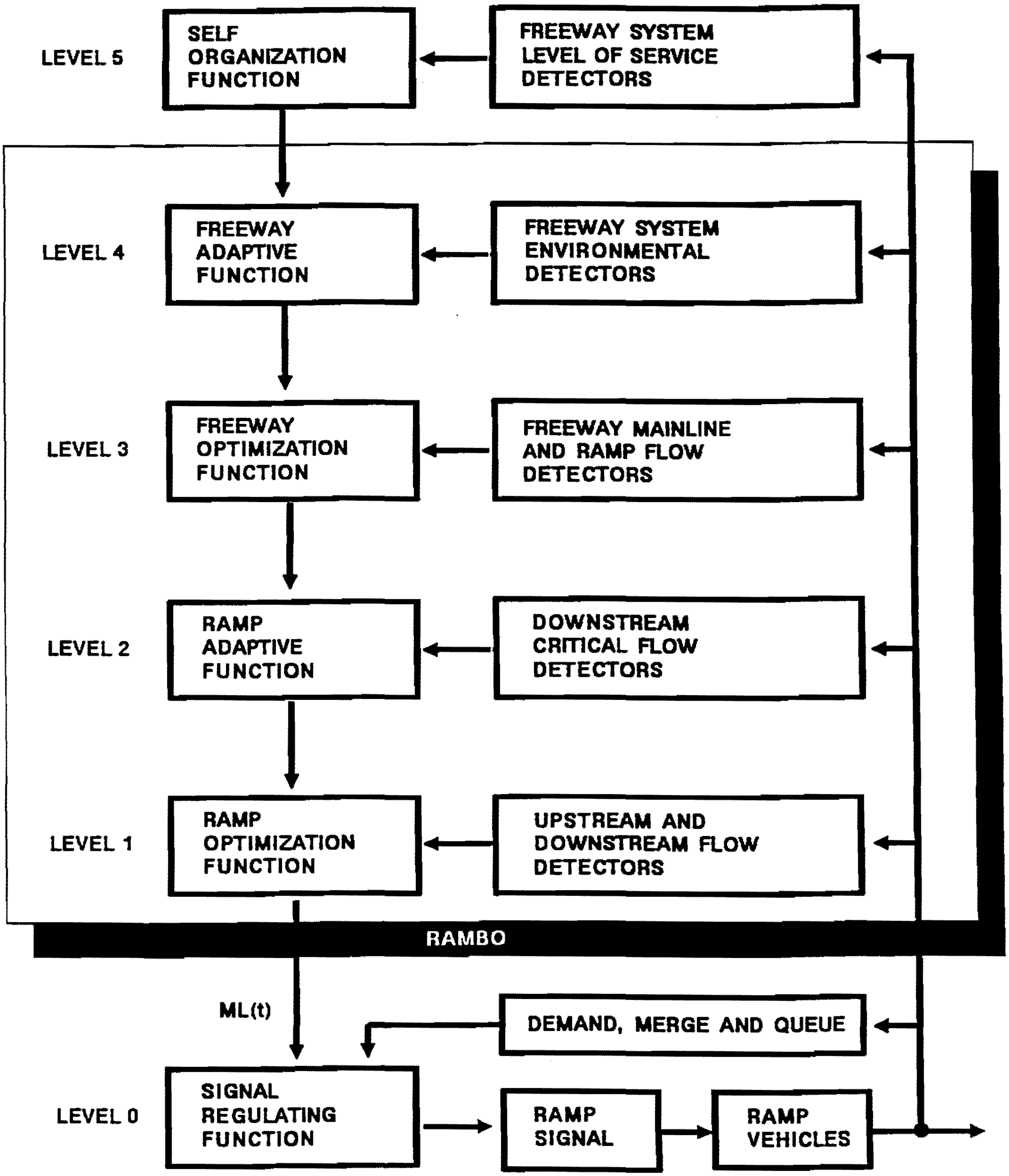


Figure 16. Multilevel Hierarchical System Metering.

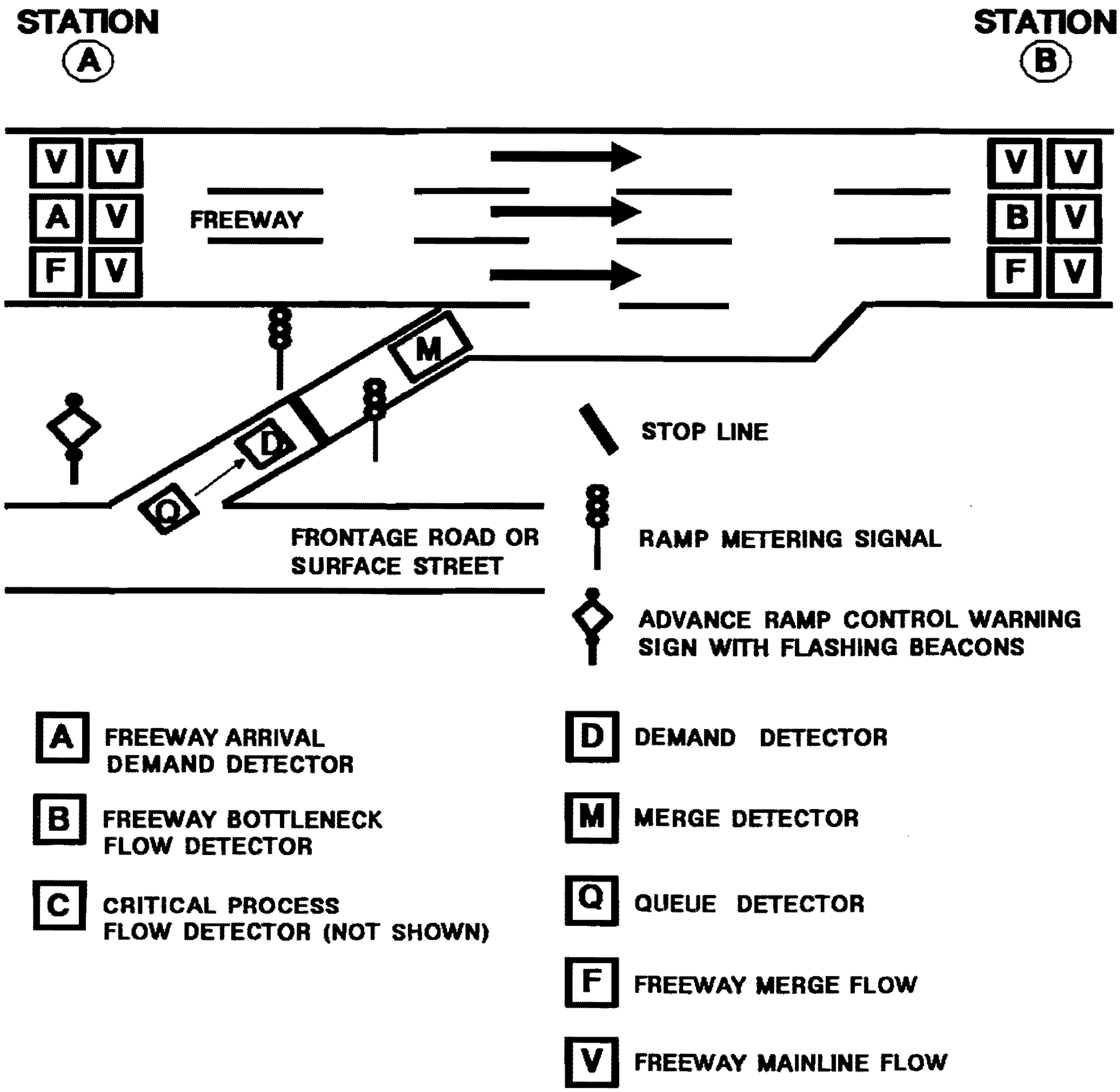


Figure 17. Typical Freeway Ramp Control Subsystem.

The database manager should provide for both synchronous and asynchronous data interchange (uploading and downloading) between control levels as required. The manager should also routinely provide for data storage buffers to hold all control variables and parameters that might be changed at any time by other levels. This data transfer should not disturb the equivalent active control variable at the moment of arrival to the control level, but rather the data should be stored in a "buffer" variable as follows: DATAB = DATA(t) arriving at time t from some other control level; then DATA = DATAB at a later time.

Time frames TF are used to collect traffic data and for which to derive average traffic flow values for the durations. The following definitions of data base time frames should apply for real-time adaptive system control:

Short-term: TFS = i = 20 seconds;
Intermediate-term: TFI = j = 1 minute; and
Long-term: TFL = k = 6 minutes.

Short-term data should consist of either a status indicator (on, off, a metering level (ml = A,B,C,D,E,F,G,H), or a data value for a variable in real-time, e.g., occupancy(TFS).

Intermediate-term data should be derived at the most convenient location, presumably at the SCU, and should be routinely calculated using exponential smoothing techniques. The following algorithm is recommended for variable data VI(j) for TFI "j":

$$VI(j) = VI(j-1) + a\{s3\ vs(i) - VI(j-1)\} \quad (23)$$

where

s3 vs(i) = sum of three (3) most recent short-term TFS observations
a = coefficient of exponential smoothing for TFI data, e.g., a = 0.3

Long-term data should be derived at the SCU. These data should be derived from data fusion of a variety of data bases using exponential smoothing. The following algorithm is recommended for forecasting all TFL variable data VL(k) for the forthcoming time frame (slice) "k". The present error, measured during the previous time slice (k-1), is

$$e(k-1) = N(k-1) - S(k-1) \quad (24)$$

which may then be used, if e(k-1) is not considered an outlier, to develop an exponentially smoothed average error value, say with a smoothing coefficient of b = 0.3, to be

$$E(k-1) = E(k-2) + b\ \{e(k-1) - E(k-2)\} \quad (25)$$

The forecasted error for the next time slice k, *e*(k), is more stably estimated by linearly projecting the exponential average errors by

$$e(k) = E(k-1) + \{E(k-1) - E(k-2)\} = 2\ E(k-1) - E(k-2) \quad (26)$$

such that the forecasted value of VL(k), given a stored value S(k), would be

$$\mathit{VL}(k) = S(k) + e(k) \tag{27}$$

The algorithm then updates S(k-1) with the error correction term e(k-1) such that the average value continues to be effectively averaged over several working days, D, say D = 20 working days per month, by selecting the exponentially smoothing coefficient "b" here to be b = 2 / (1 + D) such that b = 0.095 if D = 20 days. In general, the updating model is

$$S(k-1) = S(k-1) + \{2 / (1 + D)\}\ e(k-1) \tag{28}$$

where

e(k-1)	=	counted error between the current "volume" count during the time slice k-1 just ended and the stored "expected" volume for the time slice k-1
N(k-1)	=	observed volume count for the time slice k-1
S(k-1)	=	volume (as originally stored, then later updated) for time slice k-1
E(k-1)	=	exponentially smoothed current "error" correction for time slice k-1
e(k)	=	predicted error for the next time slice k
VL(k)	=	predicted control "volume" for next time slice (k)
S(k)	=	currently stored average volume for next time slice (k)
b	=	exponential smoothing coefficient for TFL data
	=	2/(1+N) where N is the number of samples effectively averaged, a value of 0.3 is recommended except as noted otherwise

The data base algorithm provides an efficient method for forecasting future variables and also for slowly updating the system data without the need to store a long data chain used with moving averages. Data storage arrays should be provided for each (r) TFL control variable used for control for each control interval "k". This (r x k) matrix should, as a minimum, be provided for the typical weekday. Daily values should be considered.

The above algorithm can be used to forecast entrance or exit ramp demand as shown in Table 6 in a format suited for structured programming. Assume ramp volume counts have been made for time slices k-2 and k-1, with a forecast desired for the future time slice k. The algorithm begins by measuring the error observed between the ramp count N_r and the stored value S_r for ramp r. Then a critical error (CE) threshold is estimated based on the Poisson distribution to identify probable freeway operating conditions (congested, clearing, or normal) based solely on ramp flow. If the error is greater than CE, then a major incident is probably occurring so that the forecast ramp demand is estimated to be the most recent measured flow. If the ramp count is very low with the error below negative CE, freeway congestion has probably formed and the ramp demand is at least the nominal demand (S_r) plus the stored demand during the past time slice, k-1. If the two abnormal conditions can be rejected, then nominal traffic conditions are assumed, and the exponential smoothing algorithm noted above is used to forecast ramp demand. A volume-to-demand conversion process may be added. Figure 18 illustrates the input and output of the model.

Table 6. Ramp Demand Forecasting Algorithm.

Algorithm V_r

N_r (k-1) (the just completed volume count during (k-1) for ramp r)

$e_r (k-1) = N_r (k-1) - S_r (k-1)$ (the resulting comparison error)

$CE = T_{fac} * \sqrt{S_r (k-1)}$ (the dynamic control limit for Poisson data)

IF (e_r (k-1) .GT. CE) THEN (the error exceeds upper (surge) limit)
  $V_r (k, k+1, k+2, \ldots) = N_r (k-1)$
  $E_r (k-1) = E_r (k-2)$

ELSE IF (e_r (k-1) .LT. - CE) THEN (error exceeds lower (blockage) limit)
  $V_r (k, k+1, k+2, \ldots) = S_r (k, k+1, k+2, \ldots) - E_r (k-2)$
  $E_r (k-1) = E_r (k-2)$

ELSE (acceptable data for forecasting)
  $E_r (k-1) = E_r (k-2) + b \{e_r (k-1) - E_r (k-2)\}$
  $e_r (k) = 2 E_r (k-1) - E_r (k-2)$
  $V_r (k, k+1, k+2, \ldots) = S_r (k, k+1, k+2, \ldots) + e_r (k)$
  $S_r (k-1) = S_r (k-1) + c\, e_r (k-1)$

END IF
  $V_r (k, k+1, k+2, \ldots) = CF \times V_r (k, k+1, k+2, \ldots)$ (demand estimation)

r = ramp r
k = next time slice
T_{fac} = "t" statistic for confidence level, 2.0
b = 2/(1+T) , T = 5, b = 0.3
c = 2/(1+D), D = 20, c = 0.1
CF = demand calibration factor, 1.1

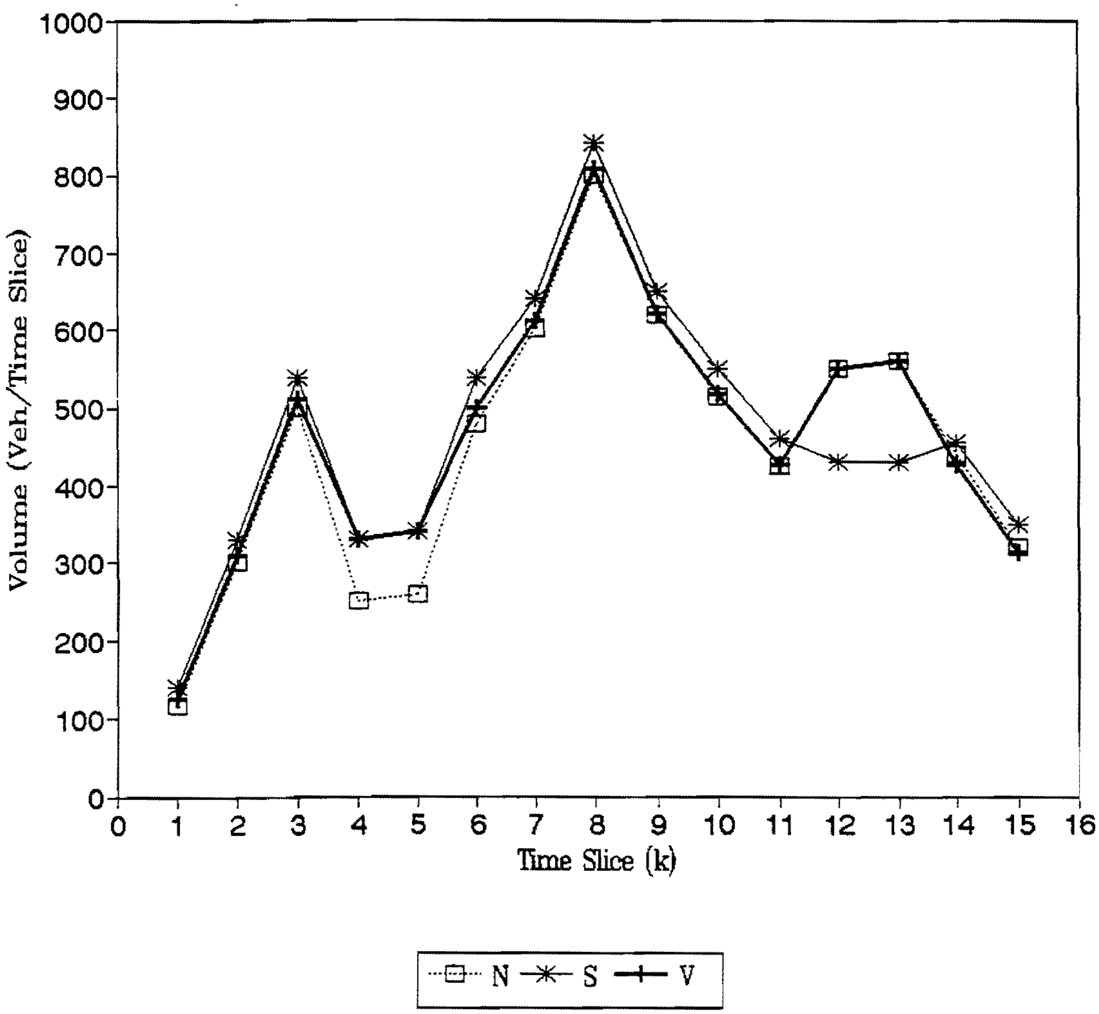


Figure 18. Application of the Ramp Metering Forecasting Algorithm.

Level 0 - Regulator Control

This lowest level of control consists of the "regulator" functions of ramp metering that are provided by the local ramp meter. These functions basically drive the traffic signals in response to demand input given the fixed and variable timing parameters of the signal indications which are derived from preselected nominal metering rates for a given metering level ml, MR(ml). The TxDOT "system" regulator produces metering rates from Table 7.

Table 7. TxDOT System Metering Rate to Metering Level Mapping Function.

Metering Levels, ML(ml)							
A *	B	C	D	E	F	G	H
Red:	2.0	2.5	3.0	4.0	5.5	8.0	13.0
s: 1	2	2	2	2	2	2	3
MML:*	B	B	C	D	D	E	F
* Denotes a Non-Metered Condition							

Seconds of Red Time is Keypad Input Only at Local Ramp Meter.

Two new rows have been added by TTI to the TxDOT regulator table. The local traffic flow "state" variable "s" has been added to define three traffic conditions: s = 1 is the free flow state; s = 2 is the metered flow region; and s = 3 is the congested stop-and-go traffic condition. Minimum metering levels, MML, to be described later, define the minimum reductions permitted during queue override from initial metering levels.

The primary output of RAMBO must be the selected "metering level" as a function of time, or ML(t) = (ml: A,B,C,D,E,F,G,H) from Table 7. As a general policy P(cl.0), control level (cl) zero (cl.0) would select metering rate (level) ML(B), if operating, unless otherwise specified by a higher level policy P(cl > 0). This statement implies

IF .NOT. ML(A) THEN (IF P(cl.0) THEN ML(t) = ML(B)) ELSE P(cl. > 0)

The prior specification simply says, given the ramp is being metered, that we initially desire to meter the ramp as fast as practically possible to smooth out short-term ramp demand variability merging onto the freeway and to minimize ramp delay, unless other higher level priorities are imposed. This is a simple but powerful control function.

The higher priority control levels of RAMBO try to produce the most desired system-based control conditions, primarily by parameter manipulation of lower control functions. This implies that metering levels (rates) derived therefrom, ML(t.cl), have higher priorities.

$$ML(t) <= MR(t): P\{ MR(0) < MR(1) < MR(2) < MR(3) < MR(4) < MR(5) \}$$

Meter rates (MR) are generated from the average cycle time, Ca, operating at the entrance ramp during the time frame of concern. Ca is the sum of the local ramp green interval G, yellow change interval Y, and average red time displayed, Ra; that is,

$$Ca = G + Y + Ra \qquad \text{(in sec)} \qquad (29)$$

Ra will be longer than the setpoint red given in Table 7 for ML(t) because the controller sometimes dwells in red due to local merging operation override detector inputs or due to a lack of ramp demand, which will extend the red time Ra on the average.

The fixed values that drive the cycle time include: G, the preset green interval for the ramp meter (nominally 1.5 sec); Y, the duration of the yellow change interval (nominally 1.5 sec); and Ra, the "running average" red time displayed at the ramp as estimated for the current time frame. The value of Ra must always exceed some minimum red time, about 1.5 seconds. Other durations of the G:Y:R interval parameters have been used, such as 1.5, 2.0 and 1.0 sec, respectively. The average red time, Ra, must be accurately estimated by the data base manager if reliable metering rates are to be produced by the control system.

The following algorithm should meter when the queue override is "on" at ramp "m" when originally metering at level "mlo." Reduce the metering level (ml) one level per TFS

$$MML(mlo) <= ML(ml) = ML(ml) - 1 \text{ ml}$$

but where the current ML(ml) is not less than the minimum metering level MML(mlo) based on the original metering level (mlo) when the queue override first came on. The minimum metering level, MML(mlo), (or rate) may be changed by any higher priority control level. The operation of this algorithm is as follows. When the local queue detector is calling, control should reduce the metering level one step each 20 seconds (an "i" time frame) (increase the metering rate) until the queue is cleared, or until the minimum metering level (maximum metering rate) is reached. No reduction may occur in some cases.

Database Requirements for Level 0, TFS

The various types of data and the nature of the data required at the SCU shown in Figure 15 to satisfy Level 0 control requirements are given in Table 8 for every entrance ramp "m" that is being metered under system control. Some of these data are the fixed timing parameters of the individual ramp meters. Others are ramp and detector pointers. Refer to Figure 17 for the definitions of detectors A, B, C, D, M, and Q. System ramp control is provided by the set <ML,MR, and s> from higher levels of control. Recall the need for buffer arrays of data that are not shown in Table 8.

Table 8. Database Requirements For Level 0.

Type	Variable	Function
Fixed		
	NR	Ramp Number
	NC	Class Number
	NS	Queue Storage
	NA	Detector A
	NB	Detector B
	NC	Detector C
	ND	Detector D
	NM	Detector M
	NQ	Detector Q
	NU	Ramp Upstream
	G	Green Time
	Y	Yellow Time
	RMIN	Red Minimum
	ML(ml)	Table 7
	R(ml)	Red Time, Table 7
	TQUE	Queue Threshold
	TMER	Merge Threshold
Variable		
TFS	MS	Merge Detector On
	QS	Queue Detector On
		(Cycle Synchronized)
TFI	SN	Cycle Count
	SR	Red Sum
	SC	Cycle Sum
	SD	Demand Sum
	SM	Merge Sum
	SQ	Queue Sum
	S	Traffic Flow State (1,2,3)
	ML	Metering Level (1-8)
	MR	Metering Rate
TFL	TES	Earliest Start
	TLF	Latest Finish
	MML(ml)	Minimum Metering Level

Level 0 - Operations

Several fundamental traffic operations and design controls should be noted when designing and operating ramp metering systems. These basic Level 0 aspects pertain to the physical kinematic parameters, human factors issues, and system impacts of ramp metering.

Vehicle Kinematics. Two vehicular kinematic operations are important operational parameters of ramp metering. These include (1) the start-and-stopping operations as vehicles move up through the queue as the leading vehicles are being processed one-at-a-time by the ramp meter, and (2) the acceleration characteristics of merging vehicles from the ramp signal toward the merge point. Discussion of each of these aspects follows.

Ramp vehicles move up one queueing position at a time in a stop-go-stop manner, much like a "frog jumps" as the ramp meter is processing the lead vehicle in queue once per meter cycle, C, in seconds. A kinematic analysis can be performed to estimate the minimum time vehicles can move up one position, given selected operational assumptions. It is important to keep these basic operational considerations in mind when selecting rapid metering rates. We assume that the cycle time, C, defines a series of operational events that must be performed by a vehicle in queue in a given cycle time:

$$C = T_p + T_r + T_a + T_r + T_d \tag{30}$$

where T_p is the perception time to the green onset of the ramp meter, T_r is the brake-to-gas (accelerator) reaction time, T_a is the time required to accelerate to speed, T_r is the accelerator-to-brake reaction time, and T_d is the time required to decelerate back to a stop over the remaining average queue spacing, L. A mean queue storage spacing is assumed to be L = 20 feet (6.1 m). Nominal response times of 1.0 seconds are assumed for both brake-accelerator transitions, T_r. It is also assumed that the acceleration and deceleration rates at these low speeds and very short travel distances are both 7 fpss (2.1 mpss), a typical value observed at traffic signals. Higher acceleration rates for movement away from ramp signals toward the freeway merge point, however, have been observed to be about 10 fpss (3.0 mpss) in Texas. Substitution of the above criteria and assumptions into the cycle time equation and solving a quadratic equation yields

$$C = T_p + 1.0 + 2 \{ L/a + T_r^2 \}^{0.5} \tag{31}$$

$$C = T_p + 2.97 \text{ sec} \tag{32}$$

Table 9 provides a summary of the available perception times, T_p, for green onset given various cycle times, metering rates, and metering levels previously given in Table 7. A rapid but well paced metering rate would seem to be about 800 vph, providing 1.53 seconds of perception time for the green signal. Metering rates in excess of 800 vph clearly are "pressurized" and may require motorists to anticipate the green onset, which would be expected to promote violations of the ramp's red signal.

Table 9. Perception Times Available for Various Metering Levels.

Metering Level ML	Metering Cycle, sec	Metering Rate, vph	Perception Time Tp, sec
A	-	-	-
B	4.0	900	1.03
C	4.5	800	1.53
D	5.0	720	2.03
E	6.0	600	3.03
F	7.5	480	4.53
G	10.0	360	7.03
H	15.0	240	12.03

This table of available perception times to the green onset of ramp meter was developed based on the following assumptions:

Reaction times = 1.0 sec.
Acceleration and Deceleration rates = 7 fpss (2.1 m/ss)
Storage distance = 20 feet per queueing vehicle (6.1 m)

Knowledge of the travel times from the ramp signal to the merge point are important to the design and operations of efficient ramp metering systems. Not a lot of information has been published on this basic kinematic problem of ramp metering, although some very detailed studies were published by TTI based on research conducted on the Gulf Freeway in Houston in the late 1960's *(19)*. An equation was proposed for estimating the mean travel time along the ramp, TT_r, in seconds since green onset as

$$TT_r = 2.4 + \{ D/5 \}^{0.5} \tag{33}$$

where D is the distance in feet from the ramp meter to the merge point. The above equation was calibrated to numerous ramps' operations. It was found that the effective average acceleration of the vehicle was 10 fpss (3.0 mpss) regardless of local ramp grades. Motorists apparently compensate for modest grades by accelerating more/less rapidly. Table 10 presents estimated travel times to the merge point and operating speed of merging vehicles at the merge point as related to travel distance from the ramp meter.

Table 10. Estimated Travel Time to and Speed at Merge Point As Related to Distance From Ramp Meter to Merge Point.

Distance to Merge		Travel Time to Merge	Speed at Merge	
feet	meters	seconds	mph	km/h
50	15.2	5.56	21.5	34.4
100	30.5	6.87	30.5	48.8
150	45.7	7.88	37.3	59.7
200	70.0	8.72	44.1	70.6
250	76.2	9.47	48.2	77.1
300	91.4	10.15	52.8	84.5

This table was developed from Equation 33 which assumes that the net vehicle acceleration rate along the ramp is 10 fpss (10 mpss)

Queueing Operations. The ramp meter will produce queueing along the ramp behind the signal. The severity of the queueing is related to the arrival flow and metering rate. Two basic queueing parameters should be closely monitored to assess the quality of ramp metering operations and its potential impact on the surrounding traffic system. These two operational parameters are queue length and ramp delay. When the upstream cross street is served by a diamond (ramp) interchange, the queue should not back into the upstream signalized intersection. When the upstream cross street is served by an X (ramp) interchange, then the more immediate queue blockage problem becomes the upstream freeway exit ramp merge point with the frontage road, if braided ramps are not installed to separate the conflicts.

The length of queueing vehicles storing spatially along the ramp should be monitored for three aspects: (1) the fraction of time when a queue is present, (2) the percentage of time when the queue is longer than a given length, probably the queue detector position, and (3) the maximum queue backup observed during the monitoring period. This maximum queue length should then be compared to the maximum queue storage available along the ramp to assess the local storage utilization factor. Ramp vehicles store at about 20 feet (6.1 m) per vehicle so a direct relationship exists between the number of vehicles stored and the expected queue storage distance utilized. Twenty (20) vehicles in queue would require about 400 feet (122 m) of ramp for storage.

A fundamental queueing theory relationship exists between queue length and ramp delay, the two principal operational measures of effectiveness. The time a vehicle takes to process though a queue depends on the length of queue (in vehicles) when the vehicle arrives multiplied by the average metering cycle time during the processing period. More explicitly, the mean queueing delay is

$$RD = Q \times C / 60 \tag{34}$$

where

RD	=	ramp delay for vehicles processing through the queue, minutes,
Q	=	queue length in vehicles for given arrival conditions, and
C	=	metering cycle length in same time units as D, seconds.

An averaging process over a metering time may be applied to calculate mean values of each variable in the above equation. Unit conversions have been applied to express RD, Q, and C in units more commonly used in practice. Other conversions, such as delay in minutes, cycle in seconds, and Q in equivalent feet of storage space, may also be useful for design purposes.

The above queueing relationship (Equation 34) can be plotted to gain valuable insight regarding the important design and operational variables involved. For example, drivers are known to lose their toleration for minimally enforced ramp meters when the ramp delay reaches about 3.0 minutes. Common ramp metering practice is to "kick in" a high metering rate when the "queue" detector becomes activated by the growing queue to address this operational problem, but marginal insight and minimal sensitivity to the variables involved may be gained. Just how high should the queue flow be? Under what set of circumstances? To address this concern, the basic queueing equation is plotted in Figure 19 and some operational aspects will be noted in the following paragraph. More detailed queueing models that address queue length will follow.

Ramp delay for vehicles processing through the meter is plotted against metering rate (MR = 3600/C) in Figure 19. A ramp delay of 3.0 minutes will be experienced when the metering rate is 200 vph and the queue is 10 vehicles (or 200 feet long). If the queue doubles to 20 vehicles, the metering rate will need to rise to 400 vph to keep the ramp delay at 3.0 minutes. The figure clearly shows the delay peril when using minimum metering rates as delay rises exponentially with linear reductions in metering rate. Delay rates being experienced in the ramp queues may quickly exceed tolerable levels. In addition and as described later, queue lengths also rise with reduced metering rates making the actual delay very sensitive to variations in demand at low metering rates. As noted below, ramp delay can be even more sensitive due to temporal correlations that may exist between ramp demand and ramp metering capacity because ramp metering capacity is inversely related to the freeway flow which in turn can be positively correlated with the local ramp demand. This is a common occurrence and operational problem at all-way stop sign operations.

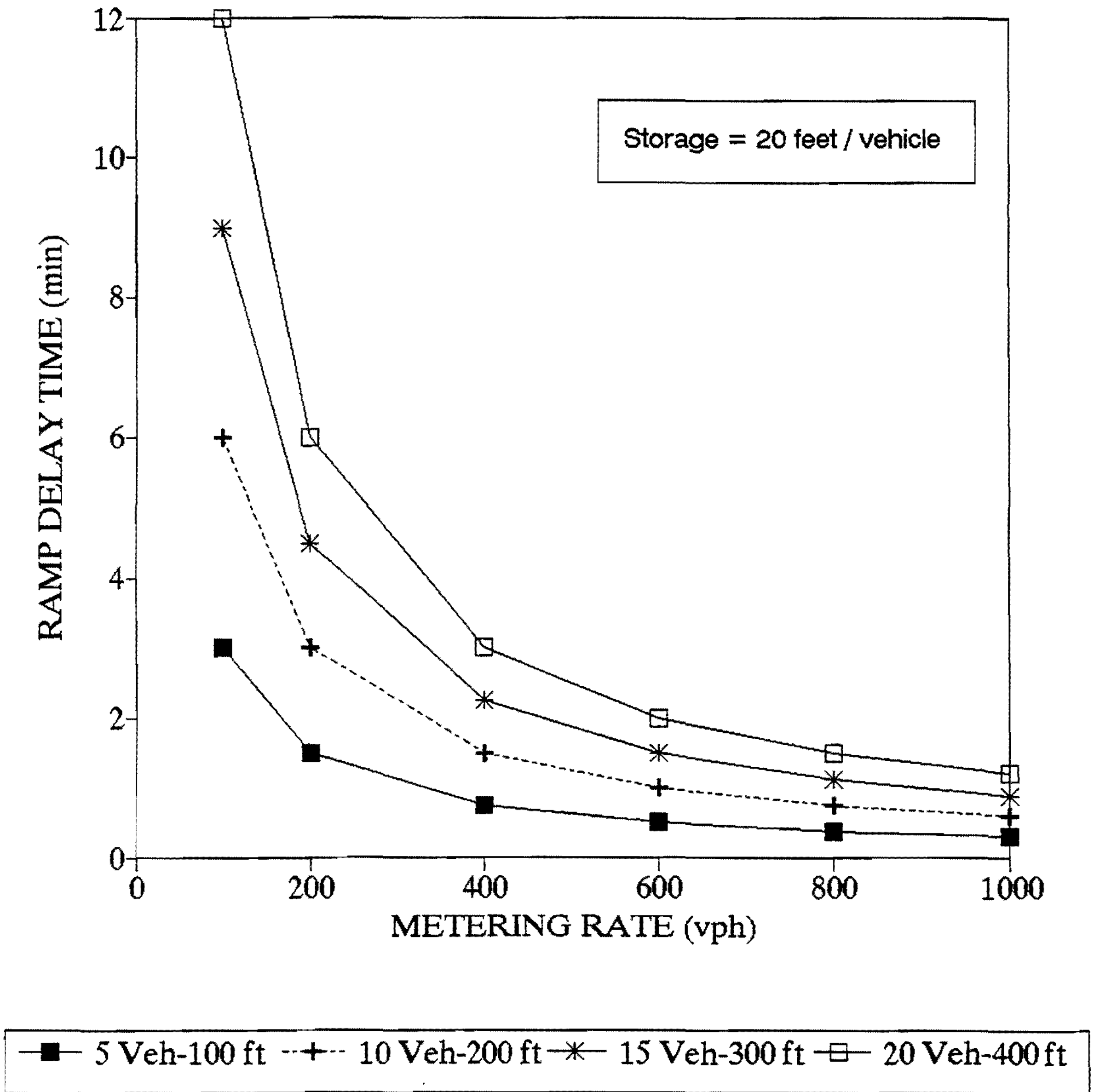


Figure 19. Ramp Delay Expected for a Given Metering Rate and Queue Length.

Queueing Models. Queueing theory provides models for estimating the average ramp queue length and metering delay for given assumptions for the arrival and departure (metering) distributions. Three basic assumptions are made: (1) that the average arrival rate (AR) is less than the average metering rate (MR) during the analysis period (i.e., during undersaturated conditions), (2) that both flows are continuous and undisturbed processes during the period, and (3) that the arrival rate is random (i.e., Poisson). It would appear that these assumptions are practically valid for ramp metering analysis if the time period of averaging is longer than five (5) traffic signal cycles (say 5 minutes) but not too long (say 15 minutes) so that traffic demands have not changed appreciably due to temporal correlations. In any case, the theory provides valuable insight into the general queueing process and performance. With the above assumptions invoked, then it can be shown that the average number of vehicles (RQ) stopped at the ramp signal can be estimated by

$$RQ = c\,AR / (MR - AR) \tag{35}$$

and the average ramp delay RD (in minutes) can be estimated by

$$RD = 60\,c / (MR - AR) \tag{36}$$

where MR is the metering rate (vph) and AR is the arrival rate (vph). Clearly, as the arrival rate (AR) grows and approaches the metering rate (MR) the delay and queue would begin to grow rapidly and go off the scale (approach infinity) as saturated conditions occur. Thus, these models are useful only to saturation ratios (AR/MR) of 0.97 or less. The coefficient "c" in the above equations reflects the degree of randomness or inversely uniformity in the output metering rate. The coefficient is estimated from the equation

$$c = 0.5 + 0.5 / a \tag{37}$$

where

Metering Condition:	Random Metering				Uniform Metering
"a"	1	2	3	. . .	∞
"c"	1.00	0.75	0.67	. . .	0.50

If the metering becomes more uniform for a given metering rate, then it is theoretically possible to reduce the queue length and delay by one-half from that expected with totally random metering. Thus, the pace of metering should be kept rather uniform over a control period. While it is theoretically possible to calculate the operating value of "a", $a = (C/s)^2$ where C is the mean metered cycle and S is the standard deviation of the individual meter cycle times during the control period, observation of the operational variables can provide direct calibration of "c" if desired.

Whereas Figure 19 provides estimates of ramp delay given a ramp queue of "RQ" vehicles, perhaps as measured by local ramp detectors, Figure 20 plots mean ramp delay (RD) versus metering rate (MR) for various arrival rates (AR). As the arrival rate approaches the metering rate, the queue size and delays increase. The operational question at any moment is always, "How much above MR is the local metering capacity?"

The above queueing theory for undersaturated conditions suggests that long queues on the order of 10-20 vehicles occur when the long-term arrival flow rate (AR) is about 95-97 % of the metering rate. Should the arrival demand exceed the metering rate, then much longer queues may form. The size of the extended queue depends on the magnitude and duration of oversaturation. Thus, special attention should be given to operating conditions when the arrival demand exceeds existing metering rate capacity. However, ramp metering capacity is not a constant, or wide-ranging independent control variable, but it depends on the freeway arrival flow at the ramp meter. As often happens, ramp demand and freeway demand may both be temporally correlated, i.e., ramp demand and freeway demand may both be increasing with time. This occurrence would make ramp queue growth very sensitive and potentially explosive as volumes approach peak-hour conditions. A ramp may be operating all right with a moderate queue, but it may then suddenly begin to suffer severe delay with a 5 % increase in demand if the freeway flow also increased 5 % during the same time, thereby producing a 10 % change in v/c (saturation) ratio.

Queue Detector Location. The location of the queue detector should provide the minimum space needed to store at least one signal phase of traffic, given that average ramp demand-to-capacity ratios less than 1.0, say 0.97, can be provided overall. In Texas, higher ramp demands will have to divert along the frontage road as there appears to be no practical way to store unlimited demand on the short ramps commonly found downstream of most diamond interchanges in large Texas cities. These criteria produce a queue overflow distance to the upstream intersection of

$$LQ = 20\ G\ (SI - MR)\ /\ 3600 \tag{38}$$

where

LQ	=	distance from queue detector to cross street, feet
G	=	duration of green signal phase at intersection, say 30 sec
SI	=	saturation flow rate at intersection, say 1900 vphgpl, and
MR	=	metering rate during queue overflow, say 900 vph

Selecting a practical set of values for the above input variables would provide some insight and guidance for design and operations personnel. We have already assumed that the storage space per vehicle is 20 feet (6.1 m). Letting the variables g, S, and MR be 30 sec, 1900 vphg for one-lane turning, and 900 vph, respectively, then L becomes 167 feet (50.8 m). Obviously, somewhat longer values are desirable, but sometimes it may not be possible to provide more than one or two vehicle storage positions behind the queue detector.

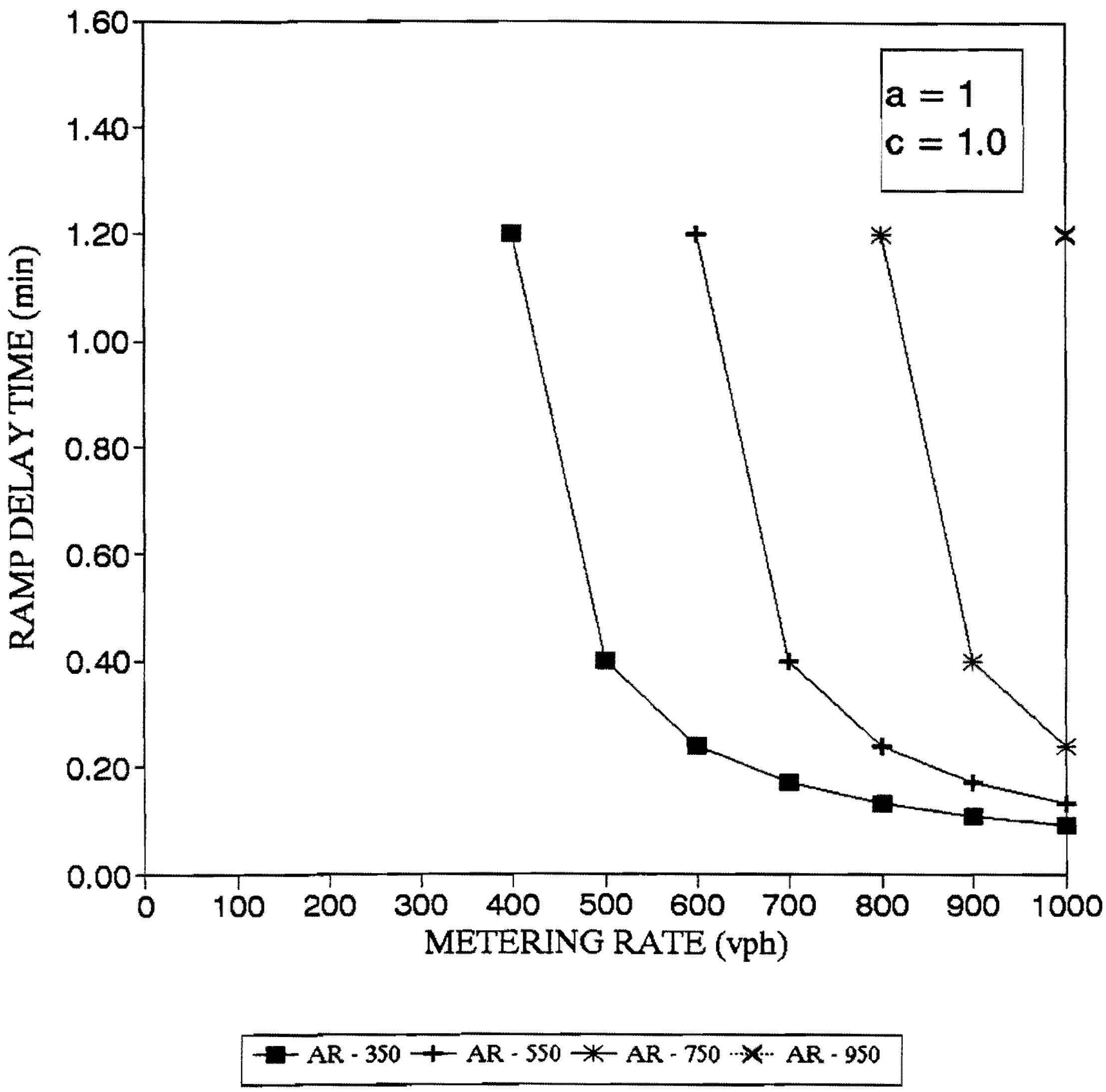


Figure 20. Theoretical Ramp Delay for Undersaturated Conditions.

Level 1 - Ramp Optimization

Three operating states of freeway traffic flow are defined at this level of ramp control: (1) free flow, (2) metered flow, and (3) congested flow. Critical occupancy values should be defined, together with the appropriate detector stations, that describe the traffic operational state at the local level. As initial critical occupancy values, upper threshold values for $O_{crit(1)}$ are about 9%, and for $O_{crit(2)}$ about 19%. Higher measured occupancy values define poorer operations and slower metering rates. Measured occupancy values less than 9 % would indicate little need for ramp metering.

Considerations of hysteresis should be made such that a 2.0% lower threshold value (ODWELL) be required to return to the prior better traffic state, and a minimum "hold" time (OHOLDJ) in the new traffic state should be provided. Perhaps OHOLDJ should equal one TFL, or six minutes. State transitions continuing in the same direction of traffic flow should not be so restricted, but perhaps limited to one state change per TFI, or one per minute. JTIC is defined as the number of times the indicator variable (occupancy) has the same trend as in previous time slices, j. A +5 would mean that the occupancy has gone up (increasing demand and density) for 5 straight j time slices. A -1 would mean that the occupancy at this time slice j is lower than the last value, but that the last value was higher than the previous one. Otherwise, the value of JTIC would have been at least -2.

For each of the seven metering levels, a Minimum Metering Level (MML) is defined, being the fastest metering level (rate) permitted for the detected freeway traffic flow. Each of these metering levels may be changed over time by higher levels of control, primarily by levels 2, 4, and 5. See Table 7 for the overall application of this control concept. Inputs would be applied explicitly to the seven metering levels used in the TxDOT ramp controller.

The following Level 1 procedures to determine optimal metering rates should be used only for the metered state defined above. The local ramp metering optimization algorithms address three control objectives in metered conditions. The basic goal is to determine the local optimal (maximum) metering rate that should not produce local traffic congestion. In addition, ramp delay will be reduced by maintaining a steady pace of freeway merging entry rate for a TFI period, or 1 minute (60 seconds) of control. No attempt will be made to adjust the metering rate to shorter-term fluctuations in measured freeway demand, such as trying to adjust metering rates to mesh into "swells between waves" (platoons) of freeway flow or trying to fit ramp vehicles into even shorter "acceptable" time gaps between vehicles. Moreover, the processor should ensure that the process calibrations are optimized.

The optimal metering rate is found from a consideration of (1) the existing ramp demand, (2) the local merging capacity, and (3) the downstream freeway bottleneck capacity. If the allowable metering rate exceeds the existing demand, then the metering rate (MR) should be set to 1.2 x the demand rate, but not less than 600 vph nor greater than 800 vph. To assess the two merging capacity constraints in real-time, a design decision has to be made regarding whether the actual freeway demand flow, either for the merging lane as F1VOL or for the total mainline demand as FVOL, will be directly measured or estimated using lane occupancy. Direct measurement of FVOL is highly recommended. However, both measures still should be assessed because lane occupancy is solely used in isolated ramp metering operations. In addition, the lane occupancy would still need to be measured to determine if free flow $\{O(t) < O_{crit(1)}\}$ or if metered/congested flow conditions exist.

To estimate freeway flow using lane occupancy O = O(t) = O(TFI), the system will have to use traffic flow theory to estimate the actual freeway flow rate. In addition, two estimates of the metering rate should be calculated and the optimal (minimum) value selected. The first estimate will be based on ramp merging capacity "N=1" and the second on total mainline capacity "N=N". The procedure first estimates the freeway density from

$$KN = 2.51 \times N \times O \qquad (39)$$

where KN is the total density for the number (N) lanes N processed, O is the measured average lane occupancy for those N lanes, and 2.51 (or other value) is the calibration value for the "trap" detector length. Then, the average freeway speed, U, is estimated from

$$U = U_f \{1 - (O/O_j)^{(l-1)}\}^{\{1/(1-m)\}} \qquad (40)$$

where "U_f" is the free speed of the freeway section and "O_j" is the jam occupancy. The shape coefficients "l" and "m" are to be determined from field data for the process or estimated in some other manner. See Figure 21 for an overview of the shape exponents.

A computerized algorithm has been prepared to estimate the shape coefficients $<l,m>$ based on estimates of freeway conditions. Figure 22 presents an example data screen of the algorithm which automatically gives the shape coefficients $<l,m>$ given the traffic flow parameters of the freeway link, either the one upstream or downstream of the metered ramp. Several related graphic files conveniently provide the traffic flow profiles together with the speed, density and occupancy as related to the flow and system state (s = 1 or s > 1). A related TTI research report provides further details on this subject (35).
Finally, the freeway flow can be calculated from the basic equation of traffic flow since freeway flow equals speed times density. Thus, flow is

$$FNVOL = U \times KN \qquad (41)$$

Given the freeway flow, either directly measured by detectors or as calculated above, the available metering capacity at the local ramp meter is calculated as the smaller of

AVAILN = minimum of a. FCAP - FNVOL .. mainlanes
b. MCAP - F1VOL .. merge

where FCAP is the local downstream bottleneck capacity or other desired target control flow level, and MCAP is the ramp merging capacity (normally about 2,000 vph). Given that FCAP

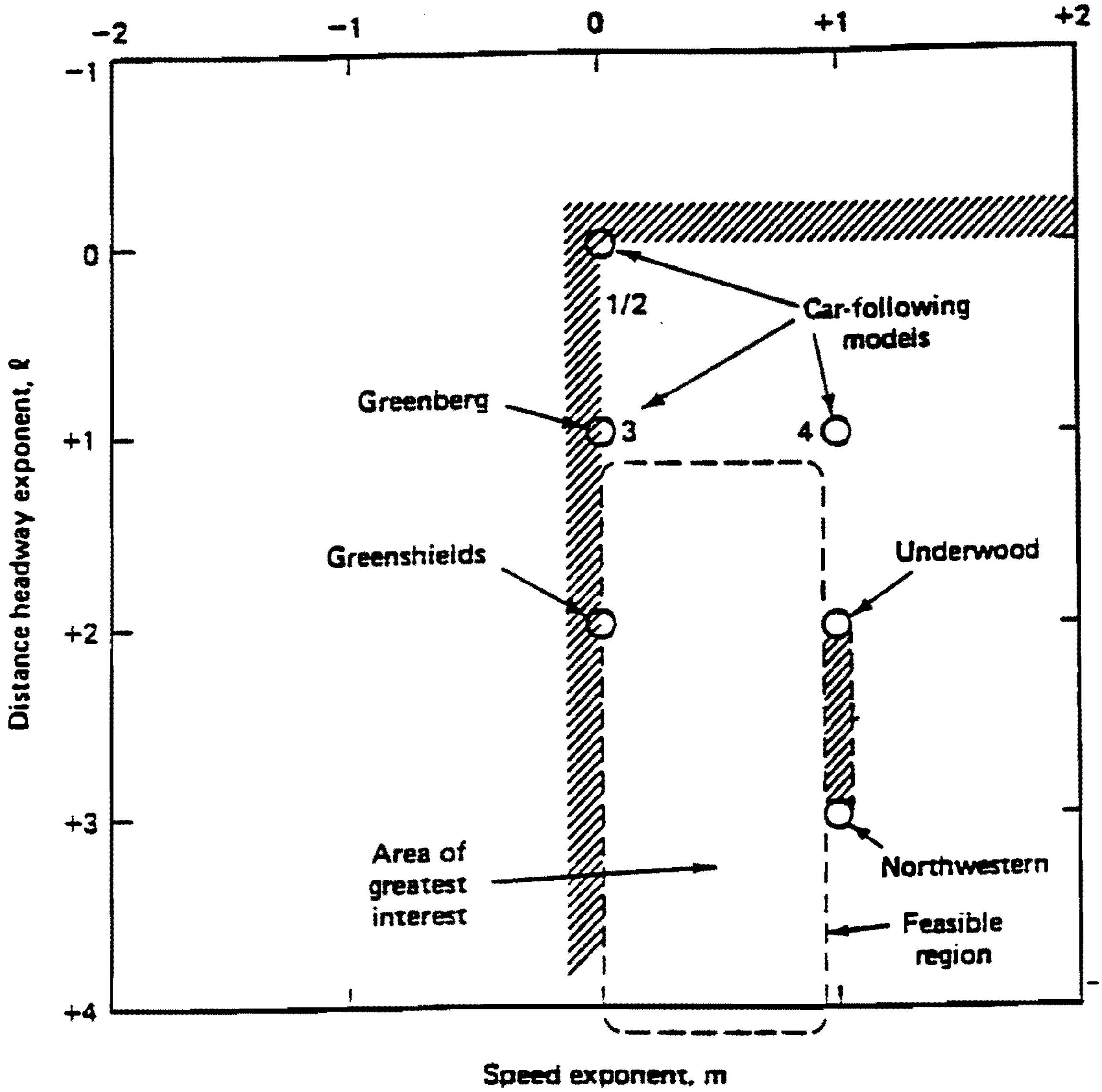


Figure 21. General Characteristics of the Shape Coefficients of Traffic Flow.

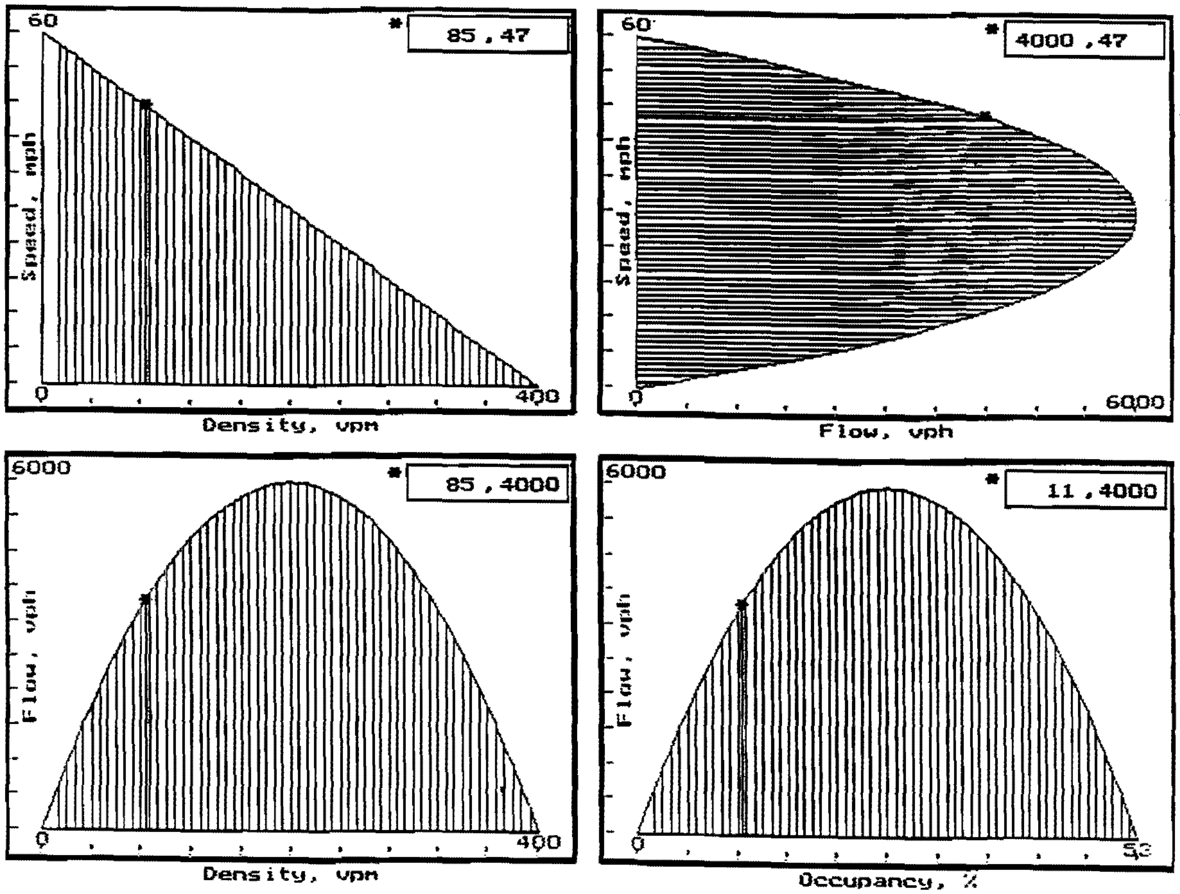


Figure 22. Traffic Assistant for Calculating Shape Coefficients.

represents the desired outflow of the metered freeway link, which may be changed over time by higher levels of control. The allowable metering rate will be dynamically changing with time as the freeway volumes change in the same time frames. Data base management techniques should provide that the FCAP and MCAP values will not be fixed values but can be modified by higher control levels ($cl = 2$ and $cl = 4$). Certainly, backup fixed values should be provided in case of real-time data processing problems.

The actively constrained metering rate (CMR) is found as the smaller of

$$CMR(t) = \min [\text{AVAIL1 or AVAILN}] \quad (42)$$

given that CMR is less than the ramp demand R. Table 11 summarizes these two ramp capacity analysis methods. Further adjustments are needed to ensure that the optimal input flow is provided in the field. The metering rate used to set the meter cycle is typically higher than the average flow by the value of ERM (1.8, or other value) if the ramp merge is active or EMF (1.2 or other value) if the freeway bottleneck is active. Improved values of EMR (for ramp merging only) and EMF (for freeway bottleneck control) will need to be determined locally by observations.

Ramp metering rates (AVAILN) for several freeway types were examined using the above traffic flow models combined with the Traffic Assistant illustrated in Figure 22. This analysis included four-, six-, eight-, and ten-lane freeways having nominal traffic flow parameters of free speed, speed at capacity, capacity flow, and jam density. The speed at capacity was assumed to be about 69% of the free speed, as compared to 50% when using Greenshield's model in Figure 21. Representative shape coefficients for the traffic flow models shown in Figure 21 were $l = 3.63$ and $m = 0.82$, respectively.

Metering rates were then calculated as the difference between the downstream capacity of the freeway and the arrival freeway input flow (as measured by the resulting lane occupancy). Thus, derived data were available between metering rate and lane occupancy. An average detection length of 21 feet (15 + 6) was assumed. Figure 23 illustrates the metering rates calculated based on average lane occupancy for several freeway conditions. These results assume that the downstream freeway bottleneck capacity is 95% of the nominal freeway capacity feeding the entrance ramp. These calculations indicate that the practical metering range for "bottleneck control" is a narrow range between about 12% and 14% occupancy for 8-lane freeways. Occupancy values below 12% would meter too fast, and values above 14% would meter too slow. Table 12 provides an extensive set of metering results for five freeway conditions (merge $l=1$ and mainline freeway $l>1$) and seven levels of bottleneck capacity constraint (70% to 100% of nominal). Again, this table was developed given the above traffic flow assumptions. Table 13 presents an important ramp metering control variable (the flow rate adjustment, FRA) where

FRA = flow rate adjustment (in vph per 1% change in lane occupancy) given the local freeway occupancy at the ramp; (43)

Table 11. Level 1 - Local Ramp Optimization Algorithms.		
Operational Estimate	Merging Capacity N = 1	Freeway Capacity N = N
1. Occupancy	O1(TSI)	ON(TSI)
2. Critical Occupancy	$< O1_{crit(1)}$? "YES"	$< ON_{crit(1)}$? "YES"
3. Density	K1 = 2.51 x O(t)	KN = 2.51 x N x O(t)
4. Speed	$U = U_f \left[1 - \left\{\frac{O(t)}{O_j}\right\}^{l-1}\right]^{\{1/(1-m)\}}$	"
5. Flow	F1VOL = U x K1	FNVOL = U x KN
6. Merge Capacity	MCAP - F1VOL	FCAP - FNVOL
7. Constrained Meter Rate	CMR = Min Merge or Freeway?	CMR=Min Mrg or Fwy?
8. Optimal Meter Rate	IF CMR > R, THEN MR =1.2 R ELSE MR = CMR	"
9. Metering Rate	MR = MR x EMR	MR = MR x EMF
10. Flow Calibration	FV = SN/SM	(See Table 13)
11. Cycle Time	C(t) = 60*Ts/{FVxMR(t)}	"
12. Red Correction Time	Rc = SC/SN - G - Y - R(ml)	(For previous "j")
13. Red Selection Time	R(ml) = C(t) - G - Y - Rc	"
14. Select Metering Level	Minimum $\lvert R_{ml} - R(ml) \rvert$	(See Table 7)

Notes: FCAP will be provided by System Control Level 3. EMR and EMF provided for each ramp. Values are about 1.8 and 1.2, respectively.

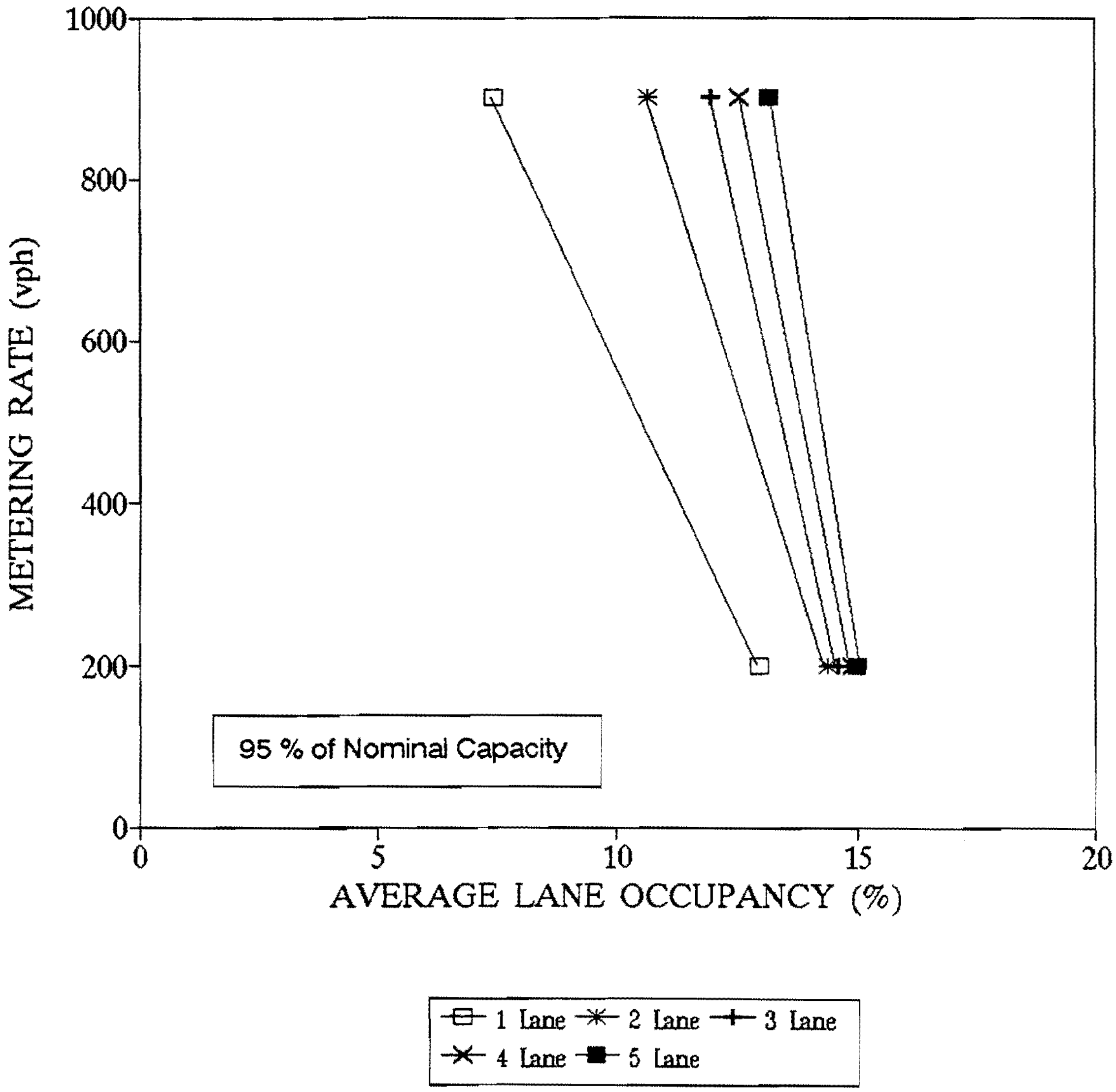


Figure 23. Ramp Metering Rates as Related to Average Lane Occupancy.

Table 12. Metering Level and Metering Rate as Related to Average Lane Occupancy for Various Freeway Lanes and Downstream Bottleneck Capacities.

Metering		Number of Freeway Lanes in Direction of Flow																				
		1[a]							2							3						
ML	Rate	Percent of Capacity							Percent of Capacity[b]							Percent of Capacity						
-	vph	70	75	80	85	90	95	100	70	75	80	85	90	95	100	70	75	80	85	90	95	100
B	900	3.9	4.5	5.2	6.0	6.6	7.4	8.2	6.8	7.4	8.1	8.9	9.9	10.7	11.7	7.8	8.5	9.3	10.1	11.1	12.0	13.2
C	800	4.4	5.2	5.8	6.6	7.4	7.9	8.9	7.2	7.8	8.5	9.3	10.1	11.1	12.2	7.9	8.7	9.7	10.3	11.5	12.4	13.6
D	720	5.0	5.6	6.4	7.2	7.8	8.7	9.3	7.4	7.9	8.9	9.7	10.7	11.5	12.4	8.1	8.9	9.7	10.7	11.7	12.6	13.8
E	600	5.8	6.4	7.2	7.9	8.7	9.5	10.3	7.6	8.5	9.3	10.1	11.1	12.0	13.2	8.5	9.1	10.1	10.9	11.8	13.0	14.4
F	480	6.4	7.2	7.9	8.7	9.7	10.5	11.5	8.1	8.9	9.5	10.7	11.7	12.6	13.8	8.7	9.5	10.3	11.5	12.4	13.6	15.0
G	360	7.4	7.9	8.9	9.7	10.7	11.5	12.4	8.5	9.3	10.1	11.1	12.2	13.2	14.6	8.9	9.9	10.7	11.7	12.6	13.8	15.6
H	240	8.1	8.9	9.7	10.7	11.7	12.6	13.8	8.9	9.9	10.7	11.7	12.6	13.8	15.6	9.3	10.1	11.1	12.0	13.2	14.4	16.1
-	200	8.5	9.1	10.1	10.9	11.8	13.0	14.4	9.1	9.9	10.9	11.8	13.0	14.4	15.9	9.3	10.3	11.1	12.2	13.2	14.6	16.5

a Merging capacity constrained ramp metering.
b A value of 95% means that the downstream freeway capacity is 95% of the freeway capacity just upstream of the merge point.

Table 12. Metering Level and Metering Rate as Related to Average Lane Occupancy for Various Freeway Lanes and Downstream Bottleneck Capacities. (cont.)

Metering		Number of Freeway Lanes in Direction of Flow													
		4							5						
ML	Rate	Percent of Capacity[b]							Percent of Capacity						
-	vph	70	75	80	85	90	95	100	70	75	80	85	90	95	100
B	900	8.1	9.1	9.9	10.9	11.7	12.6	14.0	8.5	9.3	10.1	11.1	12.2	13.2	14.6
C	800	8.5	9.1	10.1	11.1	11.8	13.0	14.4	8.7	9.5	10.3	11.5	12.4	13.6	15.0
D	720	8.5	9.3	10.3	11.1	12.2	13.2	14.6	8.9	9.7	10.7	11.5	12.4	13.8	15.0
E	600	8.7	9.7	10.5	11.5	12.4	13.6	15.0	8.9	9.9	10.7	11.7	12.6	13.8	15.6
F	480	8.9	9.9	10.9	11.7	12.6	14.0	15.6	9.1	10.1	10.9	11.8	13.0	14.4	15.9
G	360	9.1	10.1	11.1	11.8	13.0	14.4	15.9	9.3	10.3	11.1	12.2	13.2	14.6	16.5
H	240	9.5	10.3	11.3	12.4	13.2	14.6	16.6	9.7	10.3	11.5	12.4	13.6	15.0	17.1
-	200	9.5	10.3	11.3	12.4	13.6	14.9	16.9	9.7	10.5	11.5	12.4	13.6	15.0	17.1

b A value of 95% means that the downstream freeway capacity is 95% of the freeway capacity just upstream of the merge point.

Table 13. Flow Rate Adjustment as Related to Local Freeway Average Lane Occupancy and Number of Freeway Lanes.

Average Lane Occupancy %	Flow Rate Adjustment, FRA Δ vph / Δ% occ.				
	Number of Freeway Lanes				
	1 [a]	2	3	4	5
2	162.6	325.2	487.8	650.4	813.0
3	161.6	323.2	484.8	646.4	808.0
4	159.9	319.8	479.7	639.6	799.5
5	157.4	314.8	472.2	629.6	787.0
6	153.9	307.8	461.7	615.6	769.5
7	149.4	298.8	448.2	597.6	747.0
8	143.7	287.4	431.1	574.8	718.5
9	136.9	273.8	410.7	547.6	684.5
10	128.8	257.6	386.4	515.2	644.0
11	119.6	239.2	358.8	478.4	598.0
12	109.2	218.4	327.6	436.8	546.0
13	97.6	195.2	292.8	390.4	488.0
14	85.0	170.0	255.0	340.0	425.0
15	71.4	142.8	214.2	285.6	357.0
16	56.9	113.8	170.7	227.6	284.5
17	41.7	83.4	125.1	166.8	208.5
18	26.0	52.0	78.0	104.0	130.0
19	10.0	20.0	30.0	40.0	50.0
20	0.0	0.0	0.0	0.0	0.0

a Merging capacity constrained ramp metering.

A control algorithm can produce decisions only as good as the quality of the data and the assumptions upon which the decisions are based. Here we assume that the detectors are working properly. However, this does not mean that all motorists are driving according to the rules. Most metering algorithms assume that one vehicle enters the ramp per signal cycle, discounting bulk metering modes. To optimize flow control, this assumption should be assessed and calibrations in metering rates made as appropriate. The volume calibration factor for metered ramp violations can be derived from

$$FV = SN/SM \quad \text{cycles per vehicle} \tag{44}$$

where SN and SM are ramp operational variables defined in Table 8. An FV value less than 1.0 implies that the setpoint metering rate for control should be set less than the nominal value. Other tests, such as by video imaging traffic counts, may need to be performed to assess the detector counting accuracy.

The metering cycle time (seconds per vehicle) is found from the metering rate, expressed in vehicles per Ts (time slice) units of time, where Ts is the time slice (base) duration in seconds calculated from

$$C(t) = \{60 \times Ts\} / \{FV \times MR(t)\} \quad \text{in seconds} \tag{45}$$

Additionally, the metering level that should be selected must optimally account for the fact that the cycle is not actually selected but that a nominal "nearby" red time is selected. The actual average red time displayed is also unknown and must be estimated in some way. The method proposed to account for this problem in an optimal manner is as follows: determine the estimated red correction duration, Rc(t), from

$$Rc = SC/SN - G - Y - R(ml) \qquad \text{per ramp m} \tag{46}$$

in seconds for the past interval "j-1" with no queue override and $s = 1$. The above equation is equivalent to

$$Rc = G + Y + Ra - G - Y - R(ml) \tag{47}$$

such that Rc = Ra - R(ml), where Ra was the average red duration during the previous time frame, TFI, and R(ml) was the red interval of the metering level selected from Table 7 for that time frame. To illustrate the process, assume the following cycle synchronous data were known for the previous (j-1) TFI of 60 seconds:

G = 1.5 sec	ML(j-1) = F	SN = 6 cycles
Y = 1.5 sec	R(F) = 5.5 sec	SC = 57.0 sec

The value of Rc estimated for the next "j" would be based on measurements during j-1 and would be

$Rc(j) = 57.0/6 - 1.5 - 1.5 - 5.5$
$Rc(j) = 9.5 - 1.5 - 1.5 - 5.5$
$Rc(j) = 1.0$ sec

These results imply that the average ramp metering cycle was 9.5 seconds, but that the nominal cycle for metering level F was $C(F) = 1.5 + 1.5 + 5.5 = 8.5$ sec. or $Rc(j\text{-}1) = 9.5 - 8.5 = 1.0$ sec. This error suggests that if we really want an 8.0 second cycle in TFI j, then we need to set 7.0 seconds; or $R(j) = C(j) - G - Y - Rc(j)$, but satisfy $\min < R(j) < \max$. This adjustment algorithm will need to be provided at this level of ramp control for the system to produce optimal ramp flow control.

Selecting the Metering Level

The selection of the metering level with the present generation of TxDOT ramp metering controllers, whose control variables were noted in Table 7, requires the search of the nearest "setpoint" red interval for a given metering level "ml." The "ml" that provided the minimum deviation should be selected as the metering level, ml. The above ramp metering calibration procedure has been summarized in Table 11. Suffice it to note that some error in metering rate will occur due to this "round off" in expected red time, and this is not the ideal situation for control. Future ramp meter design specifications should consider providing the capability for downloading variable red times from the SCU into the meter levels shown in Table 7.

Database Requirements for Level 1, TFI

The database requirements for Level 1 - Local Ramp Metering Optimization are given in Table 14. It is assumed that the real-time traffic data have been collected during TFI time slice "j-1" and that the forthcoming TFI time frame is "j." These traffic flow data would come primarily from freeway detector stations A and B, as described in Figure 17. Upstream traffic demand is measured at station A from either detector A for mainline freeway flow or detector F for freeway merge flow. Downstream traffic status indicators (free flow, capacity flow, or forced flow) come from detector station B, and from detector B if only one lane is monitored.

Numerous other operational data will need to be collected at Level 1. Other parametric data will likewise flow into data storage registers from other higher priority levels. The principal other local operational data to be collected deals with the real-time calibration of the local ramp controller based on measurements of cycle times and displayed red times during the previous time frame (TFI), or during j-1.

Provisions for data input of parametric values should be expediently handled by the data base management system. The MML(s) parameters, in particular, are critical to the optimal operation of real-time adaptive ramp metering. Both active control MML(TFI) and update values MML(TFL) should be provided so that data exchanges can be easily made.

Table 14. Database Requirements For Level 1.

Type	Variable	Function
Fixed	NR	Ramp Number
	NC	Class Number
	NQ	Queue Storage
	NA	Detector A
	NB	Detector B
	NC	Detector C
	ND	Ramp Downstream
	G	Green Time
	Y	Yellow Time
	RMIN	Red Minimum
	ML(ml)	Meter Level, Table 6
	R(ml)	Red Time, Table 6
	Ocrit(s)	Crit. Occ. for 3 states
	ODWELL	Hysteresis Occ. Dwell
	OHOLDJ	Hysteresis Occ. Hold J
	JTICJ	Hysteresis UpTick Limit
	M	Traffic Flow, m
	L	Traffic Flow, l
	OJAM	Jam Occupancy
	UFRE	Free Speed
	MPCE	Metering PCE
Variable		
TFI	AO	Detector A Occupancy
(no TFS)	AV	Detector A Flow
	FO	Detector F Occupancy
	FV	Detector F Flow
	BO	Detector B Occupancy
	TES	Earliest Start
	TLF	Latest Finish
	MML(s)	Fastest Meter Levels
	MCAP	Merge Capacity
	FCAP	Bottleneck Capacity
	ML	Metering Level
	MR	Metering Rate (Cycle Synchronized)
	SN	Cycle Count
	SM	Merge Volume Count
	SR	Red Sum
	SC	Cycle Sum
	Ra	Red Average
	Rc	Red Correction
	JTIC	Occupancy Tick Count
TFL	MML5	Control Level 5 MML
	MML(s)	Fastest Meter Levels

Level 2 - Local Ramp Adaptive Control

This level of control provides the ramp control system with the capability to adjust values of "parameters" of the control functions without changing the functions (rules) themselves. One primary concern is whether future operations at the local ramp level are more determined by upstream or downstream factors. The overriding concern is that local ramp metering operations may become controlled by congestion backing up the freeway from a downstream traffic incident or overload.

Two downstream activities are considered as local adaptive control inputs. These include the analysis of the freeway capacity at station B in Figure 17 to better estimate FCAP, and the traffic flow state of detector station C, as also noted in Figure 17. Detector C may override other lower level indicators and define the state of the local subsystem as "s." Detector C responds every TFI.

Detector C is the "early warning" detector of the upstream progression of freeway congestion (a shockwave) which is used by ramp "m" to better determine the state of the local subsystem without having to rely on higher level system interpretations of the situation. Detector C does not have to be located at the downstream bottleneck, as does detector B. However, detector C should be located as far downstream as possible, at least as far as detector B, but not beyond the next downstream entrance ramp "m+1." In some cases, detector stations B and/or C for ramp "m" could be station A for entrance ramp m+1. The same three-state occupancy tests made at detector B would also be made at C. The critical thresholds may vary depending on if an exit ramp is located between B and C, but the basic idea is the same. If either B or C indicates "False" to traffic state 1 (implying that $s=2,3$), then a more restrictive metering level should be implemented, given "$s>1$." The technique used in RAMBO when $s>1$ is to change the upper bounds MML($s>1$) in Table 7.

Level 2 also provides the logical mechanism for adjusting or self-calibrating all of the local ramp metering parameters, such as for capacities FCAP and MCAP, that are selected for control during state of the system "s" and for reporting these adjustments to higher levels of control. These adjustments are made every TFL time frame. Categorical and continuous variable schemes (statistical in nature) may also be devised to better predict the parameters, say by day of the week, time of day, by weather conditions, etc.

Traffic volume is a basic measure of traffic flow, but it is not as simple to measure as it might appear. Consequently, the system should be able to check itself regarding volume counting, particularly between upstream and downstream count stations at each ramp. An important design decision will be whether volume (flow rate) and the capacity parameters will be based on flow units of vehicles per hour (vph), or as counted in the time frame TF, or in some equivalent time-of-occupancy-based passenger car units per hour, (pcph). In addition, the system will need to be able to make good estimates of the current bottleneck capacity, FCAP, and merge capacity, MCAP for every metered ramp each time frame TFL. Both capacity values depend on the fraction of trucks (heavy vehicles) in the traffic stream. Thus, the system will need to determine how accurately it is counting passenger cars as well as trucks, and the relative fractions of each type of vehicle.

The preferred adaptive method would appear to be to count the traffic and calibrate the capacity parameters. Firstly, a critical detection length for "cars" must be defined. This value is about 25 feet for 6 x 6 foot loop detectors operating in the presence mode. Any vehicle generating an occupancy time for a speed "u" in fps (feet per second) greater than

$$O_{car} = Lcar/u = 25/u \text{ seconds}$$

would be defined as a "truck" or "heavy vehicle." For practical purposes, the speed u can be calculated from the traffic flow equations given in Table 11 (and page 71) based on average lane occupancy. Next, the volume of vehicles counted in each of the two vehicle classes "cars" and "trucks" would be recorded as Vc and Vt, respectively, for the TFL time frame, together with the total volume, V. The average occupancy time for cars and trucks is then determined to be Oc and Ot, respectively, where presumably $Oc < Ot$. The fraction of trucks in the traffic stream is calculated as Pt, $0.00 < Pt < 1.00$.

Given the occupancy values for cars and trucks, the freeway bottleneck capacity (FCAPB) for a ramp can be estimated from the passenger cars only capacity (FCAPB0) as

$$FCAPB = \frac{FCAPB0}{1 + Pt\left(\frac{1.25 + Ot}{1.25 + Oc} - 1\right)} \tag{48}$$

where we have assumed that the minimum intervehicular headway is 1.25 seconds at capacity. This value is considered a parameter under normal traffic conditions. Similar adjustments would also apply for the merging capacity MCAP if the same traffic conditions are assumed. However, the percentage of trucks may be higher in the merging lane than for the freeway as a whole. Thus, it is recommended that separate adjustments be made for the mainline and merge capacities for each metered ramp. The initial capacity values FCAPB0 and MCAP0 for the ramp could also be changed by higher system control levels.

It is very important to note that the available downstream mainline capacity, the total allowable flow downstream from the metered ramp is FCAP, which is derived from FCAPB. Control Level 3, the system optimization function, will determine the value of FCAP from

$$FCAP = FCAPS = FCAPB - SSB \tag{49}$$

where SSB is the system-based slack for the local bottleneck that provides optimal freeway operations. SSB may be any non-negative value, but it will usually not be greater than the local ramp demand. Once FCAPS is determined by system optimization Level 3, RAMBO will simply modify the lower Level 1 control parameter, FCAP, for the next control interval "TFL = k", such that FCAP = FCAPS at control Level 1. Should telecommunications fail between

Levels 2 and 3, then Level 2 should set FCAP = FCAPB - SSBE (estimated) in Level 1. All modifications to the local ramp merging capacity MCAP are made at Level 2.

Database Requirements for Level 2, TFI & TFL

The several adaptive functions of Level 2 dictate its data requirements. Recall that this level basically serves three adaptive functions: (1) its early warning of potential shockwaves coming from downstream congestion, (2) its calculation of updated ramp merging capacity, and (3) its calculation of adjusted downstream freeway bottleneck capacity at station B. It also probably should be assigned the role of calculating the local lane distributions at stations A and B to determine if unusual traffic flow patterns are occurring that should be addressed either locally or flagged to higher levels of control. Thus, this design will assume that the fractions of vehicles using all lanes at stations A and B will be determined at Level 2 for local use as needed and uploaded to higher control levels for subsequent use therein. Table 15 summarizes the data requirements for Level 2.

Table 15. Database Requirements For Level 2.

Type	Variable	Function
Fixed	IVEH = 1.2 sec	Intervehicular Headway at Capacity
	LCAR = 25 ft	Critical Car Detection Length
	MCAP0 = 2,000 pcph	Base Merge Capacity
	FCAP0 = 2,200*N pcph	Base Bottleneck Capacity
	Ocrit1	Critical Occupancy (states 1, 2)
	Ocrit2	... Measured at Station C
	SSBE	Estimated Capacity Slack
	"*l*"	Traffic Flow Parameter, *l*
	"*m*"	Traffic Flow Parameter, *m*
Variable		
TFI	PL1A	Percent Traffic In Merge Lane
(no TFS)	PT1A	Percent Trucks in Merge Lane
		(at Stations A or B)
TFL	PLA(n)	Percent Volume Per Lane, A
	PLB(n)	Percent Volume Per Lane, B
	PTL1A	Percent Trucks In Merge Lane, A
	PTB	Percent Trucks, B
	Oc	Occupancy of Cars, B
	Ot	Occupancy of Trucks, B
	FCAPB	Bottleneck Capacity, B
	MCAP	Merging Capacity for Ramp m
	SSB	Download System Slack for B

Level 3 - System Optimization

The basic goal of freeway system optimization is to determine an optimal set of metering rates for each ramp per time frame TFL that maximizes the freeway flow without exceeding any bottleneck capacity threshold for any section along the freeway system. This system algorithm uses the linear programming formulation described in Chapter 3 and will not be repeated here. This algorithm is only used in this design when there is no congestion detected within the freeway system. This implies that the "state of the system" must be in state 2, or stable, metered-flow conditions with no latent demand (queues) already on the freeway. When the state of the system is not "2" then control reverts to other control algorithms either programmed in Levels 1 or 4 with this design.

The control outputs from Level 3 to each metered ramp are fairly simple. For each TFL, two outputs are delivered to lower levels. Level 2 receives an update of SSB. Level 1 receives an update of FCAP = FCAPS. Level 1 also may receive updates of SSM(s) and FML(s), depending on future design refinements. The control variables (VOL(B) and FCAPS) will be calculated by the linear program in Level 3 for each metered ramp:

$$SSB = FCAPB - VOL(B), \quad \text{if } RQ > 0 \tag{50}$$

$$SSB = 0, \quad \text{if } RQ = 0 \tag{51}$$

$$FCAP = FCAPS = FCAPB - SSB \tag{52}$$

where RQ is the number of vehicles not served (queued) at the ramp. Then SSB will be downloaded to Level 2; FCAPS, the system adjusted optimal throughput capacity at ramp m is then downloaded to Level 1, which will be used therein at the next update time to modify FCAP in the local ramp metering equation for freeway bottleneck capacity. Note that this design implements the system functions though the lower level control algorithms.

Database Requirements for Level 3, TFL

To provide integrated freeway system-based ramp metering, a large quantity of data needs to be provided. These data needs include the best estimate of all fixed facility and control parameters as well as present and forecasted values for the "real-time" TFL traffic variables for each time slice "k." The time slice durations should normally be 5, 6 or 15 minutes, depending on the dynamics of the local traffic as measured by the system. In any case, the time slice duration should be some integer number calculated from 3600/NTSH, where NTSH is the number of time slices per hour. Table 16 provides a comprehensive list of candidate variables required by Level 3 control. Links are the freeway mainlane sections between either entrance or exit ramps. Traffic flow and measurement parameters will be needed for each freeway link. Data have to be provided for each metered entrance ramp, including ramp metering parameters and maximum queue storage.

Table 16. Database Requirements For Level 3.

Type	Variable	Function
Fixed		
System:	NSYS	System Number
	NLKS	Number of Freeway Links
	NENT	Number of Entrance Ramps
	NEXT	Number of Exit Ramps
Control:	NTSL	Number of Time Slices
	NTSH	Number of Time Slices/Hour
	EXPB	Smoothing Factor B
Links:	LNUM	Freeway Link Number
	LTYP	Freeway Link Type (I,N,X,O)
	DISL	Link Distance, Feet
(per link)	NUTL	Link Number of Thru Lanes
	NUAL	Link Number of Auxiliary Lanes
	FCAPO	Link Base Capacity, pcph
	UFRE	Link Free Speed, mph
	DENJ	Link Jam Density, vpm
Entrance Ramps:	ERAM	Metered PCE (~1.83)
	CMRG	Nominal Merge Capacity (pcph)
	CMAX	Maximum Metering Rate (vph)
(per ramp)	CMIN	Minimum Metering Rate (vph)
	QMAX	Maximum Queue Storage (veh)
	DIVR	Ramp Queue Diversion Fraction (%)
	FMER	Nominal Fraction Merge Lane (%)
(for all exits)	EIJM	Connectivity Factor to Exit Ramps
Exit Ramps:	CXIT	Exit Ramp Capacity (vph)
Variable (TFL)		
System:	(None)	(None)
Control:	(None)	(None)

Table 16 (continued)

Links:	FCAPB	Operational Capacity
	FCAPS	Optimal Throughput Flow (vph)
	FOCCS	Optimal Throughput Occupancy (%)
	SSTAT	Flow State (s)
	QUESG	Queue Storage, (veh)
	UCAP	Capacity, Speed (mph)
	KCAP	Capacity, Density (vpm)
	OCAP	Capacity, Occupancy (%)
	"l"	Flow Exponent, l
	"m"	Flow Exponent, m
	DETC	Average Detector Length (ft)
	FLOWN	Flow Rate Input (vph)
	FLOWO	Flow Rate Output (vph)
	FOCCN	Freeway Occupancy In (%)
	FOCCO	Freeway Occupancy Out (%)
	SLOWN	Storage Flow In, veh
	SLOWO	Storage Flow Out, veh
	EFLON	Error Flow In
	EFLOO	Error Flow Out
Entrance Ramps:	VOLC	Ramp Volume Count
	VOLS	Ramp Volume Stored
	VOLP	Ramp Volume Predicted
All (TS)	QUER	Ramp Queue Detection
	RQ	Ramp Queue, veh
	FTL1	Fraction Flow Merge Lane
Exit Ramps:	VOLC	Ramp Volume Count
	VOLS	Ramp Volume Stored
All (TS)	VOLP	Ramp Volume Predicted

One major issue that must be resolved early in the data input design is whether traffic flow data will be input and displayed based on some standard rate of flow, such as in vehicles per hour (vph) which is then internally converted by software to the time base of operation, or whether the input data will be coded and displayed based on the active time base duration. System level 3 works best internally in the linear programming formulation when smaller numbers are used as long as they are not too small, perhaps less than 10. To illustrate this issue, assume that the capacity of the freeway is described as being 6,000 vph. Then the input capacity per time slice of duration TFL is 6,000/(60/TFL) or 6,000/NTSPH.

Level 4 - Freeway Adaptive Functions

Adaptive control identifies relevant changes occurring within and upon the system and the appropriate responses to those changes. Three adaptive control strategies have been identified for ramp metering: changes in traffic parameters for a given traffic state, changes in traffic state for a given system environment, and changes in system environment. In all of the above cases, the basic ramp metering control rules do not change.

Traffic Functions. Firstly, there are those adaptive functions that continually "fine tune" the traffic parameters used in the optimization algorithms for a given state of the traffic stream (s). Routine updating and fine tuning of those parameters and coefficients used in system level control (Level 3) are desired to improve system performance. The need for parameter calibrations may be due to (1) limitations in the original databases used for developing empirically-based relationships, (2) slow changes that may be occurring in the traffic stream over time, (3) changes that may be occurring in the environment of the control system itself, (4) or due to other unknown reasons.

Several adaptive "traffic" functions are recommended in this first design version of RAMBO with other functions surely to follow. The first traffic adaptive function deals with a proposed enhancement of the exponential smoothing algorithms. Two aspects of these algorithms should be regularly "fine-tuned" by the system. These deal with the assumed parameter "b" in the models, whose initial value is estimated to be 0.3 for all levels of control. The second concerns the effective lag time of the exponentially smoothed curve as derived from traffic data as measured for typical dynamic traffic scenarios. Typical traffic demand curves will be monitored to determine how far "ka" lags "ko" for typical growth/decay curves. The lag in "k" is related to the "a" coefficient in the exponential smoothing equation. This refined estimate will improve the ability to forecast future demand volumes used for system optimization.

The second traffic adaptive algorithm addresses the overall problem of continually fine tuning the estimates of the flow percentages in the merging lanes as control and traffic state parameter (s) change with time. It is essential that good estimates of PF1 be estimated for each metered ramp. An equilibrium algorithm was provided in Chapter 2 (Equation 8) which could be used in initial efforts to better estimate merging lane distributions.

A third traffic adaptive algorithm needs to be provided for real-time origin-destination estimation of traffic flow along the freeway system. An effective model is needed to provide quality estimates of current travel patterns along the freeway system from every input to each output. The algorithm should be capable of responding to local traffic, system, and environmental changes. A second generation adaptive origin-destination predictive algorithm (*23*, *36*) is under development at TTI that could serve this need.

State Functions. It is essential that the control system know each freeway link's traffic state (s) and be able to determine the overall state of the freeway system. Incident detection algorithms must be provided to determine if congestion is occurring, and resolve what is the probable cause of the congestion (e.g., accident in left lane), what will be the Freeway Traffic Management (FTM) response to this incident, and what are the best Level 4 ramp metering adaptive functions to be implemented on this freeway. Control responses are a priori defined initially, which might then be manually overridden by Level 5 inputs.

Several control parameters, including the state of the system along a string of ramps (s1, s2, s3), might be used to define selected control responses. These responses may be modified based on the queueing state at the local ramps. The FML(ml) may be modified by Level 4 based on an overall analysis of the global traffic situation based on previous experience. The Denver response algorithm described at the end of Chapter 2 should be considered as a viable candidate strategy.

Incident detection algorithms are also under development by TTI and other organizations. A related TTI report is under development which addresses exclusively the subject of freeway incident detection for FTM systems (*37*). Those incident detection and response algorithms selected would be inserted into RAMBO at this point in its design.

Environmental Functions. These adaptive functions respond to changes detected in the system environment that are affecting the control process. A myriad of impacts might arise; therefore, the wise selection of potentially significant "threats" to be monitored should be made so that limited resources can be used wisely. Initial concern should be on monitoring the design and control assumptions of the process to see if they are still valid. Then consideration should be given to foreseen and unforeseen threats to system security and performance, such as component failures, broken lines, vandalism, highway maintenance, and extreme temperatures. Then consideration could be given to forseen and detectible disturbances that might reduce system efficiency if not responded to, such as inclement weather, an unusually large number of trucks, and other special events.

Database Requirements for Level 4, TFS, TFI, & TFL

The required data for Level 4 will depend on the adaptive strategy being implemented. Higher priority items, like incident detection, may require fast response times, like TFS, due to the increasing cost of delayed response. TFI and TFL data should suffice in most cases where "catastrophes" would not occur.

5.0 CONCLUSIONS AND RECOMMENDATIONS

INTRODUCTION

This study has proposed, developed, and conducted preliminary evaluations of an advanced freeway ramp metering and control system. The hierarchical design framework provides a simple but powerful building-block approach to designing and implementing real-world freeway ramp metering systems. The multilevel hierarchical control system identifies the control tasks to be performed at each level, the data requirements for each level, and the data linkages required between each level. Efficient algorithms were presented for each level to perform the related control tasks.

These control concepts were developed without the assistance of modern real-world freeway control systems or real traffic data. Some modification of the algorithms should be expected as the true dynamics of TxDOT's Freeway Traffic Management Systems come on line in future years. To be sure, several of the recommended algorithms have their roots in the venerable Gulf Freeway Surveillance and Control Project conducted along IH-45 in Houston in the late 1960's and in the Dallas Corridor Project along North Central Expressway in the early 1970's.

CONCLUSIONS

An efficient systems-level modelling approach for determining optimal ramp metering rates was developed based on linear programming techniques. The implementation of these algorithms in real time would be through a system called "RAMBO" for Real time Adaptive Metering for Bottleneck Optimization. This methodology can develop either of two optimal freeway flow rates: either based on (1) the optimal freeway mainline output flow from each merge point along the freeway, or (2) the optimal metering rate that feeds each merge point along the freeway. RAMBO can tell which strategy is optimal. The control strategy also simultaneously maximizes system throughput and minimizes system delay (the sum of the freeway and ramp delays) over the control period as defined by the sum of the consecutive time slices.

The first version of operational analysis software for pre-implementation studies of ramp metering has been designed to handle a maximum of eight input entrance ramps over eight time slices. The Dynamic Model formulation, with queue management functions, has been successfully implemented in the pre-implementation software.

This research also compared the results produced by the optimal linear programming solutions for system-based ramp metering control developed from both the sequential solution and dynamic solution approaches. Several observations were noted from this system evaluation of the linear programming algorithms:

1. To "balance out traffic operations" and simultaneously satisfy the requirements of freeway and interconnecting arterial systems, a queue-based freeway system management scheme should (and can) be achieved.

2. The simultaneous solution approach consistently outperformed the sequential solution approach for balancing the overall freeway system input volumes and queue lengths over different ramp and time periods.

3. The diversion factor, d, was effective in adjusting on-ramp queue lengths. The diversion factor, a function of queue length, can be calibrated to the drivers' behavior at individual ramp locations.

The simultaneous solution approach clearly outperformed the sequential solution on balancing freeway volumes and queue lengths. In addition, the new simultaneous approach can be easily converted to the sequential approach if desired. In order to properly model system control needs during incidents and urban congestion, freeway-and-arterial system integration should (and can) be considered using different diversion percentage values.

Due to memory constraints, the initial experiment analyzed a freeway system with eight (8) entrance ramps, and four (4) time periods. The system will eventually be expanded to analyze a freeway system with a maximum of three (3) subsections, twenty (20) entrance ramps, and eight (8) time periods. Using the linear programming code developed by TTI the system provides results similar to the proprietary LINDO code. This study is currently investigating the design of a user-interface input/output program and several other feasible freeway ramp management alternatives to improve the integration of freeway ramp metering control with neighboring interchanges and signalized intersections.

Two types of ramp metering systems can be developed which provide hierarchical "systems-based" ramp metering: (1) a design that is based on direct flow rate calculations described in detail in Chapter 4, and (2) a design that is based on indirect estimation of flow rates using lane occupancy, as alluded to in Tables 11, 12, and 13. The following recommendations provide algorithms for both methods. In addition, some recommendations have been developed outside the immediate scope of this research report, but have been included herein for completeness.

RECOMMENDATIONS

The real-time, adaptive, integrated ramp metering system described in Chapter 4 is recommended for implementation. To expedite immediate implementation using TxDOT's ramp metering controllers, the occupancy-based design to follow is also recommended.

Some of the more important recommendations offered for real-time, adaptive, integrated ramp metering include the following items for multilevel control:

Level 0

1. Meter uniformly over extended time frames of 3-5 minutes.
2. Consequently, all real-time traffic control data should be exponentially smoothed over the desired time frame using the algorithm described in Table 6.
3. A metering rate of 600 vph provides a good tradeoff between delay time and time to respond to green onset. This metering rate should be used when the ramp demand limits the actual ramp flow.
4. Merge detector operations (when used) should be based on either (a) resume metering one cycle after merge point detector activation (with a 2 second delay time for vehicle passage over the detector), or (b) locate the merge point detector at merge queue position no. 2 since one service time is about one meter cycle.
5. Multi-point queue detectors (Qi) (perhaps provided by video imaging) would provide attractive operational features. For example, when Q1 comes on, the metering rate could rise to a desirable 600 vph; when and if Q2 comes on, the queue override metering rate could rise to 900 vph, which might relieve a queue spillback problem at some increased risk of drivers running the red signal.

Level 1

Total multilevel ramp metering optimization can be implemented in a first-generation occupancy-based ramp metering system in a straightforward manner as follows:

$$O_o = O_r + [\, o_1 + o_2 + o_3 + o_4 \,] \ <?> \ O_{ML} \qquad (53)$$

where

O_o	=	Optimal control occupancy (in percent) for ramp r. This value is then used with the nominal (100%) O_{ML} value given in Table 12 to select the metering level (ML) in Table 7, TxDOT's ramp control table.
O_r	=	Current, real-time measured lane occupancy approaching ramp r.
o_1, o_2, o_3, o_4	=	Occupancy adjustments from higher control levels to follow. These adjustments can be added to O_r as shown above or subtracted from O_{ML}, depending on which method is most convenient for selecting ML.

The Level 1 incremental occupancy adjustment, o_1, could be obtained from

$$o_1 = \frac{\text{Upstream Capacity - Downstream Capacity (in vph)}}{\text{FRA}} \qquad (54)$$

where FRA is given in Table 13. However, as shown in Table 12, this "adjustment" in occupancy is already available in modified tables of $O_{ML} = O_{ML} - o_1$, for example, as illustrated in Table 12 for 70, 75, 80, 85, 90, and 95% levels of the nominal (100%) capacity bottleneck. FRA depends on the upstream freeway occupancy and number of freeway lanes.

Level 2

Level 2 provides the local ramp metering error correction term *(35)* such that the meter will output the desired flow rate based on real-time local bottleneck calibration.

$$o_2 = \frac{\text{Output Volume Error (Measured - Target)}}{\text{FRA}} \tag{55}$$

where the "volume error" is the difference between the actual measured flow across the N freeway lanes at the downstream bottleneck and the last desired or "target" flow rate for the same time period. FRA is obtained from Table 13 for a given occupancy at the ramp. This adjustment process assumes that queueing has not backed across the downstream detectors, B. If this is not the case, i.e., $O_B > O_{crit\,2}$, then the measured flow should be set to the target flow, such that the adjustment o_2 is zero. To provide this local ramp optimization function, future TxDOT ramp metering systems should be able to count traffic at detector station A (arrivals) for the local ramp meter and be able to transmit this same volume data to the next upstream ramp meter controller. Station A becomes station B in Figure 17.

Level 3

Level 3 provides the system adjustment to insure full utilization of the freeway system without overloading any of the downstream bottlenecks. The system adjustment is the optimal reduction in flow below the capacity of the first downstream bottleneck, or

$$o_3 = \frac{\text{Downstream Capacity Flow Reduction (SSB, Eq. 50)}}{\text{FRA}} \tag{56}$$

SSB can be provided either by RAMBO (See Eq. 50.), or estimated by some other means.

Level 4

Level 4 provides the system adjustment to the occupancy algorithm that accounts for traffic congestion that may be expanding upstream from downstream detector station B.

$$o_4 = O_B(t) - O_{B\,crit\,2} > 0 \mid O_B(t) > O_{B\,crit\,2} \tag{57}$$

where

$O_B(t)$ = real-time measured occupancy at downstream bottleneck B at time t, given that this occupancy is greater than $O_{B\,crit\,2}$.

$O_{B\,crit\,2}$ = occupancy at station B that defines onset of congestion (about 23%).

Future TxDOT ramp metering systems should be able to easily transmit arrival occupancy for a ramp to the next upstream ramp controller when congestion is occurring downstream.

REFERENCES

1. Herrick, G.C. "Strategies for Improving Traffic Operations at Oversaturated Signalized Diamond Interchanges." Thesis in Civil Engineering. Texas A&M University, May 1992.

2. Hanks, Jr., J.W., and T.J. Lomax "Roadway Congestion in Major Urban Areas: 1982 to 1988." Texas Transportation Institute Research Report 90-131-3, July 1990.

3. Drew, D.R. "Traffic Flow Theory and Control." McGraw-Hill, New York, 1968, p. 218.

4. Hutchinson, T.P. "Delay at a Fixed Time Traffic Signal II: Numerical Comparisons of Some Theoretical Expressions." Transportation Science, Vol. 6, No. 3, 1972, pp. 286-305.

5. Urbanik, T. "Understanding Forced Flow." ITE Journal, October 1986, pp. 23-26.

6. Banks, J.H. "Two-Capacity Phenomenon at Freeway Bottlenecks: A Basis for Ramp Metering?" Transportation Research Record 1320, 1991, pp. 83-89.

7. Everall, P.F. "Urban Freeway Surveillance and Control, A State of the Art." Federal Highway Administration, November 1972, p. 32.

8. Goolsby, M.E. "Influence of Incidents on Freeway Quality of Service." Highway Research Record 349, 1971, p. 44.

9. Jones, E.R., Goolsby, M.E., and K.A. Brewer. "The Environmental Influence of Rain on Freeway Capacity." Highway Research Record 321, 1970, pp. 74-82.

10. Kuhne, R. "Macroscopic Freeway Model for Dense Traffic in Start-Stop Waves and Incident Detection." Proc. 9th International Symposium on Transportation and Traffic Theory, VNU Science Press, Utrecht, Germany, 1984.

11. Dudek, C.L., and C.J. Messer. "Incident Detection on Urban Freeways." Transportation Research Record 495, 1974, pp. 12-24.

12. Wattleworth, J.A. "Peak Period Control of A Freeway System--Some Theoretical Considerations." Chicago Area Expressway Surveillance Project, Report 9, 1963.

13. Wattleworth, J.A. "Systems Demand-Capacity Analysis on the Inbound Gulf Freeway." Texas Transportation Institute Research Report 24-8, 1964.

14. McCasland, W.R. "Freeway Ramp Control System." Texas Transportation Institute Report 24-26F, 1969.

15. Robinson, J., and M. Doctor. "Ramp Metering: Does It Really Work?" ITE Compendium of Technical Papers, 1989, pp. 503-508.

16. Various Project Status Reports Presented at Freeway Operations Committee Meeting of the Transportation Research Board, Washington, D.C., January 1992 and 1993.

17. "Traffic Control Systems Handbook." Federal Highway Administration, Washington, D.C., FHWA-IP-85-11, April 1985.

18. Blumentritt, C.W., Pinnell, C., and W.R. McCasland. "Guidelines for Selection of Ramp Control Systems." NCHRP 232, Washington, D.C., May 1981.

19. Drew, D.R., Brewer, K.A., Buhr, J.H., and R.H. Whitson. "Multilevel Approach to the Design of a Freeway Control System." Highway Research Record 279, 1969, pp. 40-55.

20. Sumner, R. "A Freeway Management Handbook, Volume 2, Planning and Design." Federal Highway Administration, Washington, D.C., May 1983.

21. Buhr, J.H. "Freeway Entrance Ramp Merging Control Systems." Ph.D. Dissertation in Civil Engineering, Texas A&M University, January 1967.

22. "Highway Capacity Manual." Transportation Research Board Special Report 209, 1985.

23. Messer, C.J. "A Design and Synthesis of a Multilevel Freeway Control System and a Study of its Associated Operational Control Plan." Ph.D. Dissertation in Civil Engineering, Texas A&M University, August 1969.

24. May, A.D. "Traffic Flow Theory Fundamentals." Prentice Hall, 1990, p. 306.

25. Haenel, H.E., Kosik, A.H., and B.G. Marsden. "Innovative Uses of Traffic Responsive Control in Texas." Engineering Foundation Conferences, 1983, p. 108.

26. Lipp, L.E., Corcoran, L J., and G.A. Hickman. "Benefits of Central Computer Control for Denver Ramp-Metering System." Transportation Research Record 1320, 1991, pp. 3-6.

27. Wattleworth, J.A., and D.S. Berry. "Peak-Period Control of a Freeway System - Some Theoretical Investigations." Highway Research Record No. 89, 1965, pp. 1-25.

28. Imada, T., and A.D. May. "FREQ8PE: A Freeway Corridor Simulation and Ramp Metering Optimization Model." Report UCB-ITS-RR-85-10, Institute of Transportation Studies, University of California, Berkeley, California, 1985.

29. Yuan, L.S., and J.B. Kreer. "Adjustment of Freeway Ramp Metering Rates to Balance Entrance Ramp Queues." Transportation Research 5, 1971, pp. 127-133.

30. Chen, C., Cruz, Jr., J.B., and J.G. Paquet. "Entrance Ramp Control for Travel-Rate Maximization in Expressways." Transportation Research 8, 1974, pp. 503-508.

31. Papageorgiou, M. "A New Approach to Time-of-Day Control Based On a Dynamic Freeway Traffic Model." Transportation Research 14B, 1980, pp. 349-360.

32. Kim, Y.K. "Development of Optimization Models for Signalized Intersections During Oversaturated Conditions." Ph.D. Dissertation in Civil Engineering, Texas A&M University, December 1990.

33. Kim, Y.K., and C.J. Messer. "Traffic Signal Timing Models for Oversaturated Signalized Interchanges." TTI Research Report 1148-2, January 1992.

34. Papageorgiou, M. "Lecture Notes in Control and Information Sciences." Springer-Verlag, Berlin/Heidelberg, # 50, 1983.

35. Papageorgiou, M., Hadj-Salem, H., and J.M. Blosseville. "ALINA: A Local Feedback Control Law for On-Ramp Metering." Transportation Research Record 1320, 1991, pp. 58-64.

36. Park, E. "Critical Assessment of the Features of Two Ramp Metering Optimization Models." Ph.D. Dissertation in Civil Engineering, Texas A&M University, December 1993.

37. Balke, K.N. "An Evaluation of Existing Incident Detection Algorithms." TTI Research Report 1232-20, 1993.